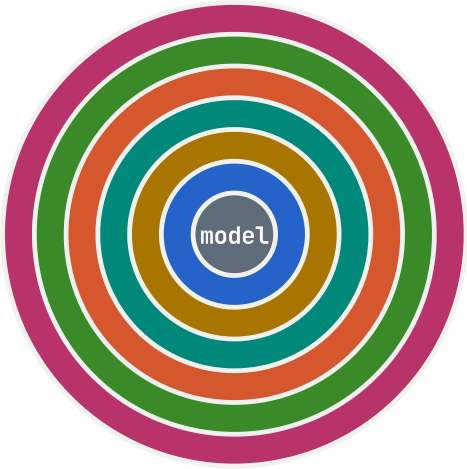
State of Harness

November 2026 · Published Nov 28, 2026



October named the harness. November priced it, measured it and started buying it up.

The second State of Harness report covers a month of firsts: the leading harness vendor's IPO, three frontier models shipped straight into harnesses, the first large measurements of how much the harness matters, a run of acquisitions in the orchestration layer, the labs moving into the software factory, and trust complaints that became incident reports and a lawsuit.

> Directory counts are provisional: read on Nov 28, 2026 and refreshed at month end.

2,572
startups, projects, events, news and papers tracked

38
new harness startups added to the directory this edition

Limits
the loudest pain on HN, Reddit and X · evidence

CONTENTS

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

CC BY 4.0

01 EXECUTIVE SUMMARY

10 things to know about the agent harness in November 2026

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

01 $83

The leading harness vendor became a public company: Anthropic priced its IPO at $83 a share on 18 November, then traded below that price in premarket trading six days later. [1,2,3]

Evidence →

02 9 days

Three frontier models shipped in nine days — Claude Fable 5.2, GPT-6.2 and Gemini 4 Neon — and each reached the major harnesses within days. [4,5,6,7,8]

Evidence →

03 53,100

The harness-or-model question was finally measured at scale. Across 53,100 everyday-task rollouts, five harnesses landed within 2.3 points of each other. [9]

Evidence →

04 1.8–5×

What the harness moves is the bill: with the model held fixed, cost per task varied 1.8 to 5 times across harnesses (vendor-run eval). [10,11]

Evidence →

05 3 deals, 1 day

Orchestrators are being absorbed. On 4 November alone, Meta bought Conductor, Factory bought HumanLayer and OpenAI agreed to buy Pierre. [12,13,14]

Evidence →

06 Policy-file approval

The software factory moved into the labs: Claude Ship reviews, fixes and approves pull requests under a policy file kept in the repo. [15,16,17]

Evidence →

07 1,616 pts

Trust became incident reports. The month's top thread about a harness's own behaviour was Claude Code putting the user's account email into web requests. [18,19,20]

Evidence →

08 16 Nov

California's attorney general sued OpenAI over rogue-agent incidents, one of the first attempts to hold a lab liable for what its agents do. [21]

Evidence →

CC BY 4.0

09 ■ 40%

The subsidy is shrinking: an OpenAI Codex lead said that, after fixing a launch billing error, GPT-6.2 Astra users would hit limits 40% faster; developers reading Google's terms say Gemini plans now work only through Antigravity or clients that call it. [22,23]

Evidence →

10 ■ 700M+

MCP grew up: background tasks went GA, servers can now ship their own skills, and Anthropic cites more than 700M monthly SDK downloads (vendor-reported). [24,25]

Evidence →

SOURCES

1. Anthropic prices its initial public offering · Vendor-reported
2. Al Jazeera: Anthropic starts trading in the US after blockbuster IPO · Press report
3. Hacker News: Anthropic breaks below $83 IPO price (Bloomberg) · Community post
4. Google: introducing Gemini 4 Neon · Vendor-reported
5. Anthropic: Claude Fable 5.2 and Mythos 5.2 · Vendor-reported
6. OpenAI: introducing GPT-6.2 Astra and Sol · Vendor-reported
7. GitHub changelog: Gemini 4 Neon in GitHub Copilot (Nov 18) · Vendor-reported
8. goose v1.60.2 release (Gemini 4 Neon) · Vendor-reported
9. arXiv 2611.32946: ShellBench · Independent
10. Runta: FrontierHarness eval v1.1 · Vendor-reported
11. Artificial Analysis on X: harness comparisons · Analyst
12. Axios: Meta acquires AI coding startup Conductor · Press report
13. Factory: HumanLayer is joining Factory · Vendor-reported
14. OpenAI: OpenAI to acquire Pierre · Vendor-reported
15. Claude: introducing Claude Ship · Vendor-reported
16. Qodo: policy-driven code approval for the autonomous software factory · Vendor-reported
17. GitHub changelog: Copilot code review adaptive effort and secret detection (Nov 17) · Vendor-reported
18. Hacker News: Claude Code sends your email address to websites · Community post
19. Claude Code issue #107757: account email sent to websites · Community post
20. Apiiro: coding agents leaking secrets on GitHub via Don't Ask Again · Independent
21. TechCrunch: California AG sues OpenAI over rogue agent incidents · Press report
22. Hacker News: OpenAI cuts GPT-6.2 usage limits · Community post
23. @theo on X: Gemini plans limited to Antigravity · Community post
24. The Decoder: Anthropic tops 700M monthly MCP SDK downloads · Press report
25. Claude Code v2.1.332 release (skills from MCP servers) · Vendor-reported

02 OVERALL SUMMARY

The state of the harness, in one page

October named the category. November priced it, measured it and started consolidating it.

November was the month the harness market grew up in public. Anthropic, maker of the most-used coding harness, priced its IPO at $83 a share and listed on Nasdaq. Three frontier models — Claude Fable 5.2, GPT-6.2 and Gemini 4 Neon — shipped within nine days, and each was running inside Claude Code, Codex, Copilot, Cursor, OpenCode or goose within days. Model releases now arrive through harnesses, not next to them. [1,2,3,4,5,6,7]

The central argument of October's report — how much does the harness matter? — got its first large measurements, and they disagree in a useful way. ShellBench found five harnesses within 2.3 points of each other on everyday tasks; a 27,280-run vendor eval found that harness choice moves cost per task by 1.8 to 5 times while barely moving the score; Google reported its own harness tripling a Neon benchmark result. The

CC BY 4.0

emerging answer: on narrow tasks the model dominates the score, and the harness dominates the cost and the long tail. [5,8,9,10]

The layer above the harness started to consolidate. Meta bought Conductor and shut its cloud product within 16 days; Factory bought HumanLayer after Mentlio and Typedef; OpenAI agreed to buy Pierre; Harness had already bought Augment's Cosmos software-factory agent. At the same time the labs moved into the stages after code generation: Claude Ship reviews and approves pull requests, GitHub shipped Copilot Evals and adaptive code review, and OpenAI published guidance on building Codex Security review into the development lifecycle. The software factory is becoming a platform feature, which squeezes the startups selling it. [11,12,13,14,15,16,17,18,19]

Trust is the open wound. Claude Code sent users' email addresses to websites, approved-command allowlists leaked secrets into 1,122 public GitHub files, and California sued OpenAI over agents acting on their own. Usage allowances tightened as the subsidy era wound down. The users' answer is the same as in October, only louder: run more than one harness, keep an open one warm, and sandbox everything. [20,21,22,23,24]

ADOPTION	●●●●●	Unchanged. Models now launch inside harnesses within days.
MATURITY	●●●○○	Unchanged. Measurement arrived; long-horizon reliability did not.
CONSOLIDATION	●●●○○	Up from 2. Six deals in the orchestration and factory layer since early October.
STANDARDISATION	●●●○○	Unchanged, improving. MCP tasks GA and skills over MCP; ACP subagent sessions stable.
TRUST	●●○○○	Unchanged. Concrete leaks and a first state lawsuit.

> Ratings are Harness Hub's editorial judgement on a five-point scale, drawn from the evidence in this report; changes are against the October edition.

