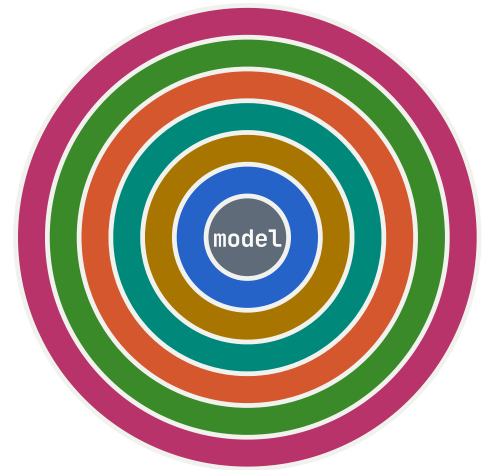


State of Harness

October 2026 • Published Oct 29, 2026

The model is the engine. The harness is the car – and in 2026 the car became the product.

The first State of Harness report defines the agent harness and reads the market around it: who uses harnesses and for what, open versus lab-owned harnesses, the move to orchestration, the pain points that decide budgets, what developers say on Hacker News, Reddit, X and Substack, and the startups building the next layer.



Directory counts are provisional: read on Oct 29, 2026 and refreshed at month end.

1,265

startups, projects, events, news and papers tracked

29

new harness startups added to the directory this edition

Cost

the loudest pain across HN, Reddit, X and Substack • [evidence](#)

CONTENTS

- | | |
|---------------------------------------|------------------------------------|
| 01 Executive summary | 02 Overall summary |
| 03 Defining the category | 04 How we got here |
| 05 The Harness Market Map | 06 Adoption & usage |
| 07 Open vs closed | 08 Orchestration |
| 09 Context, memory, tools & sandboxes | 10 Trust, evals, cost & governance |
| 11 The conversation | 12 Pain points → opportunities |
| 13 New harness startups | 14 Playbooks |
| 15 What to watch | 16 Methodology |
-

01 EXECUTIVE SUMMARY

10 things to know about the agent harness in October 2026

The 90-second read. Each claim links to the section with the evidence.

01 ■ Feb 5 → Mar 10

The category got its name in one quarter, after the market already existed. ^[1]

[Evidence →](#)

02 ■ 66%

Coding agents are now the default developer interface: two in three developers who use AI work in one. ^[2]

[Evidence →](#)

03 ■ \$2.5B+

Money followed adoption fast. Claude Code passed a \$2.5B run-rate by February (vendor-reported). ^[3]

[Evidence →](#)

04 ■ 23% → 52%

The harness moves results as much as the model: swapping only the harness took one model from 23% to 52% on SWE-bench Pro. ^[4]

[Evidence →](#)

05 ■ 4.7×

The harness is also the meter. In a small test, one harness sent about 33k tokens before reading the prompt; another sent about 7k. ^[5,6]

[Evidence →](#)

06 ■ 28%

Cost and limits are the loudest everyday pain, yet only 28% of organisations have reviewed their AI spending data. ^[2]

[Evidence →](#)

07 ■ 15 days

Open harnesses are the hedge against lock-in, and they grow faster each time: the newest reached 200K GitHub stars in 15 days. ^[7]

[Evidence →](#)

08 ■ 5–7

Orchestration is the new frontier, but practice lags the pitch: the orchestrator that markets hundreds of agents says it runs 5–7 reliably today (vendor). ^[8]

[Evidence →](#)

09 ■ 29.5%

Verification, not generation, is the binding constraint: 29.5% of teams building agents run no evals at all (vendor survey).^[9]

[Evidence →](#)

10 ■ \$144.5M

Among new harness-adjacent startups, disclosed 2026 rounds went mostly to memory and context (\$144.5M, two-thirds of it one round), security and sandboxes; orchestrators raised \$33.5M. The mega-rounds went to established vendors such as Factory and Cognition.

^[10,11,12,13,14,15,16,17,18]

[Evidence →](#)

SOURCES

1. [LangChain, The anatomy of an agent harness](#) · Vendor-reported
2. [Stack Overflow Developer Survey 2026 results](#) · Independent
3. [Anthropic, Series G announcement \(Feb 2026\)](#) · Vendor-reported
4. [Joel Niklaus on X: 10 harnesses on SWE-bench Pro](#) · Independent
5. [Systima, Claude Code vs OpenCode token overhead \(2–3 runs, one machine\)](#) · Independent
6. [Hacker News: Claude Code vs opencode token overhead \(706 pts\)](#) · Community post
7. [Star History, deepseek-harness is the fastest to 200K stars](#) · Independent
8. [Paradev \(formerly Superset\), Roadmap to 100 agents](#) · Vendor-reported
9. [LangChain, State of Agent Engineering \(survey, n=1,340\)](#) · Vendor-reported
10. [Pulse 2.0: Engram launches with \\$98M](#) · Press report
11. [Interloom, seed announcement](#) · Vendor-reported
12. [TAMradar: SageOx \\$15M seed](#) · Press report
13. [Cognee, seed announcement](#) · Vendor-reported
14. [Dealroom: Uvi raises \\$7.5M](#) · Press report
15. [Conductor, Series A announcement](#) · Vendor-reported
16. [Paradev, Superset is now Paradev](#) · Vendor-reported
17. [The Next Web on Factory's \\$5B round](#) · Press report
18. [The Next Web on Cognition's Series E](#) · Vendor-reported

02 OVERALL SUMMARY

The state of the harness, in one page

The model is the engine. The harness is the car — and in 2026 the car became the product.

By October 2026 the agent harness is no longer a hidden implementation detail. It is the loop, the tools, the context, the sandbox and the guard-rails that turn a model into a working agent, and it is the product through which most professional software now gets written. Developers adopted it before anyone named it: agent use among Stack Overflow respondents went from about 31% in 2025 to 59% in an April 2026 pulse, and in the 2026 survey 66% of developers who use AI work in coding agents.^[1,2,3]

The vocabulary arrived in a burst between February and March 2026, when Mitchell Hashimoto, OpenAI and LangChain all wrote about "harness engineering". Money had already arrived. Claude Code passed a \$2.5B run-rate (vendor-reported), SpaceX completed a roughly \$60B acquisition of Cursor, and Factory's valuation went from \$300M to \$5B in about a year.^[4,5,6,7,8,9]

What the harness decides most reliably is cost. Benchmarks moved by 13 to 29 points when only the harness changed, depending on study and benchmark, but independent

work also finds success rates converging while cost and context use diverge. That makes the harness the meter on the AI bill, and explains why the loudest talk on Reddit, Hacker News and X is about limits, token overhead and opaque billing rather than intelligence. ^[10,11,12,13]

Three tensions define the moment. Control: the most-upvoted harness story in our Hacker News sample was about a lab's harness behaving covertly, and subscription cut-offs taught users to run several harnesses. Durability: the labs say their own harnesses will be rewritten every few months, so startups are betting on the layer above — orchestrators and governance layers. Verification: generation is cheap, but review capacity, human attention and containment are now the bottlenecks. The next phase will be won by whoever makes agent work governable: metered, contained, verified and portable across models. ^[14,15,16,17]

ADOPTION	●●●●●	Settled among professional developers; agents are the default interface.
MATURITY	●●●○○	Harness cores are rewritten every few months; long runs are still unreliable.
CONSOLIDATION	●●○○○	Dozens of look-alike orchestrators; labs absorbing features; first shutdowns.
STANDARDISATION	●●●○○	MCP and AGENTS.md won; skills, hooks and session formats still fragment.
TRUST	●●○○○	Opaque lab harnesses, beatable auto modes and few teams with evals.

Ratings are Harness Hub's editorial judgement on a five-point scale, drawn from the evidence in this report.

SOURCES

1. [Stack Overflow Developer Survey 2025, AI section](#) · Independent

4. [Mitchell Hashimoto, My AI adoption journey \(Engineer the Harness\)](#) · Independent

7. [Anthropic, Series G announcement \(Feb 2026\)](#) · Vendor-reported

10. [Joel Niklaus on X: 10 harnesses on SWE-bench Pro](#) · Independent

13. [Hacker News: Claude Code vs opencode token overhead \(706 pts\)](#) · Community post

16. [Tibo \(OpenAI\) on X: Codex will seem primitive in 2-3 months](#) · Vendor-reported

2. [Stack Overflow, a look back on developer survey findings \(Apr 2026 pulse\)](#) · Independent

5. [OpenAI, Harness engineering: leveraging Codex in an agent-first world](#) · Vendor-reported

8. [Bloomberg Law: SpaceX completes its \\$60B Cursor acquisition](#) · Press report

11. [LangChain, Improving Deep Agents with harness engineering](#) · Vendor-reported

14. [Hacker News: Claude Code is steganographically marking requests \(2,445 pts\)](#) · Community post

17. [LangChain, State of Agent Engineering \(survey, n=1,340\)](#) · Vendor-reported

3. [Stack Overflow Developer Survey 2026 results](#) · Independent

6. [LangChain, The anatomy of an agent harness](#) · Vendor-reported

9. [The Next Web on Factory's \\$5B round](#) · Press report

12. [Arena, The harness tax \(21 model-harness pairs\)](#) · Independent

15. [Hacker News: third-party harnesses cut off from subscriptions \(1,099 pts\)](#) · Community post