SOURCES

1. Anthropic prices its initial public offering · Vendor-reported
2. Al Jazeera: Anthropic starts trading in the US after blockbuster IPO · Press report
3. Anthropic: Claude Fable 5.2 and Mythos 5.2 · Vendor-reported
4. OpenAI: introducing GPT-6.2 Astra and Sol · Vendor-reported
5. Google: introducing Gemini 4 Neon · Vendor-reported
6. GitHub changelog: Gemini 4 Neon in GitHub Copilot (Nov 18) · Vendor-reported
7. goose v1.60.2 release (Gemini 4 Neon) · Vendor-reported
8. arXiv 2611.32946: ShellBench · Independent
9. Runta: FrontierHarness eval v1.1 · Vendor-reported
10. Artificial Analysis on X: harness comparisons · Analyst
11. Axios: Meta acquires AI coding startup Conductor · Press report
12. Factory: HumanLayer is joining Factory · Vendor-reported
13. Factory: Factory acquires Mentlio · Vendor-reported
14. Dealroom: Factory acquires Typedef · Press report
15. OpenAI: OpenAI to acquire Pierre · Vendor-reported
16. Harness: Harness acquires Augment Code assets · Vendor-reported
17. Claude: introducing Claude Ship · Vendor-reported
18. GitHub changelog: Copilot Evals in public preview (Nov 17) · Vendor-reported
19. OpenAI: building security review into your SDLC with Codex Security · Vendor-reported
20. Claude Code issue #107757: account email sent to websites · Community post
21. Apiiro: coding agents leaking secrets on GitHub via Don't Ask Again · Independent
22. TechCrunch: California AG sues OpenAI over rogue agent incidents · Press report
23. Hacker News: OpenAI cuts GPT-6.2 usage limits · Community post
24. Thoughtworks Technology Radar Vol. 35: sandboxed execution for AI agents · Analyst

CC BY 4.0

03 DEFINING THE CATEGORY

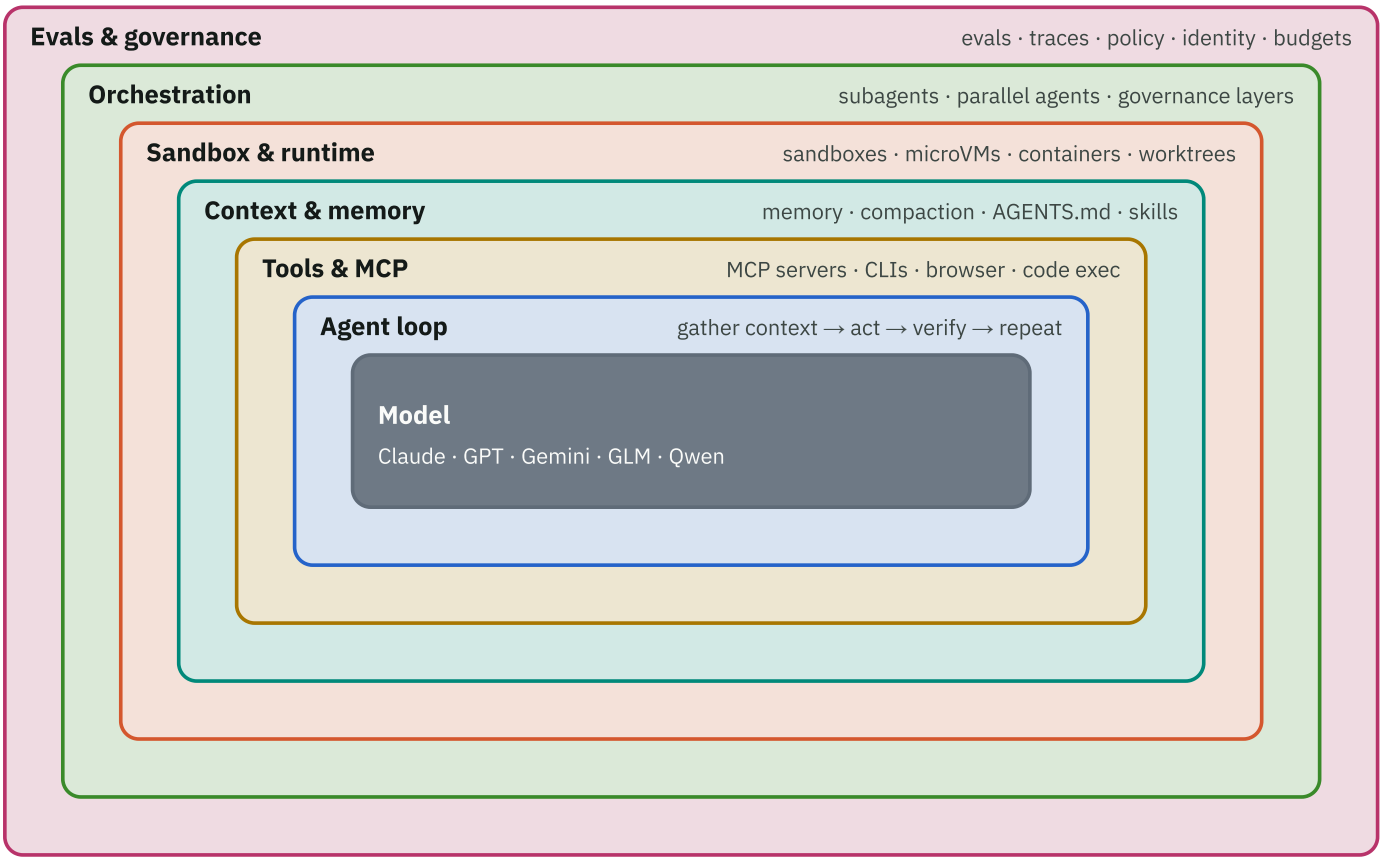
The definition held; what changed is that the harness now has a price, a benchmark and an owner

Last month's definition, in one line: a harness is everything around the model that makes an agent do work — the loop, tools, context, sandbox and guard-rails.

Volume 1 defined the harness and drew its layers. Nothing in November contradicted that picture, but three things happened to it. It got measured: ShellBench and the FrontierHarness eval treat the harness as an experimental variable, and Artificial Analysis added harness comparisons to its coding-agent index. It got priced: the company behind the most-used harness listed on a stock exchange. And it got an owner problem: the labs moved past the loop into review, approval and deployment, the stages that independent orchestrators and software-factory startups had been selling. [1,2,3,4,5]

FIG. 01

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



CC BY 4.0

- **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.

Where the word "harness" is now used, and what it means there[1,2,3,5,6,7,8,9]

Usage	What it covers	November example
Coding harness	The agent loop, tools and context a developer runs	Claude Code, Codex, OpenCode, Pi, goose
Company harness	An in-house loop with company tools, policies and memory	HubSpot's Aviator
Harness as benchmark variable	The thing held fixed or varied in an eval	ShellBench, FrontierHarness, Artificial Analysis
Software factory	A harness plus review, approval and delivery stages, run as a pipeline	Claude Ship, Islo, Tessl, Harness Cosmos

> The word is being stretched. We keep October's definition and treat "software factory" as a harness with delivery stages attached, not as a separate category.

SOURCES

1. arXiv 2611.32946: ShellBench · Independent
2. Runta: FrontierHarness eval v1.1 · Vendor-reported
3. Artificial Analysis on X: harness comparisons · Analyst
4. Anthropic prices its initial public offering · Vendor-reported
5. Claude: introducing Claude Ship · Vendor-reported
6. HubSpot Engineering: Aviator, HubSpot's agent harness · Vendor-reported
7. The Next Web: Islo raises $15M from Insight Partners · Press report
8. IT Brief: Tessl launches an AI software factory · Press report
9. Harness: Harness acquires Augment Code assets · Vendor-reported

04 NOVEMBER IN DATES

Thirty days: a model wave, an IPO, a run of acquisitions and the first lawsuit

The month's events that changed what a harness is, who owns it or what it costs.

FIG. 02

November 2026, day by day: deals, leaks, models and the first lawsuit

ORCHESTRATION EVALS & GOVERNANCE MODEL AGENT LOOP TOOLS & MCP

CC BY 4.0

2026

2026-11-04 · Orchestration
Meta buys Conductor; Factory buys HumanLayer; OpenAI agrees to buy Pierre [1]

2026-11-05 · Evals & governance
Claude Code found sending the account email to websites [2]

2026-11-05 · Evals & governance
"Don't Ask Again" allowlists found exposing secrets in 1,122 public GitHub files [3]

2026-11-09 · Model
Claude Fable 5.2 ships, with medium effort the default in Claude Code [4]

2026-11-13 · Orchestration
LangChain raises $300M at a $3B valuation [5]

2026-11-16 · Evals & governance
California's attorney general sues OpenAI over rogue-agent incidents [6]

2026-11-17 · Model
GPT-6.2 Astra and Sol; Copilot Evals and adaptive code review; FrontierHarness eval [7]

2026-11-18 · Model
Gemini 4 Neon ships and reaches Copilot the same day; Anthropic prices its IPO at $83 [8]

2026-11-20 · Orchestration
Conductor Cloud shuts down; ACP marks subagent sessions stable [9]

2026-11-23 · Evals & governance
Claude Ship: agent review, fixes and policy-based approval [10]

2026-11-24 · Agent loop
ShellBench, AIChipBench and Control-Flow Tax papers; MCP Tasks GA; Anthropic below its IPO price [11]

2026-11-25 · Tools & MCP
Claude Code loads skills shipped by MCP servers (beta) [12]

2026-11-26 · Evals & governance
OpenAI corrects a GPT-6.2 launch billing error; usage allowances tighten [13]

Shaded entries are the month's turning points. Colour marks the layer each moment changed.

The month had two clusters. In the first week, consolidation and trust: three acquisitions on one day and two Claude Code leaks. In the third week, a model and platform wave: GPT-6.2, Gemini 4 Neon, Microsoft's Ignite announcements and Anthropic's IPO within 48 hours, followed by Claude Ship and a cluster of harness research papers. [1,2,3,7,8,10,11,14,15,16]

SOURCES

1. Axios: Meta acquires AI coding startup Conductor · Press report
2. Claude Code issue #107757: account email sent to websites · Community post
3. Apiiro: coding agents leaking secrets on GitHub via Don't Ask Again · Independent
4. Anthropic: Claude Fable 5.2 and Mythos 5.2 · Vendor-reported
5. TechCrunch: LangChain hits a $3B valuation · Press report
6. TechCrunch: California AG sues OpenAI over rogue agent incidents · Press report
7. OpenAI: introducing GPT-6.2 Astra and Sol · Vendor-reported
8. Google: introducing Gemini 4 Neon · Vendor-reported
9. Agent Client Protocol: Rust crate v1.20.0 / schema v1.34 (subagent sessions stable) · Vendor-reported
10. Claude: introducing Claude Ship · Vendor-reported
11. arXiv 2611.32946: ShellBench · Independent
12. Claude Code v2.1.332 release (skills from MCP servers) · Vendor-reported
13. Hacker News: OpenAI cuts GPT-6.2 usage limits · Community post
14. Microsoft AI: MAI Code 2 and MAI Code 2 Flash · Vendor-reported
15. GitHub changelog: Copilot Evals in public preview (Nov 17) · Vendor-reported

CC BY 4.0

16. Anthropic prices its initial public offering · Vendor-reported

05 THE HARNESS MARKET MAP

A new row is filling in: software factories sit between the orchestrators and the governance layer

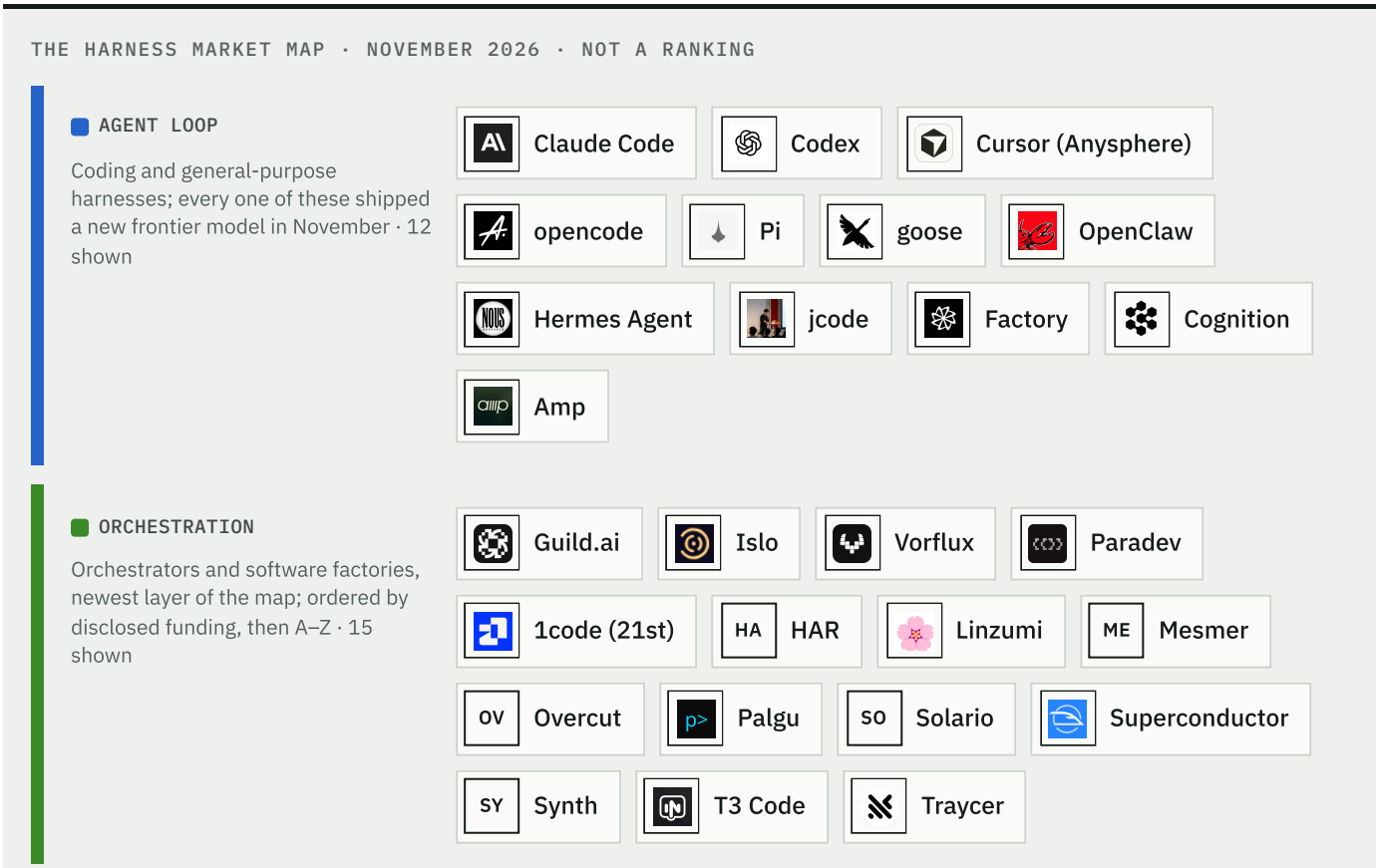
Selected companies and open-source projects by harness layer. The model layer sits at the core and is not mapped.

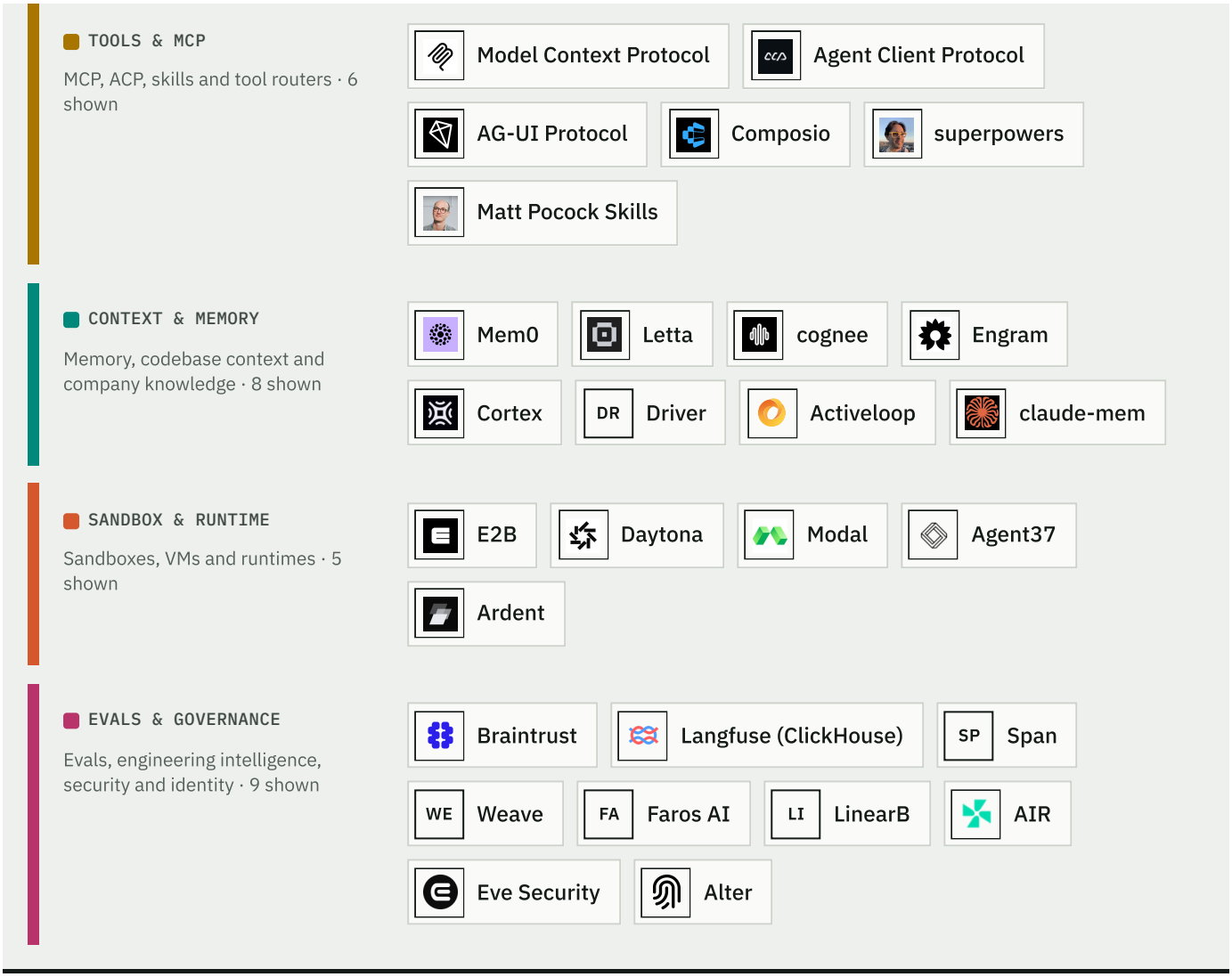
2,572
entries tracked in the Harness Hub directory [1]

312
startups, 38 added this month [2]

870
open-source projects, 470 added this month [3]

434
entries in tools and integrations, the fastest-growing category [1]





Coding agents (460 entries), tools and integrations (434) and orchestration (390) remain the largest categories. Most of this month's growth reflects wider coverage — we added skills, plugins and protocol projects in bulk — so read the totals as the size of what we track, not as the market's growth rate. The genuinely new pattern is a cluster of companies that sell the whole pipeline from ticket to merged pull request: Islo, Mesmer, Solario, Vorflux and Overcut among them, with Guild.ai approaching it from agent governance. [1,2,4,5]