03 DEFINING THE CATEGORY

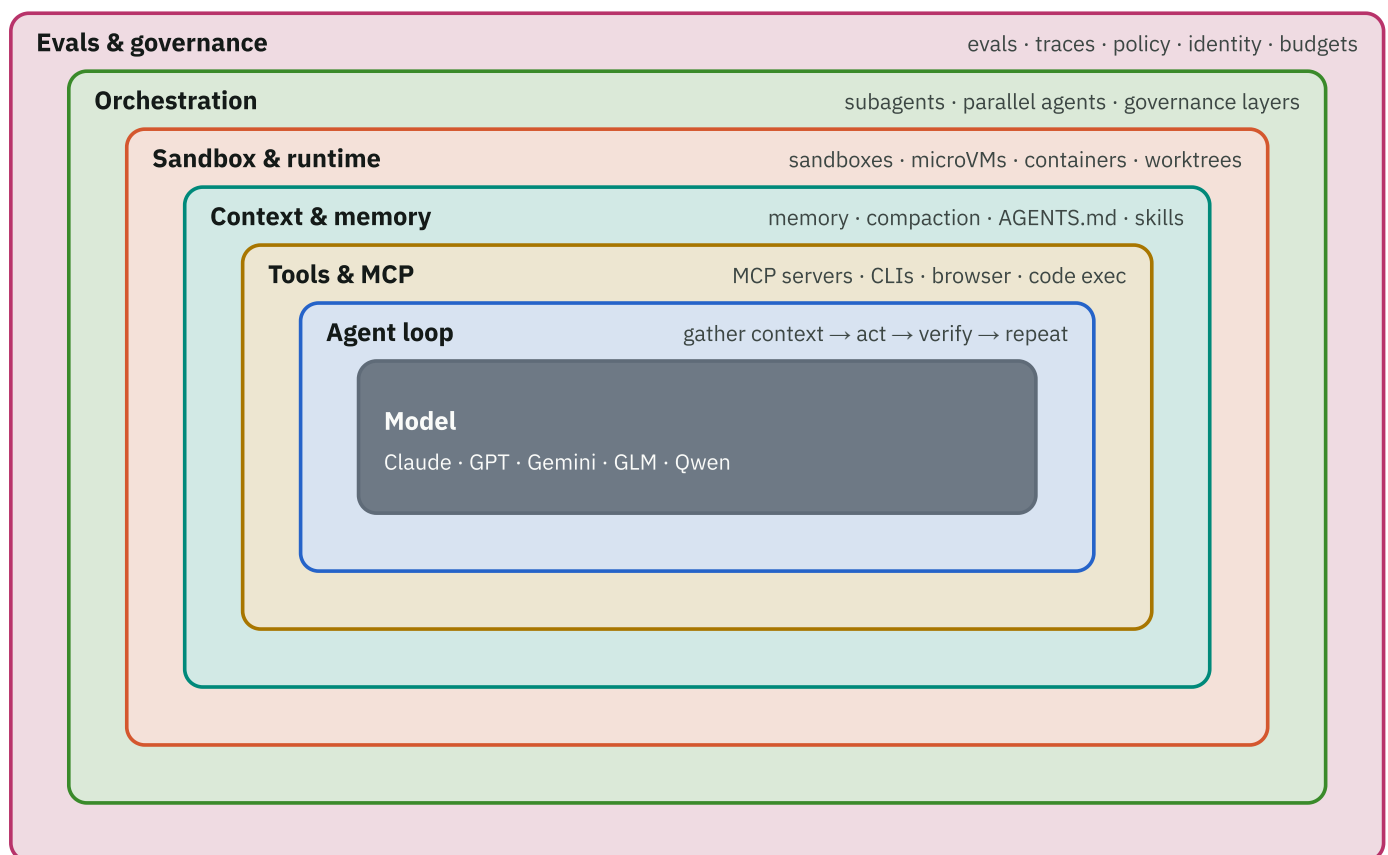
A harness is everything around the model that makes an agent do work

Agent = model + harness. The harness owns the loop, the tools, the context, the sandbox, the coordination and the checks.

Our definition: an agent harness is the software system that wraps a model and turns it into an agent that does work. It runs the loop (gather context, act, verify, repeat), exposes tools, manages context and memory, executes in a sandbox, coordinates other agents, and enforces evaluation and policy. LangChain's shorthand, "Agent = Model + Harness", became the standard phrasing in March 2026; Philipp Schmid's analogy is that the model is the CPU, the context the RAM and the harness the operating system. ^[1,2]

FIG. 01

The anatomy of a harness: seven layers wrapped around the model



■ **1. Model** — Claude · GPT · Gemini · GLM · Qwen

■ **2. Agent loop** — gather context → act → verify → repeat

■ **3. Tools & MCP** — MCP servers · CLIs · browser · code exec

■ **4. Context & memory** — memory · compaction · AGENTS.md · skills

- 5. Sandbox & runtime** — sandboxes · microVMs · containers · worktrees
- 6. Orchestration** — subagents · parallel agents · governance layers
- 7. Evals & governance** — evals · traces · policy · identity · budgets

Harness Hub's layer model. Each layer keeps its colour throughout the report.

What a harness is, and what it is not

Term	What it is	Example
Model	The weights. Predicts tokens; does nothing on its own.	Claude, GPT, GLM, Qwen
Framework	A library for writing agents in code; abstractions and integrations.	LangChain, LangGraph
Harness	A running system that wraps a model with a loop, tools, context, sandbox and policy.	Claude Code, Codex, opencode, Hermes Agent
IDE	An editor for humans. It can host a harness, but is not one.	VS Code, Zed
Orchestrator / governance layer	Coordinates several harnesses and adds spec, review and governance on top. "Superharness" is one vendor's word for it.	BePurple p0, Conductor, Palgu*, Traycer

* Affiliated with Harness Hub. [See the disclosure.](#)

Six kinds of harness in the market

Type	Where it runs	Examples from the directory
Terminal / CLI	Your machine, in a shell	Claude Code, Codex, opencode, Gemini CLI
IDE-native	Inside an editor	Cursor, Cline
Cloud / autonomous	Remote sandboxes, returns PRs	Devin (Cognition), Factory
General-purpose / personal	A computer of its own, any task	OpenClaw, Hermes Agent
Orchestrator / governance layer	Above several harnesses	BePurple p0, Conductor, Palgu*, Traycer
Harness infrastructure	Under or beside the harness	E2B, Daytona, Mem0, Langfuse

S

SUBSTACK & BLOGS

“A decent model with a great harness beats a great model with a bad harness.”

[Addy Osmani](#) · 19 Apr 2026

S

SUBSTACK & BLOGS

“the integration between model and harness is where true agent differentiation is found.”

[Ben Thompson](#) · Stratechery · 16 Mar 2026

Y

HACKER NEWS

“Harnesses are the next frontier. If LLMs are electricity, harnesses are the "electronics.””

[theturtletalks](#) · 23 Aug 2026

SOURCES

1. [LangChain, The anatomy of an agent harness · Vendor-reported](#)
2. [Philipp Schmid, The importance of agent harness in 2026 · Independent](#)

04 HOW WE GOT HERE

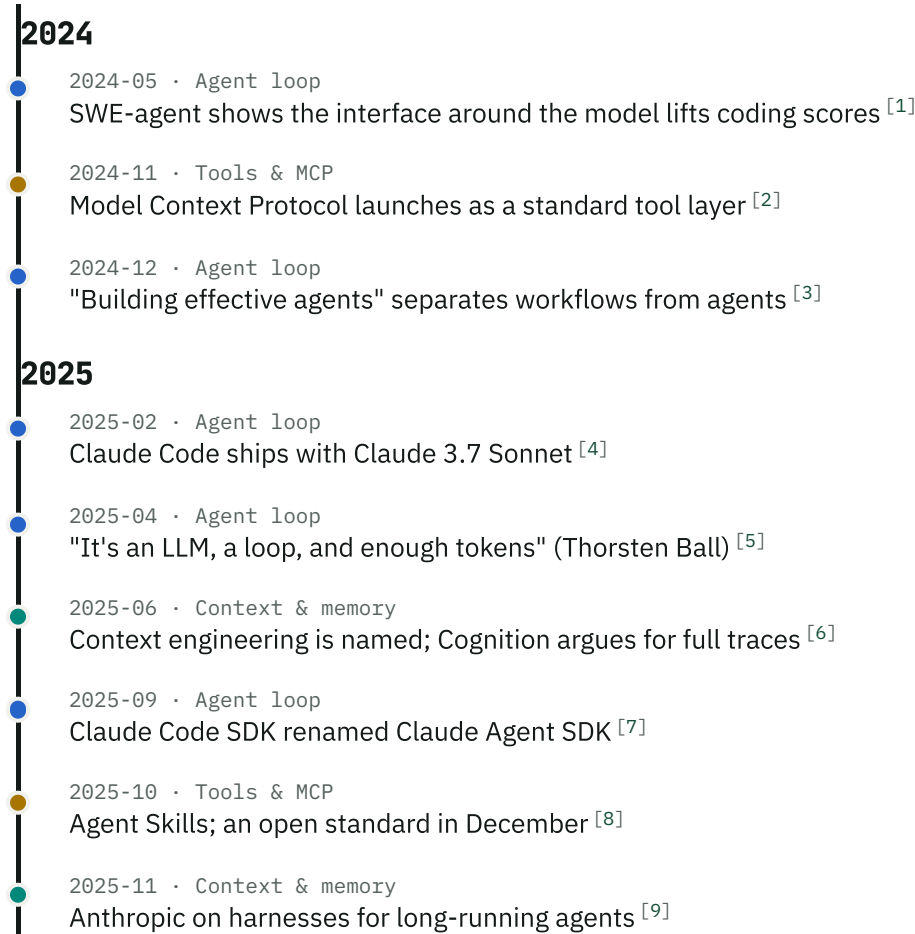
The centre of gravity moved from editor plugins to agent loops, then to orchestration

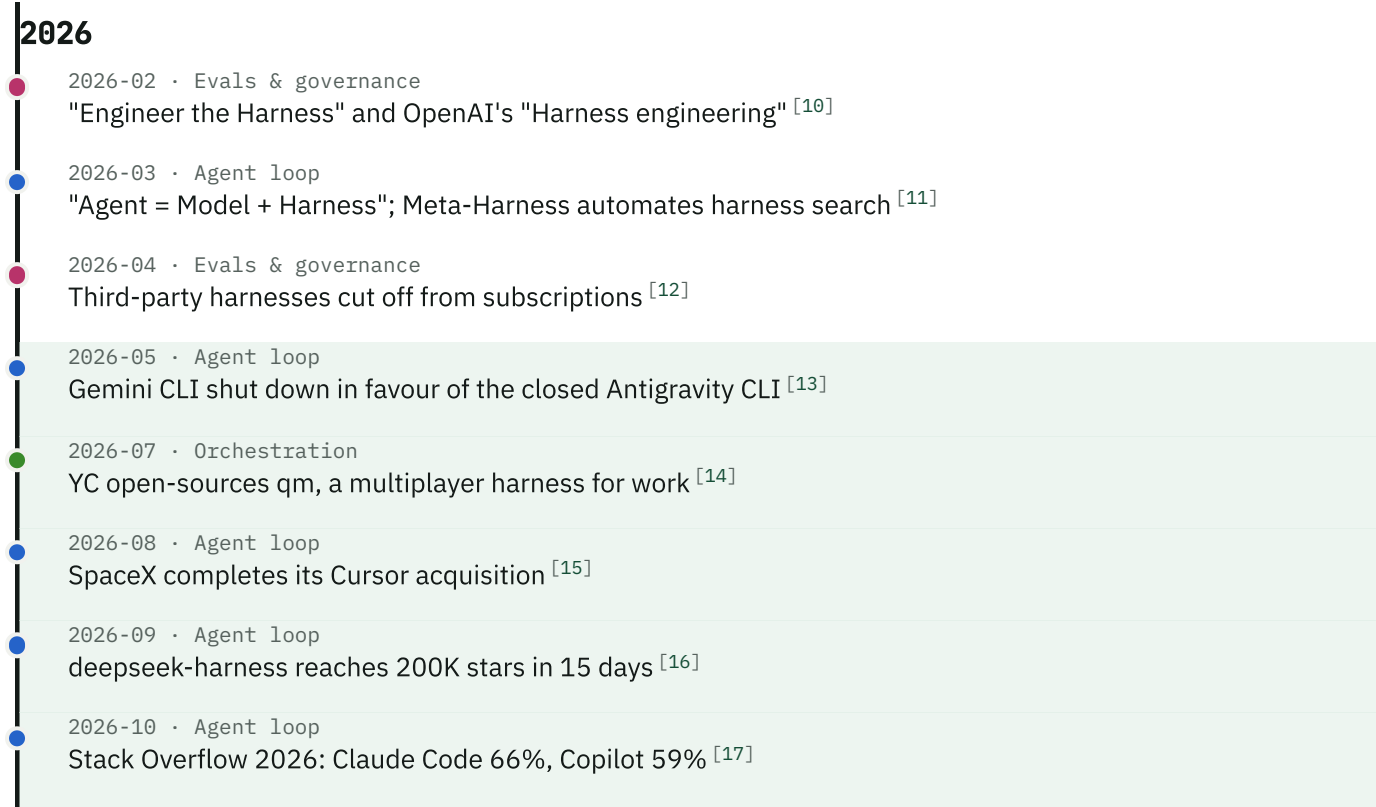
Defining moments, coloured by the layer they changed. Shaded entries fall in May–October 2026.