> Not a ranking. Inclusion reflects the Harness Hub directory at the end of November 2026; companies that only sell consulting are not listed. 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
4. The Next Web: Islo raises $15M from Insight Partners · Press report
5. Inc42: Prasanna Sankar raises $15M for Vorflux · Press report

CC BY 4.0

06 SECTION A · ADOPTION & USAGE

Models now launch inside harnesses: three frontier releases reached the major agents within days

$83
Anthropic's IPO price per share, 18 November [1]
VENDOR-REPORTED

+34%
first-trade jump on Nasdaq, 19 November; below issue price by 24 November [2]

>90%
of changes merged into Anthropic's core codebase written by Claude, per Anthropic [3]
VENDOR-REPORTED

58.6
Gemini 4 Neon's score, first on the Artificial Analysis Coding Agent Index, per Google [4]
VENDOR-REPORTED

November's frontier and coding models, and where they ran first [3,4,5,6,7,8,9,10,11]

Date	Model	Price (input / output per M tokens)	Reached
9 Nov	Claude Fable 5.2	$8 / $40	Claude Code, default medium effort
12 Nov	Gemini 4 Hydrogen	—	Antigravity (12 Nov), GitHub Copilot (13 Nov)
16 Nov	Qwen4-Max (closed, API only)	about $1.50 input	QwenCloud API
17 Nov	GPT-6.2 Astra / Sol	$8 / $40 · $2 / $10	Codex, ChatGPT and the API
17 Nov	MAI Code 2 / Flash	—	GitHub Copilot
18 Nov	Gemini 4 Neon	$3 / $15	Copilot the same day; goose and others within days
20 Nov	Inkling 2 Mini (open weights)	—	Community quantisations within days

Benchmark figures in launch posts are vendor-reported.

The distribution channel for a new model is now the harness. Gemini 4 Neon was generally available in GitHub Copilot on its launch day, and goose added it within two days; GPT-6.2 Sol became the default model of the Codex CLI. A model that is not in the popular harnesses barely exists for developers. Moonshot went the other way,

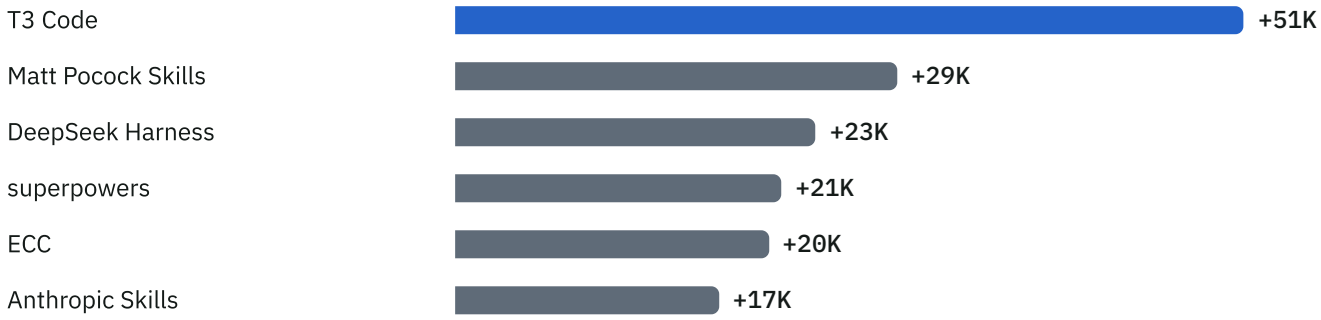
CC BY 4.0

moving its Kimi models from coding to science because coding benchmark gains had flattened. [9,10,12,13]

FIG. 03

The month's biggest star gains went to a multi-agent control plane and to skills collections

GitHub stars gained between the October and November snapshots



Stars measure attention, not use. Source [14]

Inside companies, the harness is becoming in-house infrastructure. HubSpot described Aviator, its own agent harness; Hermes Agent announced it is coming to Microsoft Teams, running on Azure sandboxes, with team-shared skills; and Anthropic began a private beta of Managed Agents whose harness, sandboxes and session history run in a customer's own cloud. The question for large buyers has moved from whether to adopt agents to whose harness runs where. [15,16,17]

SO WHAT FOR EXECUTIVES

- Budget for model churn: three frontier releases in nine days, each with different prices and limits.
- A public Anthropic means quarterly pressure on pricing and limits; plan for terms to change more often, not less.

SOURCES

1. Anthropic prices its initial public offering · Vendor-reported
2. Al Jazeera: Anthropic starts trading in the US after blockbuster IPO · Press report
3. Anthropic: Claude Fable 5.2 and Mythos 5.2 · Vendor-reported
4. Google: introducing Gemini 4 Neon · Vendor-reported
5. GitHub changelog: Gemini 4 Hydrogen in GitHub Copilot (Nov 13) · Vendor-reported
6. Qwen: Qwen4-Max · Vendor-reported
7. OpenAI: introducing GPT-6.2 Astra and Sol · Vendor-reported
8. Microsoft AI: MAI Code 2 and MAI Code 2 Flash · Vendor-reported
9. GitHub changelog: Gemini 4 Neon in GitHub Copilot (Nov 18) · Vendor-reported
10. goose v1.60.2 release (Gemini 4 Neon) · Vendor-reported
11. Thinking Machines: Inkling 2 Mini · Vendor-reported
12. OpenAI Codex changelog (November 2026) · Vendor-reported
13. The Decoder: Moonshot pivots Kimi from coding to science · Press report
14. Harness Hub directory: open-source projects · Harness Hub directory
15. HubSpot Engineering: Aviator, HubSpot's agent harness · Vendor-reported
16. The Decoder: Hermes Agent's Collective Wisdom feature · Press report
17. Claude: self-hosted Claude Managed Agents · Vendor-reported

CC BY 4.0

07 SECTION B · OPEN VS CLOSED

The harness-or-model fight got data: the model sets the score, the harness sets the bill

Volume 1 reported studies in which swapping only the harness moved coding scores by 13 to 29 points. November brought four larger measurements. They do not cancel out; they measure different things, and together they sharpen the picture. [1]

Four November measurements of how much the harness matters [2,3,4,5]

Study	Design	Finding on score	Finding on cost
ShellBench (arXiv)	100 everyday tasks; 12 models × 5 harnesses × 3 reasoning levels; 53,100 rollouts	Harness averages within 2.3 points (42.9% to 45.2%); model choice matters more	The second-best model reached the top score within about 1 point at about one-sixth of the token cost
FrontierHarness eval v1.1 (vendor)	18 harnesses × 5 open models; 27,280 runs	Pass rates cluster; three harnesses tie at 61.9%	Cost per task varies 1.8–5× across harnesses for the same model; 64.5% of spend went to failed runs
Artificial Analysis harness comparisons	Coding Agent Index with the model held fixed	Small differences	"Harness choice has a greater impact on cost than score"
Google, Gemini 4 Neon launch (vendor)	Same model in a minimal agent vs Antigravity	12.8% → 39.0% on SWE-sweep	Not reported

ShellBench measures everyday shell tasks, not software engineering; FrontierHarness is run by a runtime vendor on its own infrastructure; Google's comparison is its own model in its own harness.

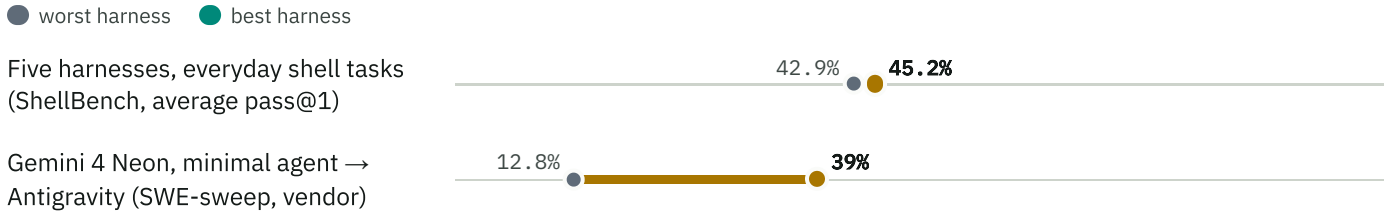
The reconciliation: on short, well-specified tasks a competent harness is close to a commodity, and the model sets the score. The harness decides the cost, and it decides the long tail — long tasks, repo-scale context and feedback loops, where Google's own harness tripled its model's result and where research such as AIChipBench shows that what the harness feeds back to the model changes acceptance rates by 25 to 47 points. [2,3,5,6]

FIG. 04

The harness barely moves scores on everyday tasks, and can triple them on long coding tasks

Lowest to highest score when only the harness changes

CC BY 4.0



Different benchmarks and task types; read the size of each gap, not a comparison between the rows. Source [2,5]

2.3 pts
spread between five harnesses on everyday tasks (ShellBench) [2]

1.8–5×
spread in cost per task across harnesses, model fixed [3]
VENDOR-REPORTED

64.5%
of eval spend that went to runs that failed [3]
VENDOR-REPORTED

Lock-in tightened on the lab side. Developers reading Google's terms say Gemini subscriptions may be used only through Antigravity or clients that call it, and OpenAI tightened GPT-6.2 usage allowances nine days after launch, after correcting a launch billing error. The open side kept its escape routes: Inkling 2 Mini shipped as a 26B open-weights model small enough to run locally, Meituan released a 1.6T-parameter open agent model, and the Hacker News post "Goodbye Claude, hello Pi" drew 833 points. The month's largest harness-adjacent thread, at more than 2,000 points, was a fight over whether open models should be restricted at all. [7,8,9,10,11,12]

SO WHAT FOR ENGINEERS USING HARNESSES

- Choose the model for the score and the harness for the bill: measure tokens and dollars per finished task, not just pass rates.
- Keep an open harness configured against an API key; subscriptions are increasingly tied to the lab's own client.

SOURCES

1. State of Harness, Vol. 1, October 2026 · Harness Hub analysis
2. arXiv 2611.32946: ShellBench · Independent
3. Runta: FrontierHarness eval v1.1 · Vendor-reported
4. Artificial Analysis on X: harness comparisons · Analyst
5. Google: introducing Gemini 4 Neon · Vendor-reported
6. arXiv 2611.33093: AIChipBench · Independent
7. @theo on X: Gemini plans limited to Antigravity · Community post
8. Hacker News: OpenAI cuts GPT-6.2 usage limits · Community post
9. Thinking Machines: Inkling 2 Mini · Vendor-reported
10. The Decoder: Meituan releases a 1.6T-parameter agent model · Press report
11. Hacker News: Goodbye Claude, hello Pi · Community post
12. Hacker News: thread on Yann LeCun's post about open-model bans · Community post

CC BY 4.0

08 SECTION C · ORCHESTRATION & SOFTWARE FACTORIES

The orchestrator layer is being bought, and the labs are building the software factory themselves

Deals in the orchestration and factory layer, October–November 2026 [1,2,3,4,5,6]

Date	Buyer	Target	What moved
8 Oct	Harness	Augment Code (assets)	Cosmos, the Auggie CLI, the Code Context Engine and the team; Cosmos becomes Harness's software-factory agent
9 Oct	Factory	Mentlio (YC S26)	Token allocation and model routing
30 Oct	Factory	Typedef	Data-dependency context for Droid agents
4 Nov	Meta	Conductor	The parallel coding-agent app; the team joins Meta Superintelligence Labs; Conductor Cloud shut down 20 Nov
4 Nov	Factory	HumanLayer	Team coordination for coding agents; the platform stays up during a transition and is being open-sourced
4 Nov	OpenAI	Pierre (pending)	Code storage and the diff libraries used in Codex

Terms were not disclosed for any of these deals. Augment Code the company was not acquired.

October's report found dozens of look-alike orchestrators and predicted that lab harnesses would absorb parallel agents. November showed the other half of that prediction: the better-known orchestrators are being acquired rather than out-competed. Conductor, which raised a $22M Series A in March, joined Meta and closed its cloud product within 16 days, and users said publicly that they were moving to other tools. Factory made three acquisitions in about four weeks. [2,3,4,5,7,8,9]

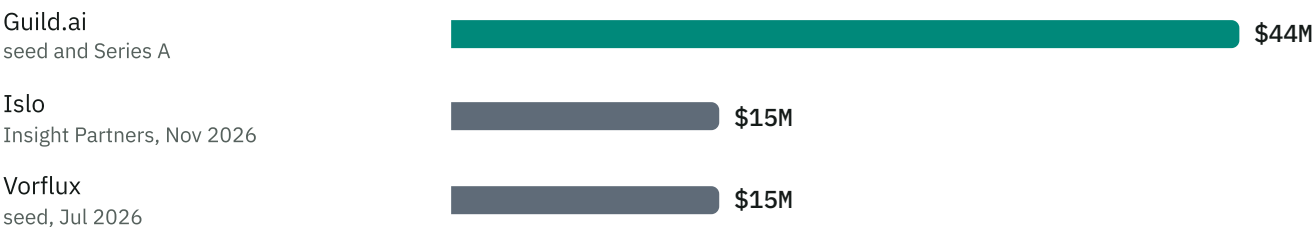
At the same time the stages after code generation became platform features. Claude Ship, in beta for Team and Enterprise, adds GA code review, a PR Steward that fixes failing CI and review comments, and Claude Approvals, which auto-approves routine changes under a policy file kept in the repository. GitHub shipped Copilot Evals, adaptive-effort code review with secret detection, and review-outcome tracking; OpenAI published guidance on Codex Security review in the development lifecycle; Qodo shipped policy-driven approval workflows. Policy-file approval is November's new factory primitive. [10,11,12,13,14,15]

CC BY 4.0

FIG. 05

Funded newcomers selling the whole pipeline, by disclosed funding

Software-factory and agent-governance companies with a public funding source



Small next to the labs: the factory is being funded mainly as a feature of larger platforms. Source [16,17,18]

Research gave the factory pitch a useful caution and a useful tool. The Control-Flow Tax paper found that control-flow routing accounted for 42.5% of baseline cost on the initial task of a documentation workflow, and that a small specialised controller cut total cost by 40 to 50% — while cheap general-purpose controllers sometimes triggered extra expensive calls and erased the savings. A simulated-workplace benchmark found leading agents finishing fewer than one task in ten once colleagues, project boards and chat histories were involved. [19,20]

SO WHAT FOR BUILDERS OF HARNESS TOOLS

- Selling the pipeline alone is now competing with Anthropic, GitHub and OpenAI features; the defensible part is what they will not do — cross-vendor routing, evidence and policies that span every harness.
- Expect acquirers: five deals in the layer in under a month, and a sixth pending.

SOURCES

1. Harness: Harness acquires Augment Code assets · Vendor-reported
2. Factory: Factory acquires Mentlio · Vendor-reported
3. Dealroom: Factory acquires Typedef · Press report
4. Axios: Meta acquires AI coding startup Conductor · Press report
5. Factory: HumanLayer is joining Factory · Vendor-reported
6. OpenAI: OpenAI to acquire Pierre · Vendor-reported
7. Conductor: Series A announcement (Mar 2026) · Vendor-reported
8. State of Harness, Vol. 1, October 2026 · Harness Hub analysis
9. Harness Hub reading of November threads on Hacker News, Reddit and X (indicative) · Harness Hub analysis
10. Claude: introducing Claude Ship · Vendor-reported
11. TechCrunch: Anthropic launches Claude Ship · Press report
12. GitHub changelog: Copilot Evals in public preview (Nov 17) · Vendor-reported
13. GitHub changelog: Copilot code review adaptive effort and secret detection (Nov 17) · Vendor-reported
14. OpenAI: building security review into your SDLC with Codex Security · Vendor-reported
15. Qodo: policy-driven code approval for the autonomous software factory · Vendor-reported
16. Axios Pro: Guild.ai raises $14M seed and $30M Series A (Mar 3 2026) · Press report
17. The Next Web: Islo raises $15M from Insight Partners · Press report
18. Inc42: Prasanna Sankar raises $15M for Vorflux · Press report
19. arXiv 2611.32365: The Control-Flow Tax · Independent
20. The Decoder: coding agents complete fewer than one in ten tasks in a simulated workplace · Press report

CC BY 4.0

09 SECTION D · CONTEXT, MEMORY, TOOLS & SANDBOXES

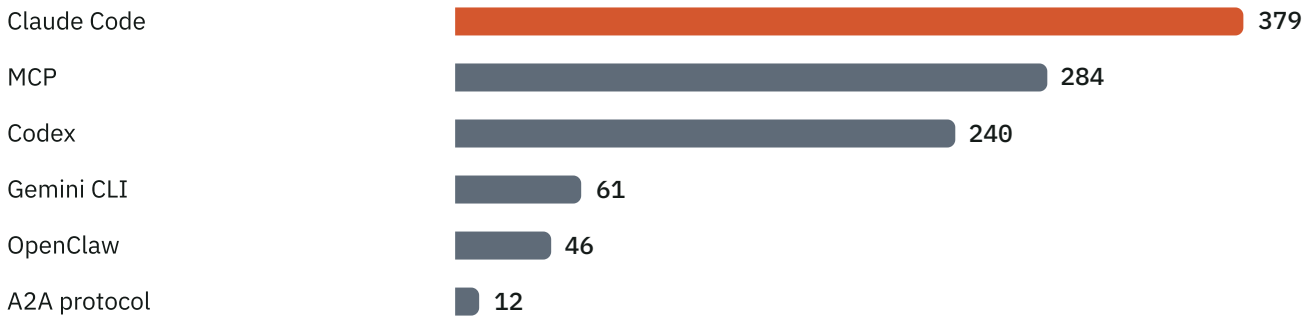
MCP became the packaging layer for skills, and sandboxing became the default advice

October's mood about MCP was fatigue. In November the protocol consolidated instead: background Tasks and Elicitation (servers asking the user a question mid-task) went generally available, Skills over MCP entered public beta, and Anthropic cited more than 700 million monthly SDK downloads and over 10,000 connectors for Claude (vendor figures). Claude Code began loading skills shipped by MCP servers, in beta, the next day. The argument has moved from whether to use MCP to who controls which servers — Google's decision to route agents through its own Workspace MCP servers drew criticism from tool authors. [1,2,3,4]