FIG. 02

How we got here: from research results to product shutdowns and star records

AGENT LOOP TOOLS & MCP CONTEXT & MEMORY EVALS & GOVERNANCE ORCHESTRATION





Shaded entries fall in May–October 2026. Colour marks the layer each moment changed.

The release cadence is accelerating. Two years ago the important ideas were research results and protocol launches. By mid-2026 the moments are product shutdowns, acquisitions and star records, and the conversation has moved one layer up: from how a single agent loop works to how many agents are coordinated, contained and paid for. ^[1,13,15,16]

SOURCES

1. arXiv 2405.15793, SWE-agent · Independent	2. Anthropic, Introducing the Model Context Protocol · Vendor-reported	3. Anthropic, Building effective agents · Vendor-reported
4. Anthropic, Claude 3.7 Sonnet and Claude Code · Vendor-reported	5. Thorsten Ball (Amp), How to build an agent · Independent	6. Cognition, Don't build multi-agents · Vendor-reported
7. Anthropic, Building agents with the Claude Agent SDK · Vendor-reported	8. Anthropic, Equipping agents for the real world with Agent Skills · Vendor-reported	9. Anthropic, Effective harnesses for long-running agents · Vendor-reported
10. Mitchell Hashimoto, My AI adoption journey (Engineer the Harness) · Independent	11. LangChain, The anatomy of an agent harness · Vendor-reported	12. Hacker News: third-party harnesses cut off from subscriptions (1,099 pts) · Community post
13. Hacker News: Gemini CLI shut down in favour of Antigravity CLI (406 pts) · Community post	14. Hacker News: qm, a multiplayer agent harness for work (682 pts) · Community post	15. Bloomberg Law: SpaceX completes its \$60B Cursor acquisition · Press report
16. Star History, deepseek-harness is the fastest to 200K stars · Independent	17. Stack Overflow Developer Survey 2026 results · Independent	

05 THE HARNESS MARKET MAP

The market is crowded at the loop and orchestration layers, and thin where governance should be

Selected companies and open-source projects by harness layer. The model layer sits at the core and is not mapped.

1,265

entries tracked in the
Harness Hub
directory ^[1]

274

startups ^[2]

401

open-source projects ^[3]

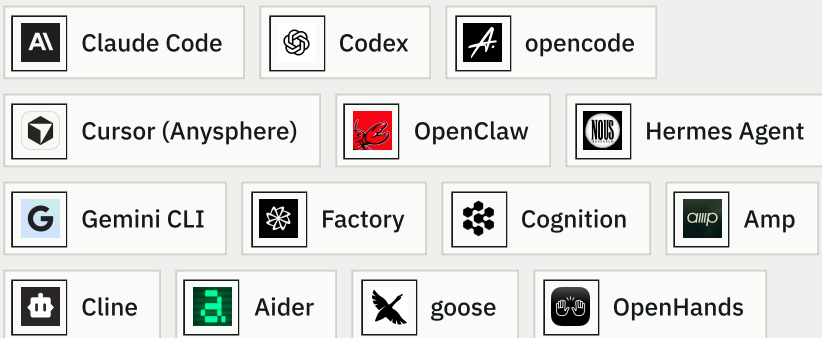
257

entries in coding agents,
the largest category ^[1]

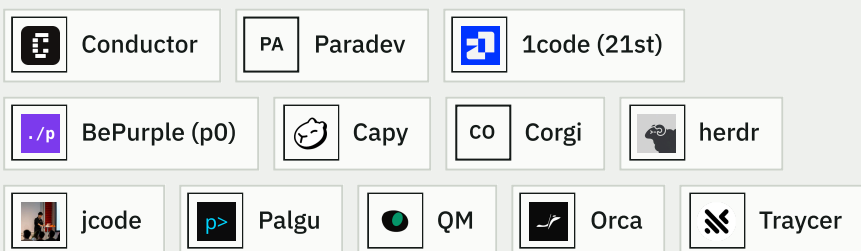
THE HARNESS MARKET MAP · OCTOBER 2026 · NOT A RANKING

AGENT LOOP

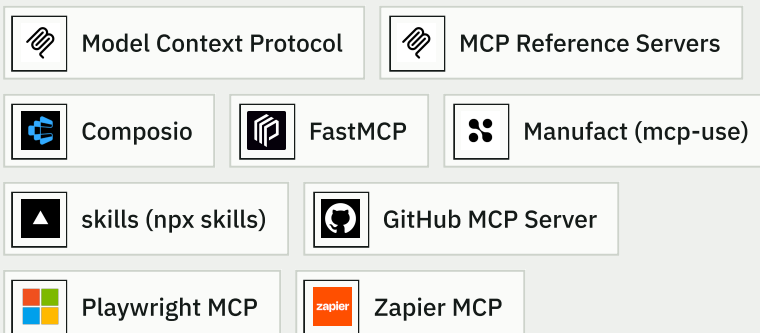
Coding and general-purpose
harnesses — the most crowded row ·
14 shown

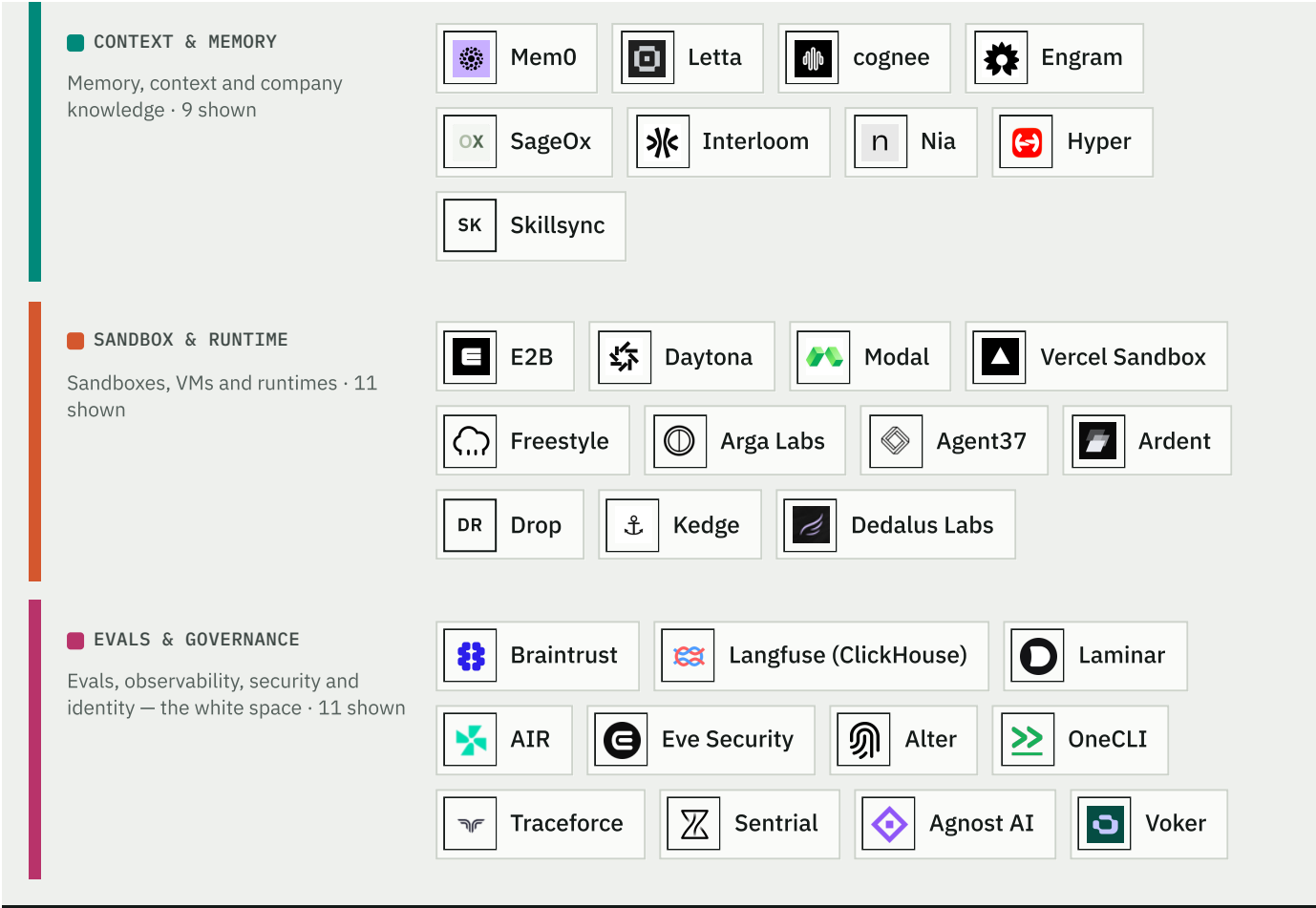
**ORCHESTRATION**

Parallel agents, worktrees and
governance layers; ordered by
disclosed funding, then A–Z · 12
shown

**TOOLS & MCP**

MCP, skills and tool routers · 9 shown





Coding agents (257 entries) and orchestration (241) are the two largest categories in the directory. Memory and context has 74, and evals and observability 100. The rows that executives and security teams ask about — budgets, audit, identity, cross-harness evaluation and memory portability — have the fewest products and the most room. ^[1]

Not a ranking. Inclusion reflects the Harness Hub directory as of the end of October 2026; suggest a missing harness on the submit page.