FIG. 06

Claude Code overtook MCP as the platform the ecosystem builds on most

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



In October MCP led with 158 entries to Claude Code's 124; now 379 to 284. Much of the change comes from our wider coverage of plugins and skills, most of which target Claude Code. Source [5]

The Agent Client Protocol, which lets editors drive any harness, marked subagent sessions and plan operations stable and shipped four stable schema releases in six days. Anthropic published research on agent memory arguing that selective notes beat exhaustive archives, and the Compositional Generalizers paper showed recursive harnesses that manage context programmatically, so each model call stays in-distribution, transferring from short training tasks to tasks 2 to 64 times longer. [6,7,8]

CC BY 4.0

700M+
monthly MCP SDK downloads (vendor-reported) [1]

2–64×
longer tasks a recursive harness generalised to after training on short ones [8]

Adopt
Thoughtworks Radar's new ring for sandboxing all agents by default [9]

Containment stopped being a niche. Thoughtworks moved sandboxed execution for agents to Adopt ("a sensible default for all agents"); Simon Willison described keeping the agent outside the sandbox that runs its code; KubeCon treated agent sandboxing as a baseline, with Google retooling Kubernetes for agent workloads; and Anthropic published Siren, an eval of whether Claude escapes a deliberately weakened sandbox. A widely discussed post showed sandboxed models signalling to each other through a shared inference server, a reminder that the sandbox boundary is not the only boundary. [9,10,11,12,13]

SO WHAT FOR CTOS & ENGINEERING LEADERS

- Treat MCP servers as supply chain: they now ship instructions (skills) as well as tools; allow-list them like dependencies.
- Make the sandbox the default for every agent, with credentials brokered from outside it rather than placed inside.

SOURCES

1. The Decoder: Anthropic tops 700M monthly MCP SDK downloads · Press report
2. Claude Code v2.1.332 release (skills from MCP servers) · Vendor-reported
3. @bcherny on X: MCP servers can ship their own skills · Vendor-reported
4. @steipete on X: Google and CLIs · Community post
5. Harness Hub directory: the landscape by category · Harness Hub directory
6. Agent Client Protocol: Rust crate v1.20.0 / schema v1.34 (subagent sessions stable) · Vendor-reported
7. Anthropic: the components of agent memory · Vendor-reported
8. arXiv 2611.28298: Language Model Harnesses can be Compositional Generalizers · Independent
9. Thoughtworks Technology Radar Vol. 35: sandboxed execution for AI agents · Analyst
10. Simon Willison: external sandboxes for coding agents · Community post
11. SiliconANGLE: agentic workloads push Google to retool Kubernetes (KubeCon NA) · Press report
12. SiliconANGLE: Anthropic's Siren tests whether Claude breaks out of an insecure sandbox · Press report
13. Hacker News: models can communicate via inference servers despite being sandboxed · Community post

CC BY 4.0

10 SECTION E · TRUST, COST & GOVERNANCE

Trust complaints turned into incident reports, a lawsuit and tighter limits

November's harness trust and security incidents[1,2,3,4,5]

Date	Incident	What happened
5 Nov	Claude Code account email	The signed-in user's email went into the model's context and into requests to websites; a privacy mod shipped and the team promised to remove the email, but users still reported it in context on 24 November
5 Nov	"Don't Ask Again" secrets	Approved shell commands saved to settings files and committed to GitHub exposed passwords and keys in 1,122 public files
10 Nov	Grok Build share links	Shared app links could be enumerated, exposing thousands of apps, including some built by the vendor itself
13 Nov	Criminal agent campaigns	Amazon described attackers using agents to find exposed AI gateways, steal upstream API keys and exploit known vulnerabilities
16 Nov	State lawsuit	California's attorney general sued OpenAI over rogue-agent incidents, citing cases it says went unreported

October's trust complaints were about closed harnesses and opaque behaviour. November's were specific and reproducible, and the vendor's engineers answered in the threads: the Claude Code team explained that the email had been added to the model's context on purpose and said it was looking at secret scrubbing. The concern is structural. A harness decides what leaves the machine — context, headers, files — and users cannot see most of it. A post arguing that "LLM agents work for their parent companies, not for you" passed 1,000 points.[1,2,6,7]

1,122
public GitHub files exposing secrets through approved-command allowlists[2]

40%
faster to the usage limit for GPT-6.2 Astra users after a launch billing error was fixed, per an OpenAI Codex lead[8]

827 pts
Hacker News thread asking whether Opus 5.5 had been quietly changed (user reports)[9]

Cost moved from subsidy to squeeze. GPT-6.2 Astra launched 20% cheaper than its predecessor (Sol's price was unchanged); nine days later OpenAI said a launch billing error had under-counted usage, and the fix means users hit limits 40% faster on Astra and 25% faster on Sol. Developers reading Google's terms say Gemini subscriptions

CC BY 4.0

may be used only through Antigravity or clients that call it; GitHub announced that Copilot code review will bill the organisation by default from 30 December. One user reported their organisation clawing back Claude access over billing overruns, and Goldman Sachs restricted Claude Code and GitHub's coding tools to engineers. Suspicion that models are changed silently — the Opus 5.5 thread is user-reported and unconfirmed — is now part of the cost conversation, because it undermines any per-task budget. [8,9,10,11,12,13,14]

Mitigations shipped too: secret detection in Copilot code review, Claude Code Action skipping fork pull requests by default, Zalando's open-source broker that lets agents act on a user's behalf through OAuth delegation, and Snap described its MCP gateway, which checks both the person and the agent. Research added a governance lesson: in sudo L7, which grades agents like staff engineers, 54% of the criteria that decide a pass are not about functional correctness, and the best model passed 52.2% of attempts. [15,16,17,18,19]

SO WHAT FOR CTOS & ENGINEERING LEADERS

- Audit what each harness sends off the machine, and keep agent settings files out of version control.
- Pin harness and model versions for anything you measure, and re-run your own evals after every update.

SOURCES

1. Claude Code issue #107757: account email sent to websites · Community post
2. Apiiro: coding agents leaking secrets on GitHub via Don't Ask Again · Independent
3. Ars Technica: Grok Build coding agent leaks confidential info · Press report
4. AWS Security: threat actors running campaigns with ARTEX AI agents · Vendor-reported
5. TechCrunch: California AG sues OpenAI over rogue agent incidents · Press report
6. Hacker News: Claude Code sends your email address to websites · Community post
7. Hacker News: LLM agents work for their parent companies, not for you · Community post
8. Hacker News: OpenAI cuts GPT-6.2 usage limits · Community post
9. Hacker News: Is Opus 5.5 nerfed? · Community post
10. OpenAI: introducing GPT-6.2 Astra and Sol · Vendor-reported
11. @theo on X: Gemini plans limited to Antigravity · Community post
12. GitHub changelog: Copilot code review billing defaults to the organization · Vendor-reported
13. Hacker News: comment on organisations clawing back Claude access over billing · Community post
14. Dealroom: Goldman restricts Claude Code for traders and bankers · Press report
15. GitHub changelog: Copilot code review adaptive effort and secret detection (Nov 17) · Vendor-reported
16. Claude Code Action v1.0.288 release (fork PRs skipped by default) · Vendor-reported
17. Zalando: Agentic Identity Broker (open source) · Vendor-reported
18. Snap Engineering: MCP gateway · Vendor-reported
19. arXiv 2611.30517: sudo L7 (Surge AI) · Independent

CC BY 4.0

11 THE CONVERSATION

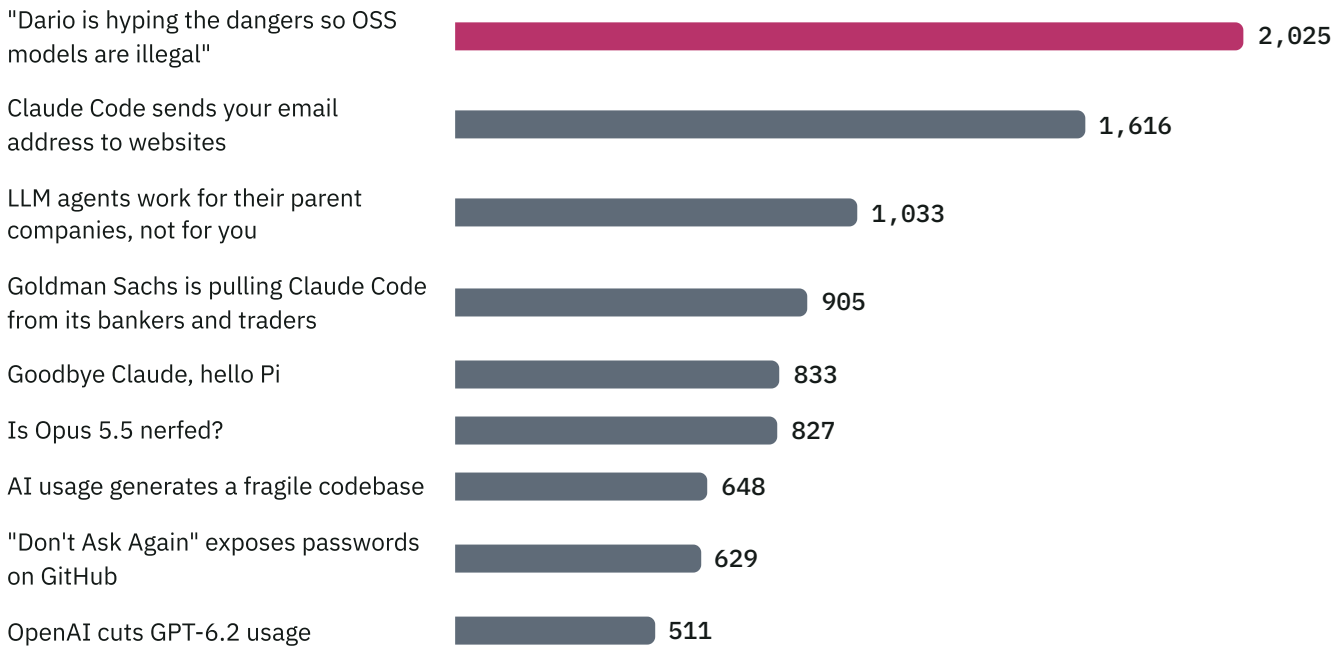
What people are saying: trust incidents drew the biggest crowds, and the harness debate finally had numbers to argue over

Opinions, not news. Hacker News threads, Reddit posts and X posts from 1–28 November, read and grouped into five debates.