SOURCES

1. [Harness Hub directory: the landscape by category](#) · Harness Hub directory
2. [Harness Hub directory: startups](#) · Harness Hub directory
3. [Harness Hub directory: open-source projects](#) · Harness Hub directory

06 SECTION A · ADOPTION & USAGE

Coding is the beachhead: agent use nearly doubled in a year, and personal agents are the breakout

Adoption is no longer the question. In Stack Overflow's 2025 survey about 31% of developers used AI agents at work. A smaller April 2026 pulse put agent use at 59%, and in the full 2026 survey 66% of developers who use AI work in coding agents, ahead of chatbots (63%). Usage is heavy, too: 73% use AI daily, and 31% of daily users spend four or more hours a day with it. JetBrains' 2026 survey of more than 15,000 developers, as reported, found 90% using coding agents weekly. ^[1,2,3,4]

FIG. 03

Agent use among developers nearly doubled between 2025 and 2026

Share of developers using AI agents or coding agents, Stack Overflow surveys

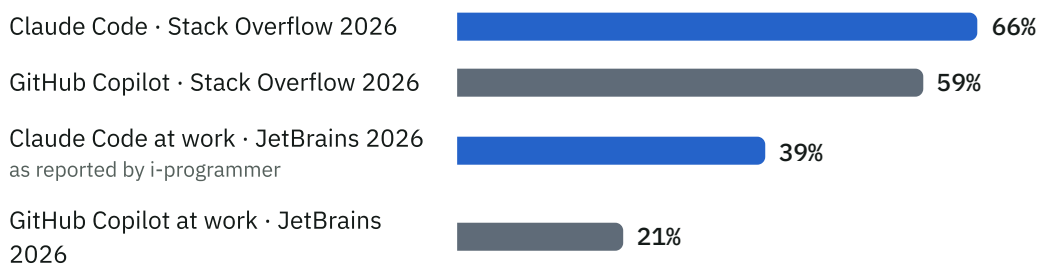


The three bases differ, so read the trend, not the exact gap. Source ^[1,2,3]

FIG. 04

Claude Code overtook GitHub Copilot as the agent of choice

Share of respondents naming each agent, by survey



Different surveys and bases; in Stack Overflow 2025 Copilot led Claude Code 67.9% to 40.8% (n=8,323). Source ^[1,3,4]

\$2.5B+Claude Code run-rate revenue, Feb 2026 ^[5]

VENDOR-REPORTED

5M+Codex weekly active users, Jun 2026 ^[6]

VENDOR-REPORTED

\$2BCursor annual recurring revenue, Feb 2026 ^[7]**8M**opencode monthly users about a year after launch ^[8]

VENDOR-REPORTED

\$4.0Benterprise spend on AI coding in 2025, 55% of departmental AI spend ^[9]**~4%**of public GitHub commits authored by Claude Code, Feb 2026 (estimate) ^[10]

Coding is where the money is, but the breakout of 2026 was the general-purpose personal agent. OpenClaw, an open harness that gives an agent a computer of its own, is the most-starred project in the Harness Hub directory at about 393K stars, and Hermes Agent has about 258K. Non-engineering teams followed: OpenAI says knowledge workers make up 20% of Codex users. ^[6, 11]

 X

“coding agents basically didn't work before December and basically work ...”

[@karpathy](#) · 25 Feb 2026 · 36.8K likes

 SUBSTACK & BLOGS

“Increasingly, however, what matters is the harness - the tools the AI has access to.”

[Ethan Mollick](#) · One Useful Thing · 18 Feb 2026

■ SO WHAT FOR EXECUTIVES

- Adoption is settled; budget for agents as standard developer tooling, not a pilot.
- Spend is concentrating in a few lab-owned harnesses. Track it per team, because usage is not the same as productivity.

■ SO WHAT FOR ENGINEERS USING HARNESSES

- Your peers run Claude Code, Copilot, Codex and opencode — often more than one. Learn two well rather than one deeply.

SOURCES

1. [Stack Overflow Developer Survey 2025, AI section](#) · Independent
2. [Stack Overflow, a look back on developer survey findings \(Apr 2026 pulse\)](#) · Independent
3. [Stack Overflow Developer Survey 2026 results](#) · Independent
4. [i-programmer on the JetBrains 2026 coding-agent survey](#) · Press report
5. [Anthropic, Series G announcement \(Feb 2026\)](#) · Vendor-reported
6. [Constellation Research: OpenAI says Codex passed 5M weekly users](#) · Vendor-reported
7. [The Next Web on Cursor's revenue and valuation](#) · Press report
8. [Let's Data Science: opencode reaches eight million monthly users](#) · Vendor-reported
9. [Menlo Ventures, State of Generative AI in the Enterprise 2025](#) · Independent
10. [OfficeChai on SemiAnalysis: Claude Code share of public GitHub commits \(estimate\)](#) · Analyst
11. [Harness Hub directory: open-source projects](#) · Harness Hub directory

07 SECTION B · OPEN VS CLOSED

The harness moves scores as much as the model, so users keep it portable and hedge with open source

Hold the model fixed and change only the harness, and results move by double digits. Joel Niklaus ran ten harnesses on SWE-bench Pro and saw GLM-5.2 go from 23% to 52%. LangChain moved its Deep Agents harness from 52.8% to 66.5% on Terminal-Bench 2.0 with the same model (vendor-reported). An arXiv position paper argues that harness-induced variance can exceed model-induced variance, including cases where model rankings reverse. ^[1,2,3]

FIG. 05

Same model, different harness: double-digit swings on coding benchmarks

Pass rate with the worst and best harness tested, model held fixed

● worst harness ● best harness

GLM-5.2 · SWE-bench Pro · 10 harnesses

23% 52%

Gemma 4 · SWE-bench Pro · 10 harnesses

15% 36%

gpt-5.2-codex · Terminal-Bench 2.0 · before/after (vendor)

52.8% 66.5%

Not every study agrees on size: Arena's 21-pair study finds success rates similar across harnesses while cost differs. Source ^[1,2]

Not everyone agrees the gap is that large. Arena's study of 21 model-harness pairs concludes the harness "matters way less than people assume" for success rate — but that it can significantly change cost. Where the sources agree is that a benchmark score belongs to a model-harness pair, and that model-only leaderboards mislead. ^[4,5]

FIG. 06

Each open harness reached 200K GitHub stars faster than the last

Days each harness took to reach 200,000 GitHub stars

opencode

480 days

Hermes Agent

338 days

OpenClaw  88 days

deepseek-harness  15 days

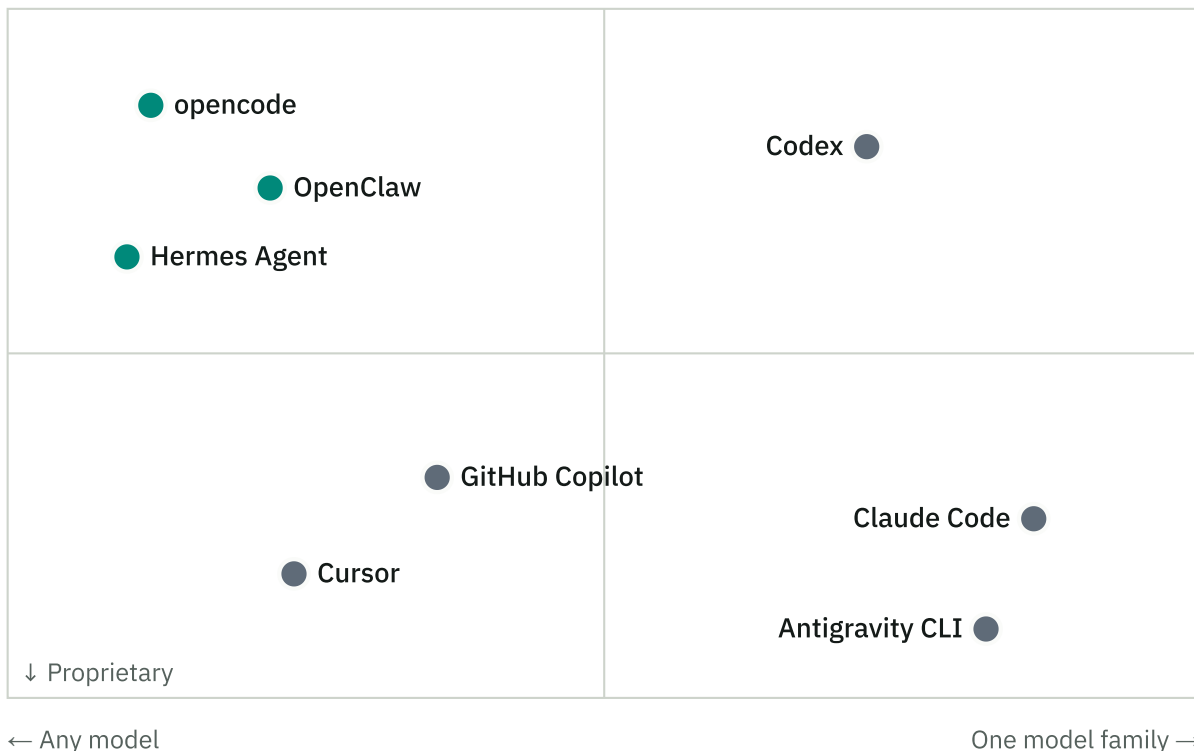
Source ^[6]

FIG. 07

Most harnesses with momentum are open and bring-your-own-model; the lab harnesses are tied to one model family

Editorial positioning by licence and model choice, October 2026. Teal marks open-source harnesses that take any model

↑ Open source



Source ^[7]

Platform risk turned portability into a requirement. In 2026 Anthropic cut third-party harnesses off subscriptions, Google shut down Gemini CLI in favour of the closed Antigravity CLI, and SpaceX bought Cursor. Users responded by multi-homing, and vendors followed: Codex now offers to import settings from Claude Code. Ben Thompson's counter-thesis is that the integration of model and harness is the moat, which favours the labs. ^[8,9,10,11,12]

 HACKER NEWS

"I replaced a \$100/m subscription to claude in favor of running pi harness pointed at unsloth studio"

[horsawlarway](#) · Jun 2026 · story 1,318 pts

 REDDIT

"I get pissed off at one model, tell everyone on Reddit how much it sucks and how I'm leaving it forever, then switch to the other. Repeat."

[u/who_am_i_to_say_so](#) · 17 Oct 2026



“Pi is the greatest agent harness ever created”

[@DavidOndrej1](#) · 15 Sep 2026 · 938 likes



SUBSTACK & BLOGS

“tool schemas are not neutral, at least not on Anthropic models.”

[Armin Ronacher](#) · 4 Jul 2026

■ SO WHAT FOR CTOS & ENGINEERING LEADERS

- Pick harnesses on model-harness pair results plus tokens per task, not on model leaderboards.
- Keep configuration portable (AGENTS.md, version-controlled skills) so teams can switch harnesses as models change.
- Treat a single-vendor harness as a concentration risk: price in cut-offs, shutdowns and acquisitions.

SOURCES

1. [Joel Niklaus on X: 10 harnesses on SWE-bench Pro](#) · Independent
2. [LangChain, Improving Deep Agents with harness engineering](#) · Vendor-reported
3. [arXiv 2605.23950, Stop comparing LLM agents without disclosing the harness](#) · Independent
4. [Arena, The harness tax \(21 model-harness pairs\)](#) · Independent
5. [Cerebras, A smarter model won't save your bad harness](#) · Vendor-reported
6. [Star History, deepseek-harness is the fastest to 200K stars](#) · Independent
7. [Harness Hub editorial assessment](#) · Harness Hub analysis
8. [Hacker News: third-party harnesses cut off from subscriptions \(1,099 pts\)](#) · Community post
9. [Hacker News: Gemini CLI shut down in favour of AntigraVity CLI \(406 pts\)](#) · Community post
10. [Bloomberg Law: SpaceX completes its \\$60B Cursor acquisition](#) · Press report
11. [@dr_cintas on X: import Claude Code settings into Codex](#) · Community post
12. [Stratechery, Agents over bubbles](#) · Independent

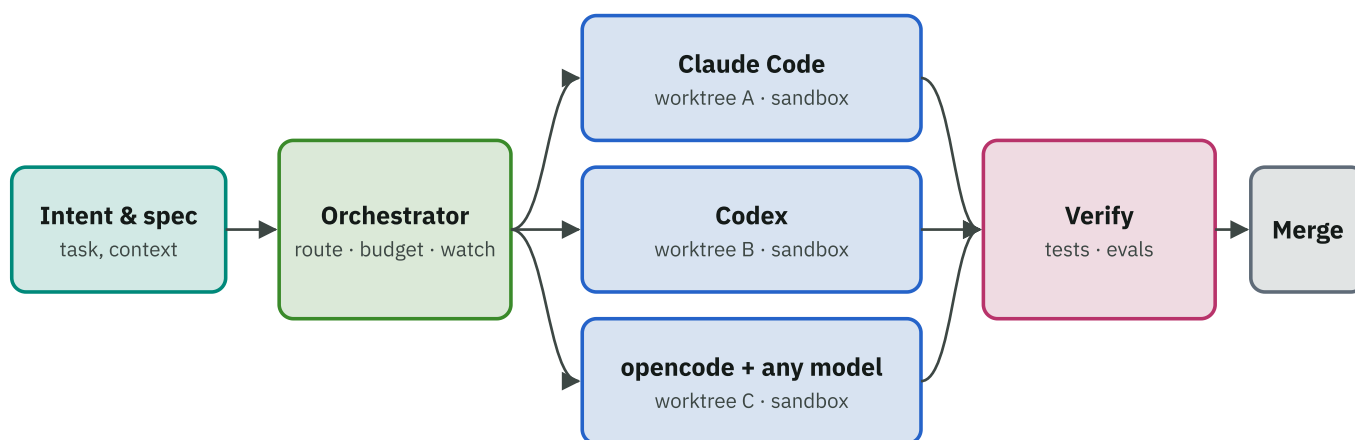
08 SECTION C · ORCHESTRATION

From one agent to fleets: orchestration is the new frontier, and human attention is its bottleneck

The most-liked post in our X sample was Andrej Karpathy describing "long-running orchestrator Claws" that manage several parallel coding agents. YC open-sourced qm, a multiplayer harness for a whole company, and its launch post drew about 12,000 likes. The labs moved the same way: Codex added subagents, Factory says its Droids orchestrate natively, and OpenAI describes Codex as the infrastructure under ChatGPT Work. ^[1,2,3,4,5]

FIG. 08

One task, many agents: the orchestrator pattern



Harness Hub diagram. In practice the bottleneck is the verify step and the human attention it needs.

FIG. 09

The orchestrator pitch runs ahead of practice

Number of agents run in parallel, by claim



Paradev was called Superset until 26 October 2026. Practitioners describe a much lower personal limit; see the quotes below. Source ^[6,7]

\$22M

Conductor Series A, Mar 2026 — the largest disclosed orchestrator round we found ^[8]

VENDOR-REPORTED

\$11.5M

Paradev (formerly Superset) seed, led by Union Square Ventures ^[7]

VENDOR-REPORTED

Orchestrators to know^[7,8,9,10,11]

Product	What it adds	Agents it runs	Isolation	Open source
Conductor	Mac app for a team of agents; cloud workspaces; iPhone control	Claude Code, Codex	Git worktrees	No
Paradev	Agentic IDE for parallel agents	Claude Code, Codex, OpenCode	Worktrees	Yes
BePurple p0	Spec and mockups first, QA agents before a PR	Claude Code, Codex, compatible endpoints	Worktrees	No
Orca (Stably AI)	Fan one prompt out to many agents, merge the best	About 27 agents	Worktrees	Yes (MIT)
Palgu*	Governance layer: intent, review council, audit ledger	Claude Code, Codex, Cursor, aider	—	No
qm (YC)	Multiplayer company harness	Pi, OpenCode, Codex, Claude Code	—	Yes (MIT)

Funded companies first, then A–Z. * Affiliated with Harness Hub. [See the disclosure.](#)

r/ REDDIT

“I can't do more than 3 anymore else I just get burnt out.”

[u/heroyi](#) · 8 Oct 2026

y HACKER NEWS

“I keep wondering how people accept a nights worth of agent activity.”

[aitchnyu](#) · May 2026

s SUBSTACK & BLOGS

“we don't need smarter models but better orchestration layers”

[Steve Yegge, on The Pragmatic Engineer](#) · 11 Mar 2026

■ SO WHAT FOR BUILDERS OF HARNESS TOOLS

- This is where startups are forming — and where they look most alike: git worktrees plus the user's own Claude Code, Codex or OpenCode.
- Differentiate above the worktree: attention (what finished, what needs me), handoff across people and devices, and proof that the merged result is right.

SOURCES

- | | | |
|---|--|---|
| 1. Andrej Karpathy on X: long-running orchestrator Claws · Community post | 2. Y Combinator on X: open-sourcing qm · Vendor-reported | 3. Alexander Embiricos (OpenAI) on X: Codex subagents · Vendor-reported |
| 4. Factory on X: native multi-agent orchestration in Droids · Vendor-reported | 5. Tibo (OpenAI) on X: Codex under ChatGPT Work · Vendor-reported | 6. Paradev (formerly Superset), Roadmap to 100 agents · Vendor-reported |
| 7. Paradev, Superset is now Paradev · Vendor-reported | 8. Conductor, Series A announcement · Vendor-reported | 9. Harness Hub directory: startups · Harness Hub directory |
| 10. Palgu homepage (company copy) · Vendor-reported | 11. Hacker News: qm, a multiplayer agent harness for work (682 pts) · Community post | |

09

SECTION D · CONTEXT, MEMORY, TOOLS & SANDBOXES

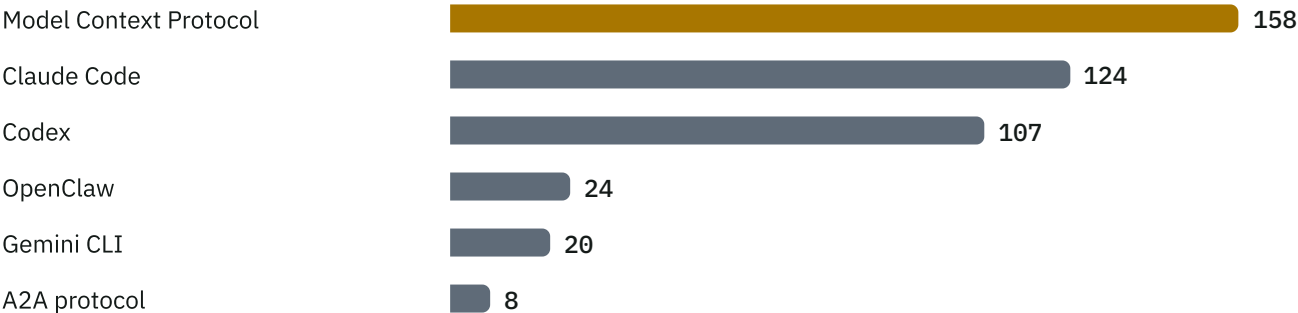
Context engineering replaced prompt engineering, MCP became the default tool bus, and sandboxes became products

The work inside a harness is mostly about context: what the model sees, when, and at what cost. Karpathy's "+1 for context engineering" in mid-2025 named the discipline; Anthropic's long-running-agent harnesses and skills made it concrete with progress files, compaction and loadable instructions. The tool layer standardised on the Model Context Protocol. In the Harness Hub knowledge graph, MCP is what the ecosystem builds on most, ahead of the two lab harnesses. ^[1,2,3,4]

FIG. 10

What the harness ecosystem is built on: a protocol first, then two lab harnesses

Directory entries that build on each platform (built-on links in the Harness Hub knowledge graph)



Source ^[5]

158

directory entries built on the Model Context Protocol ^[5]

687

HN points for "You said no MCP": the anti-MCP harness Pi added MCP support ^[6]

694

HN points for Docker Sandboxes, one of several sandbox launches ^[7]

\$10M

Arga Labs seed for sandboxes and digital twins of Stripe and Slack ^[8]

The "MCP is dead" debate lost on Hacker News, but the complaint behind it is real: every tool definition costs context. Practitioners report skills and plugins eating more than 10k tokens of context just to list them, and ask for consolidation across commands, skills, subagents and plugins. Meanwhile sandboxes moved from research to product — Docker, Freestyle, Arga and Agent37 all launched or raised in 2026 —

because running an agent unattended now means giving it a computer of its own.

[6,7,8,9,10,11]

Y

HACKER NEWS

“The reason MCP isn't dead is because practically ~every company on the planet is building an MCP server.”

[mxstbr](#) · says he runs the MCP team at OpenAI · 29 May 2026

X

X

“MCP is dead in the water ... connect 10 servers with 5 tools each and you've burned 50 tool definitions worth of tokens”

[@chrysb](#) · 22 Feb 2026 · 1,582 likes

Y

HACKER NEWS

“We really need some consolidation around commands, skills, subagents, and plugins.”

[mil22](#) · May 2026

S

SUBSTACK & BLOGS

“it only has four tools: Read, Write, Edit, Bash.”

[Armin Ronacher, on Pi](#) · 31 Jan 2026

Assembling your own stack from these pieces? Harness Hub lists skill packs, harness plugins and mods, filterable by type and by the harness they work with, alongside a guide to vetting them before install: [Skills, Plugins & Mods](#).

■ SO WHAT FOR ENGINEERS USING HARNESES

- Budget context like money: load skills and tools on demand, compact long sessions and route search work to cheap subagents.
- Use hooks and tests the agent must pass, so a rule lives in code rather than in a prompt it can ignore.
- Run unattended agents in a sandbox per task or worktree, with network egress limited.