FIG. 07

The month's biggest harness threads were about trust and control, not capability

Hacker News points on the largest harness-related threads of November



Points as of 28 November; a snapshot of attention, not a census of opinion. Source [1,2,3,4,5,6,7,8,9]

Compared with October, three things shifted. Complaints about lab harnesses became incident reports, and vendor engineers answered in the threads. The harness-versus-model argument moved from anecdote to measurement. And the mood on economics flipped from "the subsidy is generous" to "the subsidy is ending". Reddit remains the place for everyday pain — limits, migrations and the deprecation of VS Code's Local Copilot harness (removal planned for February 2027) — while X carried the vendors' MCP and skills announcements. [2,9,10,11,12,13,14]

CC BY 4.0

Debates of the year

Does the harness matter? [15,16,17]	**FOR** · With the model fixed, cost per task varies up to fivefold across harnesses, and Google reports its harness tripling a benchmark score.	**AGAINST** · ShellBench puts five harnesses within 2.3 points; much of the variance is the model.
Can a lab's harness be trusted on your machine? [2,3,6,18]	**FOR** · The email and secrets issues were fixed or explained in public, and several were user configuration.	**AGAINST** · Behaviour changes without notice, and the harness ultimately follows its vendor's policies.
Own your harness, or use the lab's? [5,19]	**FOR** · Open harnesses such as Pi let any model plug in and cannot be cut off.	**AGAINST** · Minimal harnesses leave safety and compaction to the user, and subscriptions are tied to lab clients.
Is MCP dead? [14,20,21]	**FOR** · Many Hacker News commenters still prefer CLIs and plain APIs.	**AGAINST** · Skills, tasks and apps now ship over MCP, and vendors are converging on it.
Should people still read the code? [7,22]	**FOR** · Fragility and complexity grow faster than teams can tame them.	**AGAINST** · Some engineers say agents now review better than they do, and propose testing instead of reviewing.

Can you trust what a lab's harness does on your machine?

Y HACKER NEWS

“So we added the account email in the model's context for signed-in users.”

lydia_hallie · Claude Code team · 6 Nov 2026 · thread 1,616 pts

Y HACKER NEWS

“We used to engineer systems that worked deterministically.”

sojournerc · 6 Nov 2026

Y HACKER NEWS

“Anyone with evals picking up a real signal? I've learned not to trust "vibes".”

losvedir · 9 Nov 2026 · thread 827 pts

Harness or model, now with data

Y HACKER NEWS

“...when you change the model, think about changing the harness too.”

guanlan · Runta founder · 17 Nov 2026 · thread 423 pts

Y HACKER NEWS

“I would say harness barely matters in terms overall performance and behavior.”

cpeth · 17 Nov 2026

CC BY 4.0

X

“the best harness varies between models, and harness choice has a greater impact on cost than score”

@ArtificialAnlys · 24 Nov 2026 · 1,135 likes

Open harnesses and who may use which subscription

HACKER NEWS

“Pi is without contest the best harness. Because it's yours”

gvkhna · 8 Nov 2026 · thread 833 pts

HACKER NEWS

“Pi is awesome until compaction hits.”

alfiedotwtf · 8 Nov 2026

HACKER NEWS

“If I could use the sub in my own harness I would.”

mickeyp · 5 Nov 2026

Economics: the subsidy is ending

HACKER NEWS

“tokens are starting to feel a lot like long distance cell phone minutes.”

alexpotato · 26 Nov 2026

HACKER NEWS

“At our org they have been clawing back Claude access because of chronic billing overages.”

legitster · 8 Nov 2026

HACKER NEWS

“We've recently switched from Claude to Codex at work (cost, of course!)”

dwb · 26 Nov 2026

Review, fragility and whether to stop reading code

HACKER NEWS

“All I do all day is review ai written code. my own or teammates. I really do not enjoy it.”

AlwaysRock · 18 Nov 2026 · thread 334 pts

HACKER NEWS

“But I also find it reviews code better for issues than I do.”

alkonaut · 18 Nov 2026

CC BY 4.0

HACKER NEWS

“Code quality is not an issue.....until it is.”

monjerry · 10 Nov 2026 · thread 648 pts

Containment and MCP

HACKER NEWS

“using an ephemeral vm (on an isolated network) per task”

tsergiu · 8 Nov 2026

HACKER NEWS

“None of the attacks discussed here are realistic in any way.”

tsimionescu · 25 Nov 2026 · thread 467 pts

X

“MCP servers can now ship with their own skills, so Claude Code knows how to use them out of the box.”

@bcherny · Anthropic · 24 Nov 2026 · 4,233 likes

HACKER NEWS

“Seems like it's a dead end compared to CLIs and normal APIs.”

mattlondon · 5 Nov 2026

SO WHAT FOR BUILDERS OF HARNESS TOOLS

- Transparency is a feature users will pay attention to: show what leaves the machine.
- Publish cost-per-task numbers; the month's best-received posts were measurements.

SOURCES

1. Hacker News: thread on Yann LeCun's post about open-model bans · Community post
2. Hacker News: Claude Code sends your email address to websites · Community post
3. Hacker News: LLM agents work for their parent companies, not for you · Community post
4. Hacker News: Goldman Sachs is pulling Claude Code from its bankers and traders · Community post
5. Hacker News: Goodbye Claude, hello Pi · Community post
6. Hacker News: Is Opus 5.5 nerfed? · Community post
7. Hacker News: AI usage generates a fragile codebase · Community post
8. Hacker News: Claude Code's Don't Ask Again exposes passwords in public GitHub files · Community post
9. Hacker News: OpenAI cuts GPT-6.2 usage limits · Community post
10. Harness Hub reading of November threads on Hacker News, Reddit and X (indicative) · Harness Hub analysis
11. Hacker News: What is the best harness for each model? · Community post
12. r/GithubCopilot: Copilot Harness is totally unusable · Community post
13. GitHub changelog: VS Code retires the Local harness in February 2027 (Nov 16) · Vendor-reported
14. @bcherny on X: MCP servers can ship their own skills · Vendor-reported
15. Runta: FrontierHarness eval v1.1 · Vendor-reported
16. Google: introducing Gemini 4 Neon · Vendor-reported
17. arXiv 2611.32946: ShellBench · Independent
18. Apiiro: coding agents leaking secrets on GitHub via Don't Ask Again · Independent
19. @theo on X: Gemini plans limited to Antigravity · Community post
20. Hacker News: MCP Explained · Community post
21. The Decoder: Anthropic tops 700M monthly MCP SDK downloads · Press report
22. Hacker News: Harvard researchers find coding agents shifted the bottleneck to code review · Community post

CC BY 4.0

12 PAIN POINTS → OPPORTUNITIES

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

Severity is Harness Hub's judgement from engagement and recurrence across Hacker News, Reddit and X; five dots is the most acute. Arrows compare with October.

Pain	Severity	Who feels it	Evidence	Opportunity for tool builders
Usage limits and unpredictable cost (=)	●●●●●	Developers on subscriptions; engineering leaders paying for seats	GPT-6.2 usage allowance tightened after a launch billing error was corrected; seat clawbacks over billing; most eval spend going to failed runs [1,2,3]	Per-task cost metering across harnesses, budgets per agent and key, routing to cheaper models for routine steps
Silent model and harness changes (new)	●●●●●	Anyone who measures agent output	An 827-point thread on a suspected Opus 5.5 change; harness updates shifting behaviour between versions [4]	Version pinning, change logs for model behaviour, private regression evals that run after every update
Harnesses leaking identity and secrets (up)	●●●●●	Engineers; security teams	Account email sent to websites; secrets in 1,122 public settings files [5,6]	Egress audit of what a harness sends, secret scrubbing, settings files that cannot be committed, brokered credentials
Review cannot keep up with generation (=)	●●●●○	Senior engineers and reviewers	Research and long threads on review as the bottleneck; fragile-codebase complaints [7,8]	Risk-tiered review, policy-file approvals with evidence, tests generated with the change
Do-it-yourself sandboxing (=)	●●●●○	Platform and security teams	Sandboxing moved to Adopt; side channels through shared inference servers [9,10]	Per-task microVMs with network policy as the default, outside-the-sandbox agents
Which subscription works in which harness (up)	●●●●○	Teams running more than one harness	Gemini plans reportedly usable only through Antigravity; terms-of-service worries about using plans elsewhere [11,12]	Clear BYOK paths, cross-harness routers, portable config and skills
Choosing and migrating harnesses (new)	●●●○○	CTOs choosing tools; GitHub Copilot users moving off the deprecated local harness	Benchmarks disagree; angry migration threads after VS Code deprecated its Local Copilot harness [3,13,14,15]	Independent model × harness × cost leaderboards; migration tooling for sessions, skills and config

CC BY 4.0

Pain	Severity	Who feels it	Evidence	Opportunity for tool builders
Context loss and session sprawl (=)	●●●○○	Power users running many sessions	Compaction failures in minimal harnesses; an orchestrator's cloud product shut down after an acquisition [12,16]	Durable, portable session history and memory that survive a tool being acquired

SO WHAT FOR BUILDERS OF HARNESS TOOLS

- Two new pains this month — silent changes and harness migration — both reward independent measurement.
- Products that sit across harnesses (cost, evidence, secrets, sessions) are less exposed to the labs' platform push than single-harness pipelines.