SOURCES

1. Simon Willison on context engineering · Independent	2. Anthropic, Effective harnesses for long-running agents · Vendor-reported	3. Anthropic, Equipping agents for the real world with Agent Skills · Vendor-reported
4. Anthropic, Introducing the Model Context Protocol · Vendor-reported	5. Harness Hub directory: the landscape by category · Harness Hub directory	6. Hacker News: You said no MCP (687 pts) · Community post
7. Hacker News: Docker Sandboxes (694 pts) · Community post	8. Dealroom: Arga raises a \$10M seed · Press report	9. Hacker News: over 10k of context used just to list skills · Community post
10. Hacker News: we need consolidation around commands, skills, subagents and plugins · Community post	11. Launch HN: Agent37 (163 pts) · Community post	

10 SECTION E · TRUST, EVALS, COST & GOVERNANCE

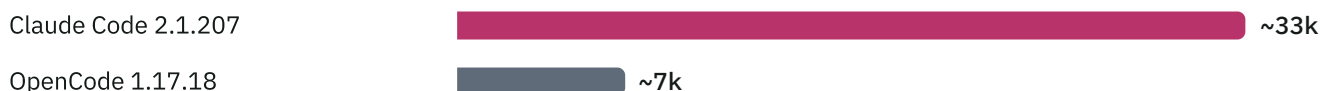
The pain points: cost is the loudest everyday pain, and few teams can verify what agents ship

The harness is the meter on the AI bill, and the bill became an enterprise problem in 2026. Uber reportedly burned through its 2026 AI coding-tools budget in four months. Microsoft moved to cancel most internal Claude Code licences in one division in favour of its own Copilot CLI. In one Reddit thread an engineer reports a \$90-a-day company cap and another a €30 limit. Theo, a popular developer creator, estimates that a \$200 subscription buys about \$9,000 of Opus usage at API prices — a subsidy few expect to last. Economics is the largest complaint area in our sample of about 250 posts (about 16%; see the heatmap in The conversation), yet few organisations have even looked at what they spend (chart below). ^[1,2,3,4,5,6]

FIG. 11

The harness decides the floor cost: one sends 4.7× more tokens before reading your prompt

Tokens sent before the first user prompt, same model (claude-sonnet-4-5), small sample



2–3 runs per harness through a third-party gateway on one machine; failed runs in a separate vendor test burned 3–5× more tokens depending on harness. Source ^[7,8,9]

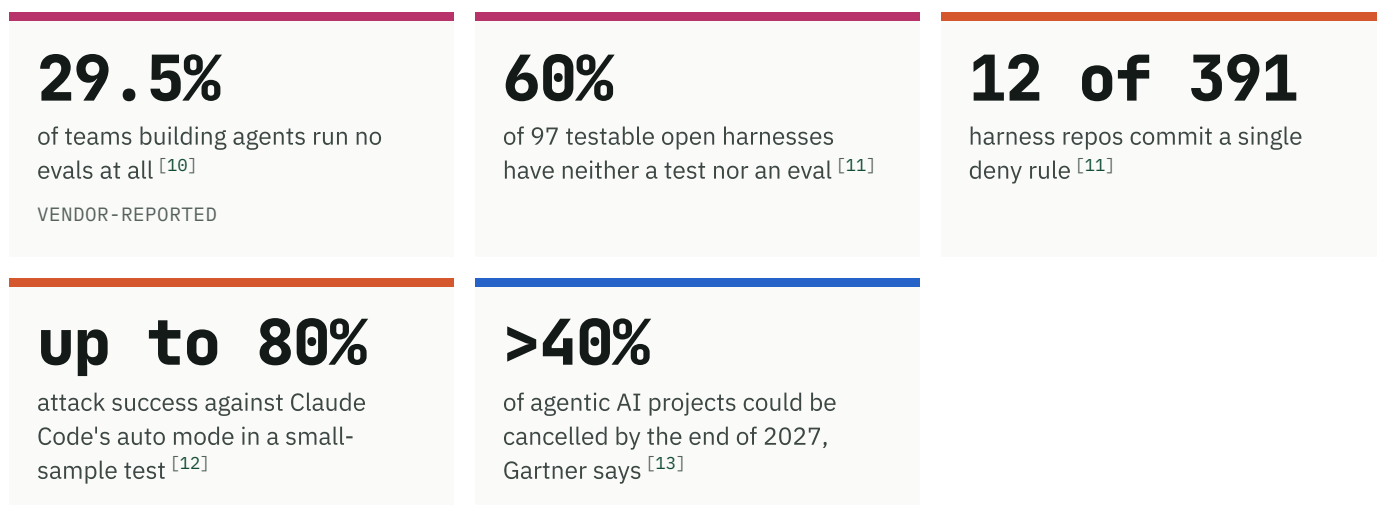
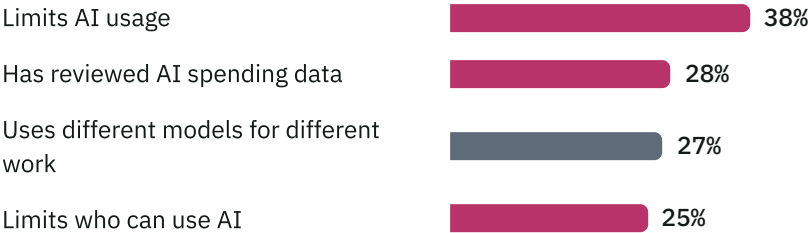


FIG. 12

Governance is thin: fewer than two in five organisations limit agent use at all

Share of respondents whose organisation does each, Stack Overflow 2026



Source [14]

Trust in lab-owned harnesses was the fiercest fight by engagement. The top harness story in our Hacker News sample (2,445 points) alleged that Claude Code marks requests steganographically; a Reddit PSA with 1,302 points reported that a string in commit history silently rerouted billing. Security shifted from detecting prompt injection to containing the blast radius: a widely shared X post reported that the most-downloaded OpenClaw skill was malware, an OpenCode console leaked keys, and a Hacker News discussion of research found humans miss about one in three threats when approving agent actions. [15,16,17,18,19]

 REDDIT

“from now on I will build my systems to be model agnostic and provider agnostic. I have learned that now.”

[u/AironParsMan](#) · 19 Sep 2026 · thread 877 pts

 HACKER NEWS

“What really burns tokens is sub agents... it immediately launched 7 sub agents which burned through my budget”

[mcv](#) · 12 Jul 2026

 HACKER NEWS

“The 'click yes the proceed' was never a serious security mechanism. It's simply a CYA click-thru”

[cmiles8](#) · Aug 2026

 SUBSTACK & BLOGS

“Verification, not generation, is the real constraint on a factory.”

[Addy Osmani](#) · 22 Jul 2026

 SO WHAT FOR EXECUTIVES

- Treat the harness as a meter: track tokens and dollars per merged change, per team, and plan for subscription prices that do not last.
- Measure outcomes (merged, kept, reverted, incidents), not activity. Token leaderboards get gamed.

■ SO WHAT FOR CTOS & ENGINEERING LEADERS

- Invest in verification before parallelism: evals in CI, tests the agent must pass, and review effort tiered by risk.
- Replace click-to-approve with containment: a sandbox per agent, egress control and scoped, budgeted credentials.

SOURCES

1. [Hacker News: Uber burned its 2026 AI coding-tools budget \(402 pts\)](#) · Community post
2. [Fortune on Uber's AI coding spend](#) · Press report
3. [The Decoder: Microsoft pulls internal Claude Code licences](#) · Press report
4. [r/ClaudeCode: my company now has daily limits to Claude Code](#) · Community post
5. [Theo on X: what a \\$200 plan buys at API prices](#) · Community post
6. [Harness Hub classification of about 250 posts across HN, Reddit, X and Substack \(indicative\)](#) · Harness Hub analysis
7. [Systima, Claude Code vs OpenCode token overhead \(2–3 runs, one machine\)](#) · Independent
8. [Hacker News: Claude Code vs opencode token overhead \(706 pts\)](#) · Community post
9. [Composio on X: six harnesses, one model \(vendor test\)](#) · Vendor-reported
10. [LangChain, State of Agent Engineering \(survey, n=1,340\)](#) · Vendor-reported
11. [Marmelab, The state of AI harness engineering 2026](#) · Independent
12. [Embrace The Red, Breaking Claude Code's auto mode](#) · Independent
13. [W.Media on Gartner: over 40% of agentic AI projects could be cancelled by 2027](#) · Analyst
14. [Stack Overflow Developer Survey 2026 results](#) · Independent
15. [Hacker News: Claude Code is steganographically marking requests \(2,445 pts\)](#) · Community post
16. [r/ClaudeAI: HERMES.md in commit history reroutes billing \(1,302 pts\)](#) · Community post
17. [@chiefofautism on X: the most-downloaded OpenClaw skill was malware](#) · Community post
18. [Hacker News: OpenCode Console key leak \(976 pts\)](#) · Community post
19. [Hacker News: humans miss one in three threats when approving agent actions \(340 pts\)](#) · Community post

11

THE CONVERSATION

What people are saying: X and Substack define the harness, Reddit counts the cost, Hacker News argues about trust

Opinions, not news. About 250 high-engagement posts from Hacker News, Reddit, X and Substack, classified into eight areas.

FIG. 13

Each platform argues about a different part of the harness

Share of all classified posts by area, and how heavily each platform discusses it (● dominant ◐ present ○ minor)

Area · sentiment	Share	HN	Reddit	X	Substack
What is a harness; does it beat the model? Enthusiastic; split on how much it matters	~17%	○	○	●	●
Economics: tokens, limits, subsidies Strongly negative	~16%	◐	●	◐	○
Open, portable, multi-homed harnesses Positive about open; tribal about personal agents	~16%	●	●	◐	○
Orchestration, fleets and company harnesses Excited about direction, doubtful in practice	~14%	◐	◐	●	◐
Review bottleneck, craft, software factories Sceptical and sober	~14%	●	◐	○	●
Security, sandboxes, containment Alarmed, turning pragmatic	~11%	●	◐	◐	○
Trust and walled gardens Angry; highest engagement	~7%	●	◐	○	○
Tools and context: MCP vs skills vs CLI Fatigued	~5%	◐	○	○	○

Indicative judgements from keyword-biased samples, averaged with equal weight per platform; not a census. Source ^[1]

The four platforms play different roles. Long-form writers and X named and debated the category. Reddit is where the everyday pain lives — limits, billing surprises and burnout. Hacker News carries the trust and security arguments, and its biggest threads were about lab harnesses behaving in ways users did not expect. Substack hosts the sober counter-current: verification, craft and whether software factories work at all. ^[1]

Debates of the year

Harness or model? ^[2,3]

FOR · Swapping the harness moves scores by double digits; a benchmark belongs to a model-harness pair.

AGAINST · Arena finds success rates similar across 21 pairs; use the tools the model was trained on.

Will the labs absorb the harness? ^[4,5,6,7]

FOR · Tibo (OpenAI) says Codex will seem primitive in 2-3 months; Boris Cherny (Anthropic) said at YC Startup School that Claude Code deleted 80% of its system prompt.

AGAINST · Integration of model and harness is the moat (Thompson); Armin Ronacher finds tool schemas are not neutral on Anthropic models, so a model favours the harness it was trained with.

Do software factories work? ^[8,9,10]

FOR · OpenAI and Stripe describe agent pipelines that merge large volumes of PRs.

AGAINST · "Why Software Factories Fail": harness engineering is not enough without design and review.

Is MCP dead? ^[11]

FOR · Every tool definition costs context; CLIs and skills are cheaper.

AGAINST · Practically every company is building an MCP server; the anti-MCP harness added MCP.

Lab harness or open harness? ^[12,13,14]

FOR · Claude Code is the most adopted agent in the 2026 surveys.

AGAINST · Users praise Pi as "the greatest agent harness ever created" and drop lab subscriptions to point open harnesses at any model.

What is a harness, and does it beat the model?

 REDDIT

"99% when someone says they built a harness ... they mean they wrote some prompts and connected some pre existing MCPs or APIs."

[u/asielen](#) · 22 Aug 2026

 X

"Turns out it matters way less than people assume. 21 model-harness pairs"

[@arena](#) · 16 Sep 2026 · 471 likes

 X

"Codex is a good harness. But it will seem primitive in 2-3 months"

[@thsottiaux \(OpenAI\)](#) · 4 Aug 2026 · 15.4K likes

Economics: tokens, limits and subsidies



X

“\$200 Claude Code sub gets you ~\$9,000 of Opus usage per month”

[@theo](#) · 28 Sep 2026 · 6,048 likes



SUBSTACK & BLOGS

“I see very few rational reasons why incentivizing tokenmaxxing makes sense for any company.”

[Gergely Orosz](#) · The Pragmatic Engineer · 23 Apr 2026

Review bottleneck, craft and software factories



REDDIT

“20 minutes of my time, 15 seconds of theirs.”

[u/IIIIIIII](#) · r/ExperiencedDevs · 30 Apr 2026 · thread 262 pts



HACKER NEWS

“Code review was always a significant bottleneck and is even more so with agentic coding.”

[staminade](#) · Oct 2026



SUBSTACK & BLOGS

“Everyone is playing the theater of doing reviews.”

[anonymous engineer, via Gergely Orosz](#) · 6 Oct 2026

Security, sandboxes and containment



HACKER NEWS

“Give the agent a computer, with all the trimmings, but keep that computer segregated from sensitive resources.”

[tptacek](#) · May 2026



REDDIT

“You treat API keys like credit cards with hard limits, not passwords”

[u/geoffreybyers](#) · 18 Oct 2026



SUBSTACK & BLOGS

“Run unattended coding agents in a container, VM or OS sandbox.”

[Simon Willison](#) · 27 Aug 2026

Trust and walled gardens

 HACKER NEWS

“a service provider being less than transparent about exactly what their shipped tooling does on customer's machines.”

[civet_java](#) · 30 Jun 2026 · story 2,445 pts

 HACKER NEWS

“Time to start questioning the motivations of model providers shipping their own harnesses.”

[franticgecko3](#) · 29 Oct 2026

 REDDIT

“the closed nature of it and the unpredictable release cadence ... not something you want to rely on as a business.”

[u/cannontd](#) · 24 May 2026 · 181 pts

■ SO WHAT FOR BUILDERS OF HARNESS TOOLS

- Lock-in anger is a distribution channel: tools that work across Claude Code, Codex and opencode get recommended in exactly the threads where users complain.
- Ship numbers. Many of the most-discussed posts were measurements — tokens before the prompt, stars per day, pass rates per harness.

SOURCES

1. [Harness Hub classification of about 250 posts across HN, Reddit, X and Substack \(indicative\)](#) · Harness Hub analysis
2. [Joel Niklaus on X: 10 harnesses on SWE-bench Pro](#) · Independent
3. [Arena, The harness tax \(21 model-harness pairs\)](#) · Independent
4. [Tibo \(OpenAI\) on X: Codex will seem primitive in 2-3 months](#) · Vendor-reported
5. [Y Combinator on X: why Claude Code deleted 80% of its system prompt](#) · Vendor-reported
6. [Stratechery, Agents over bubbles](#) · Independent
7. [Armin Ronacher, Better models: worse tools](#) · Independent
8. [OpenAI, Harness engineering: leveraging Codex in an agent-first world](#) · Vendor-reported
9. [Stripe, Minions: one-shot, end-to-end coding agents](#) · Vendor-reported
10. [HumanLayer, Why software factories fail](#) · Independent
11. [Hacker News: You said no MCP \(687 pts\)](#) · Community post
12. [Stack Overflow Developer Survey 2026 results](#) · Independent
13. [@DavidOndrej1 on X: Pi is the greatest agent harness ever created](#) · Community post
14. [Hacker News: replacing a \\$100/month Claude subscription with Pi on a local model](#) · Community post

12 PAIN POINTS → OPPORTUNITIES

Eight pains, ranked, and the product each one is asking for

Severity is Harness Hub's judgement from cross-platform weight and the evidence cited; five dots is the most acute.

Pain	Severity	Who feels it	Evidence	Opportunity for tool builders
Unpredictable cost and limits	●●●●●	Developers on subscriptions; FinOps and engineering leaders	Daily caps per engineer; a token floor that varies widely by harness; few organisations review spend (Section E) ^[1,2,3]	Tokens-per-task metering across harnesses, per-key and per-agent budgets, circuit breakers, spend dashboards
Review and verification bottleneck	●●●●●	Senior engineers, reviewers, engineering managers	Many teams run no evals; most open harnesses ship untested (Section E) ^[4,5]	Risk-tiered review, cross-harness agent review, provenance tags, evals as CI, spec capture
Security and blast radius	●●●●●	Engineers in auto modes; security and platform teams	Auto mode beaten up to 80% in a small test; marketplace malware; leaked keys ^[6,7,8]	MicroVM sandboxes per worktree, egress firewalls, secret injection, skill scanning, per-user agent identity
Lock-in and platform risk	●●●●○	Teams standardising on one vendor; procurement	Subscription cut-off; Gemini CLI shutdown; opaque harness behaviour ^[9,10,11]	Portable config, cross-harness session and memory portability, BYOK routers, audit of what a harness sends
Human attention and parallel-agent burnout	●●●●○	Power users and early orchestrator adopters	Burnout past a few sessions; "zero idea which one finished" ^[12,13,14]	Attention interfaces: summaries, notifications, handoff across devices and people
Configuration sprawl	●●●○○	Engineers maintaining harness setups; platform teams	Skills lists eating 10k+ tokens; calls for consolidation ^[15,16]	Skills package managers with evals, context-budget linters, team-shared harness config
No trustworthy harness benchmarks	●●●○○	CTOs choosing tools; harness vendors	Conflicting results; most benchmarks are vendor-run ^[17,18]	Independent model × harness × cost leaderboards that report dollars per solved task

Pain	Severity	Who feels it	Evidence	Opportunity for tool builders
Measuring productivity	●●●○○	Executives and boards	METR: 19% slower in 2025; 2026 estimates too uncertain to rely on ^[19,20]	Outcome analytics — merged, kept and reverted code, incidents — instead of token or line counts

■ SO WHAT FOR BUILDERS OF HARNESS TOOLS

- The white space users name most is metering, verification, containment, portability and independent benchmarks.
- Build on plugin APIs rather than forks, and assume the harness core changes every 90 days.

SOURCES

1. [r/ClaudeCode: my company now has daily limits to Claude Code](#) · Community post

4. [LangChain, State of Agent Engineering \(survey, n=1,340\)](#) · Vendor-reported

7. [@chiefofautism on X: the most-downloaded OpenClaw skill was malware](#) · Community post

10. [Hacker News: Gemini CLI shut down in favour of Antigravity CLI \(406 pts\)](#) · Community post

13. [@davj on X: 10 coding agents in parallel, zero idea which one finished](#) · Community post

16. [Hacker News: we need consolidation around commands, skills, subagents and plugins](#) · Community post

19. [METR, early-2025 AI and experienced open-source developers](#) · Independent

2. [Hacker News: Claude Code vs opencode token overhead \(706 pts\)](#) · Community post

5. [Marmelab, The state of AI harness engineering 2026](#) · Independent

8. [Hacker News: OpenCode Console key leak \(976 pts\)](#) · Community post

11. [Hacker News: Claude Code is steganographically marking requests \(2,445 pts\)](#) · Community post

14. [Paradev \(formerly Superset\), Roadmap to 100 agents](#) · Vendor-reported

17. [Hacker News: we really need better harness benchmarks](#) · Community post

20. [METR, uplift study update \(Feb 2026\)](#) · Independent

3. [Stack Overflow Developer Survey 2026 results](#) · Independent

6. [Embrace The Red, Breaking Claude Code's auto mode](#) · Independent

9. [Hacker News: third-party harnesses cut off from subscriptions \(1,099 pts\)](#) · Community post

12. [r/ClaudeCode: how are you managing several parallel sessions?](#) · Community post

15. [Hacker News: over 10k of context used just to list skills](#) · Community post

18. [Arena, The harness tax \(21 model-harness pairs\)](#) · Independent

13 NEW HARNESS STARTUPS

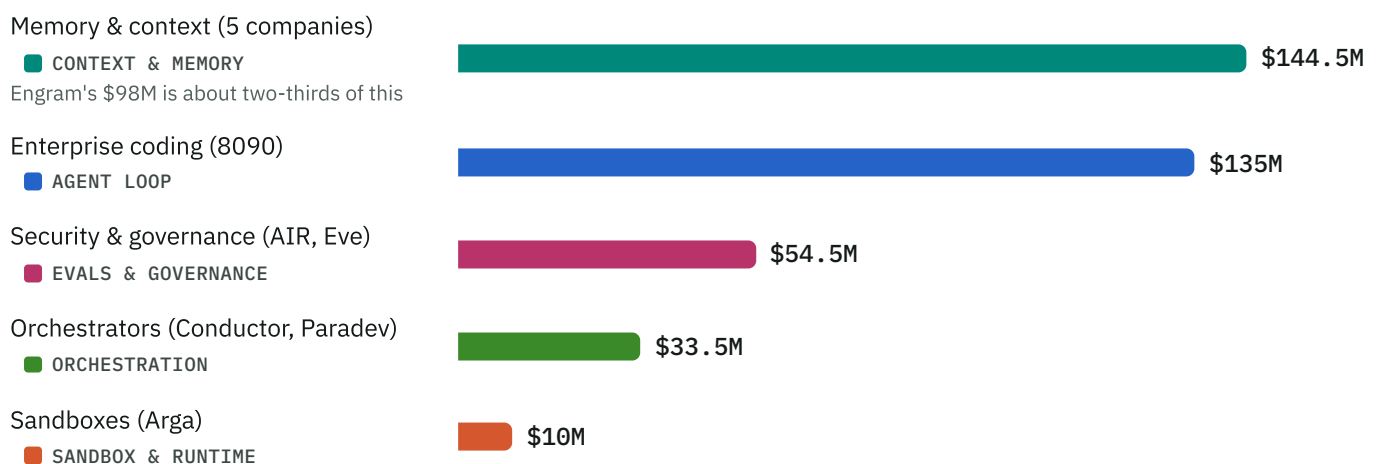
Among new harness startups, disclosed money went mostly to memory, security and sandboxes; the orchestrator crowd is thinly funded

Companies that launched or raised in the past 18 months, by layer; within a layer, funded companies first, then A–Z. Funding is shown only where a public source confirms it.

FIG. 14

Memory and context drew the most disclosed money among new harness-adjacent startups – two-thirds of it in one round

Disclosed 2026 funding rounds of new harness-adjacent startups, by layer



Rounds a company announced; undisclosed and unverified amounts are left out. Five rounds make up the memory total, so one large raise moves the ranking. Source [1,2,3,4,5,6,7,8,9,10,11]

These are new companies. The largest money of the year went to established harness vendors: Factory raised \$200M at a \$5B valuation in September, Cognition raised more than \$2B at \$48B in the same month (company-reported), and SpaceX completed a roughly \$60B acquisition of Cursor in August. Read together, investors are paying for proven harnesses at the top and for infrastructure layers among newcomers, while the new orchestrators raise little. [12,13,14]



Conductor

ORCHESTRATOR

Run a team of coding agents in parallel isolated workspaces, locally or in the cloud.

Jul 2025 · \$22M Series A^[9]



Paradev

AGENT IDE

Agentic IDE for running Claude Code, Codex and other CLI agents in parallel worktrees.

Dec 2025 · \$11.5M seed^[10]
Formerly Superset; renamed 26 Oct 2026.



1code (21st)

ORCHESTRATOR

An orchestration layer for coding agents such as Claude Code and Codex, from YC-backed 21st.

YC W26



BePurple (p0)

ORCHESTRATOR / SPEC AND QA

p0: a desktop app that orchestrates coding agents to bring software development in-house.

2026



Corgi

AGENT MULTIPLEXER

A helpful agent and terminal multiplexer for running several coding agents side by side.

Oct 2026



Palgu

ORCHESTRATOR / GOVERNANCE

An engineering 'superharness' that runs coding agents as one governed team.

2026

Disclosure: Palgu is built by the team behind Harness Hub. It gets the same card as every other startup, and it has no measurable community traction yet.



Orca

AGENT IDE (OPEN SOURCE)

Stably AI's open-source environment for running fleets of coding agents in parallel worktrees.

2026



Traycer

ORCHESTRATOR / SPEC

Nerve Center for Agentic Coding

2025



8090 Labs

ENTERPRISE CODING HARNESS

A "Software Factory" for enterprise teams that wants production-quality software, not vibe-coded prototypes.

Founded 2024 · \$135M Series A^[6]



Bullet

CODING AGENT

A faster coding agent, from the Y Combinator Summer 2026 batch.

Aug 2026



Twill

BACKGROUND AGENTS

Build locally, delegate to the cloud: runs Claude Code and Codex in cloud sandboxes and returns pull requests.

Apr 2026



Hoplite

CLOUD CODING AGENTS

Effortless cloud coding agents that carry your local sessions, memories and MCP servers to the cloud.

Aug 2026

 Engram MEMORY A learned memory layer for enterprise agents that prepares memories ahead of time to cut token costs. Jun 2026 · \$98M ^[1]	 Interloom MEMORY Turns expert workflow knowledge into lasting memory for AI agents. Mar 2026 · \$16.5M seed ^[2]
 Sage0x TEAM MEMORY The hivemind for human-agent teams: shared team memory that gives coding agents team context. May 2026 · \$15M seed ^[3]	 Nia CONTEXT FOR CODING AGENTS A context layer for coding agents that indexes code and documentation for Cursor, Claude Code and MCP clients. Dec 2025
 Skillsync SESSION SEARCH ACROSS AGENTS Searches past sessions across coding agents, so skills and memories are not locked into one provider. Sep 2026	 Arga Labs SANDBOXES & DIGITAL TWINS Sandboxes with stateful twins of the services agents act on Founded 2025 · \$10M seed ^[11]
 Agent37 MULTI-HARNESS SANDBOX HOST Isolated cloud sandboxes that host Hermes, OpenClaw, Claude Code, Codex, OpenCode, Pi and QM. Oct 2026	 Ardent DATABASE SANDBOXES Database sandboxes for you and your coding agents, so agents test against a realistic Postgres. May 2026
 AIR AGENT FIREWALL An inline firewall for AI agents and the skills, plugins and MCP servers they use. Sep 2026 · \$50M ^[7]	 Eve Security AGENT RUNTIME SECURITY Agentic AI runtime security: detects threats and enforces policy on enterprise agents in real time. Sep 2026 · \$4.5M seed ^[8]
 Alter AGENT ACCESS MANAGEMENT Privileged access management for AI agents, with per-end-user identity instead of one shared key. Oct 2026	 Sentrial AGENT MONITORING Observability for AI agents: catch drift and silent regressions across sessions and tool calls. Mar 2026

Churn is already visible: shutdowns and pivots in 2026 ^[15,16,17,18,19]

Company	What happened
---------	---------------

Company	What happened
Klaus	Hosted OpenClaw on a VM; the service ended on 1 September 2026
tilde.run	Agent sandbox with a versioned filesystem; shutdown announced 1 July 2026
Agent37	Started as an OpenClaw wrapper; now hosts many harnesses in sandboxes
Terminal Use	From filesystem-based agent hosting to rebuilding whole departments AI-first
Runtime	From sandboxed coding agents for teams to an agent harness for payment teams

 HACKER NEWS

“We get locked into a single provider and their agent as we invest in skills and memories over time.”

[Skillsync founders, Launch HN](#) · 17 Sep 2026

 HACKER NEWS

“the agent connects to the CRM or billing system with one shared API key, so the agent can do anything that the key can do.”

[Alter founders, Launch HN](#) · 21 Oct 2026

- SO WHAT FOR INVESTORS
- The biggest 2026 rounds went to established harness vendors (Factory, Cognition, Cursor). Among newcomers, disclosed money favoured memory, security and sandboxes, on a small number of rounds.
 - Orchestrator UIs and wrappers face labs adding the same features; their funding so far is small. Agent identity, insurance and FinOps look like a category in formation.

SOURCES

1. [Pulse 2.0: Engram launches with \\$98M](#) · Press report

4. [Cognee, seed announcement](#) · Vendor-reported

7. [Let's Data Science: AIR emerges with \\$50M](#) · Press report

10. [Paradev, Superset is now Paradev](#) · Vendor-reported

13. [The Next Web on Cognition's Series E](#) · Vendor-reported

16. [lakeFS: we recently shut down tilde.run](#) · Vendor-reported

19. [Runtime homepage](#) · Vendor-reported
2. [Interloom, seed announcement](#) · Vendor-reported

5. [Dealroom: Uvi raises \\$7.5M](#) · Press report

8. [SaaSRIse: Eve Security \\$4.5M seed](#) · Press report

11. [Dealroom: Arga raises a \\$10M seed](#) · Press report

14. [Bloomberg Law: SpaceX completes its \\$60B Cursor acquisition](#) · Press report

17. [Launch HN: Agent37 \(163 pts\)](#) · Community post
3. [TAMradar: SageOx \\$15M seed](#) · Press report

6. [TechCrunch: 8090 raises a \\$135M Series A](#) · Press report

9. [Conductor, Series A announcement](#) · Vendor-reported

12. [The Next Web on Factory's \\$5B round](#) · Press report

15. [Klaus: the service ended on September 1, 2026](#) · Vendor-reported

18. [Terminal Use homepage](#) · Vendor-reported

14 PLAYBOOKS

What this means for you

Five readers, five short playbooks. Colours match the takeaway boxes throughout the report.

■ Executives

CARES ABOUT: ROI, COST PREDICTABILITY, VENDOR RISK, TALENT

- The adoption question is closed; what decides value now is governance and unit economics.
- Plan the budget for the end of flat-rate subscriptions, and judge teams on what ships and stays shipped.
- Keep a second vendor warm: 2026 brought a cut-off, a shutdown and an acquisition.

■ CTOs & engineering leaders

CARES ABOUT: QUALITY, SECURITY, STANDARDISATION, THROUGHPUT

- Sequence the spend: verification capacity first, then containment, then more parallel agents.
- Expect engineers to run more than one harness, and make config, budgets and provenance work across all of them.

■ Engineers using harnesses

CARES ABOUT: FLOW, LIMITS, RELIABILITY, CRAFT

- Measure tokens per task per harness; route search and grunt work to cheap subagents; cap fan-out.
- Use a second harness as a reviewer.
- Keep to a few parallel sessions unless your verification can absorb more.
- Never run auto or yolo modes outside a sandbox; give API keys hard limits.
- Keep the design yourself and let the agent implement it.

■ Builders of harness tools

CARES ABOUT: GAPS, DEFENSIBILITY, DISTRIBUTION

- The gaps users name: cost metering, verification and evals, containment and identity, cross-harness portability, independent benchmarks, attention and handoff.
- Worktree orchestration is the most crowded new segment, and lab harnesses are adding parallel agents themselves; a new entrant needs a reason beyond the worktree.
- Design for many harnesses from day one and publish measurements: users trust numbers more than positioning.
- Whatever the layer, the burden of proof is on outcomes: merged, kept and reverted work, not agent counts.

■ Investors

CARES ABOUT: WHERE VALUE ACCRUES, TIMING, DURABILITY

- Established harness vendors drew the largest rounds and prices; among newcomers, infrastructure layers drew most disclosed money, on few rounds.
- Open harnesses have reach; whether they capture revenue at scale is unproven.
- Valuations repriced fast (Factory \$300M to \$5B in about a year); a subsidy reset would reprice harness economics again.

15 WHAT TO WATCH

Seven predictions for the next 12 months, to be scored in October 2027

Each has a threshold so it can be marked right or wrong. Next year's edition opens with the scorecard.

- | | | |
|----|---|------------|
| 01 | At least 1 of Anthropic and OpenAI moves its flagship coding subscription to metered, token-based billing for heavy users by Q2 2027. | ●●○ MEDIUM |
| 02 | An independent model × harness × cost leaderboard that reports dollars per solved task is cited by at least 3 of the top 5 harness vendors. | ●●○ MEDIUM |
| 03 | Parallel-agent orchestration ships built into at least 2 lab harnesses, and at least 3 standalone orchestrator startups shut down or pivot. | ●●● HIGH |
| 04 | Sandboxing on by default (microVM or container) ships in at least 2 major harnesses, replacing click-to-approve as the main safeguard. | ●●○ MEDIUM |
| 05 | AGENTS.md plus a portable skills format is supported by all 5 of the most-used harnesses. | ●●● HIGH |
| 06 | At least 1 agent security or identity company raises a round of \$100M or more. | ●●○ MEDIUM |
| 07 | An open-weight model in an open harness matches a lab model-harness pair on 1 major coding benchmark. | ●○○ LOW |

16 METHODOLOGY

How we made this report

State of Harness combines three kinds of evidence. First, public research: developer surveys (Stack Overflow, JetBrains, DORA, LangChain), independent studies (METR, Arena, arXiv papers), analyst and venture reports, company announcements and press coverage. Second, the conversation: about 250 high-engagement posts from Hacker News, Reddit, X and Substack, mostly from 2026, read and classified into eight discussion areas. Third, the Harness Hub directory of startups, open-source projects, events, news and papers, and the knowledge graph that links them.

Every external number and quote in this report was checked against its source page before publication; claims we could not confirm were left out. Vendor-reported figures are labelled as such, and anecdotes are presented as anecdotes. Discussion shares are indicative: the samples favour high-engagement posts and harness keywords, Reddit scores are archive snapshots, and they are not a census. Ratings, severity scores and the open-versus-closed positioning are our editorial judgement.

Affiliation: Palgu is built by the team behind Harness Hub. It is covered with the same rules as every other company, is listed in alphabetical or funding order, and is marked * in tables.

Licence: text and charts are CC BY 4.0. Cite as "Harness Hub, State of Harness, Vol. 1, October 2026". Spot an error or a missing harness? Use the submit page and we will correct it in the next edition.