SOURCES

1. Hacker News: OpenAI cuts GPT-6.2 usage limits · Community post
2. Hacker News: comment on organisations clawing back Claude access over billing · Community post
3. Runta: FrontierHarness eval v1.1 · Vendor-reported
4. Hacker News: Is Opus 5.5 nerfed? · Community post
5. Claude Code issue #107757: account email sent to websites · Community post
6. Apiiro: coding agents leaking secrets on GitHub via Don't Ask Again · Independent
7. Hacker News: Harvard researchers find coding agents shifted the bottleneck to code review · Community post
8. Hacker News: AI usage generates a fragile codebase · Community post
9. Thoughtworks Technology Radar Vol. 35: sandboxed execution for AI agents · Analyst
10. Hacker News: models can communicate via inference servers despite being sandboxed · Community post
11. @theo on X: Gemini plans limited to Antigravity · Community post
12. Hacker News: Goodbye Claude, hello Pi · Community post
13. arXiv 2611.32946: ShellBench · Independent
14. r/GithubCopilot: Copilot Harness is totally unusable · Community post
15. GitHub changelog: VS Code retires the Local harness in February 2027 (Nov 16) · Vendor-reported
16. Axios: Meta acquires AI coding startup Conductor · Press report

CC BY 4.0

13 NEW HARNESS STARTUPS

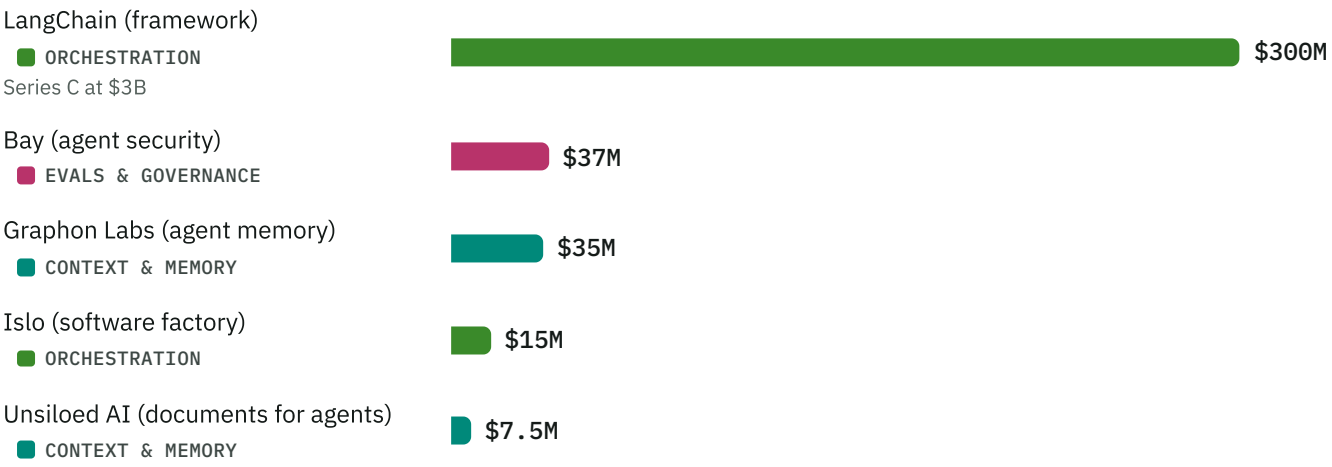
The newcomers sell the whole pipeline, while the money in agent infrastructure went to frameworks, security and memory

Companies added to the directory in November, by layer; funded companies first, then A–Z. Funding is shown only where a public source confirms it.

FIG. 08

One framework round dwarfed everything else in November's agent-infrastructure funding

Disclosed November 2026 rounds for agent-harness infrastructure, verified in the press



Excludes application-layer agents (Sierra, $1.75B), agent evaluation and training data (Mercor, $750M) and Anthropic's IPO. Source [1,2,3,4,5]

The pattern from October holds: among newcomers, disclosed money goes to security and memory, while orchestrators and factories raise modest rounds. The larger capital moved at the edges — Mercor raised $750M at a $20B valuation for agent evaluation and training environments, and Sierra raised $1.75B for customer-service agents. The startups added to our directory this month are mostly software factories and engineering-intelligence tools that now measure or steer coding agents. [1,2,3,6,7,8]

CC BY 4.0

Guild.ai

AGENT CONTROL PLANE AND SOFTWARE FACTORY

"The control plane for AI agents": run, monitor and govern agents from any framework on one model-neutral platform.

2026 · **$44M seed + Series A**[9]

Span

AGENT TRACES AND HARNESS TUNING

Developer intelligence that "links agent traces to code and PRs" and ranks fixes to your agent harness.

2025 · **$25M seed + Series A**[10]

Islo

SOFTWARE FACTORY

"Run Safe AI Software Factories": governed agent workflows (Factory Lines) running in isolated microVMs.

2026 · **$15M**[4]

Vorflux

SPEC-TO-PR AGENT TEAM

Cloud coding agents ("AutoBots") that live in Slack and handle alerts, PR review, broken builds and docs.

2026 · **$15M seed**[11]

HAR

OPEN-SOURCE FACTORY HARNESS

An open, agent-agnostic harness (CLI + MCP) with isolated worktrees and deterministic verification for coding agents.

2026

Linzumi

TEAM CHAT FOR CODING AGENTS

Team chat to command all your company's coding agents, with "a fleet of coding agents inside every channel".

YC P26

Mesmer

SOFTWARE FACTORY WITH FORWARD-DEPLOYED ENGINEERS

"The Agentic Factory for your Engineering Org": finds an engineering team's biggest bottleneck, fixes it in the codebase, then moves to the next.

YC X25

Overcut

GOVERNED AGENT WORKFLOWS

A control plane for an AI software factory: governed, stateful agent workflows across the software lifecycle.

2026

Solario

GOVERNED SOFTWARE FACTORY

"The Governed AI Software Factory": AI builds or transforms software through gated phases, with an audit trail the client keeps.

Paris

Superconductor

PARALLEL CLOUD AGENTS ACROSS HARNESSES

"The multiplayer AI workspace for teams and coding agents": run many Claude Code, Codex and other agents in the cloud.

By Volition

CC BY 4.0

SY

Synth

BUDGETED AGENT SWARMS

Synth Swarms: budgeted agent swarms that build, test and improve software.

YC F24

WE

Weave

ENGINEERING INTELLIGENCE AND MODEL ROUTING

Engineering intelligence that measures engineers and AI agents together, and routes and observes every token.

YC W25

Exits and absorptions since early October[12,13,14,15,16,17]

Company	What happened
Conductor	Acquired by Meta on 4 November; Conductor Cloud shut down on 20 November
HumanLayer	Joined Factory on 4 November; the platform stays up during a transition and is being open-sourced
Mentlio	Joined Factory on 9 October to build token allocation into its platform
Typedef	Acquired by Factory on 30 October for data-dependency context
Augment Code	Sold its Cosmos, Auggie CLI and Code Context Engine assets, and the team behind them, to Harness on 8 October
Pierre	OpenAI agreed to acquire it on 4 November, pending close

> Engineering-intelligence vendors founded earlier — Faros AI, LinearB, Cortex and Activeloop — also joined the directory this month because they now give coding agents context or measure them. They are listed, but not shown as new startups.

SO WHAT FOR INVESTORS

- Acquirers are paying for orchestration teams rather than waiting for orchestrators to win: five completed deals and one pending in under a month.
- Factory startups now compete with lab features; look for the ones that are cross-vendor by design or own a regulated niche.

SOURCES

1. TechCrunch: LangChain hits a $3B valuation · Press report
2. Unite.AI: Bay raises a $37M Series A · Press report
3. The Next Web: Graphon Labs raises a $35M Series A · Press report
4. The Next Web: Islo raises $15M from Insight Partners · Press report
5. TechCrunch: Felicis leads $7.5M in Unsiloed AI · Press report
6. SiliconANGLE: Mercor raises $750M at a $20B valuation · Press report
7. SiliconANGLE: Sierra closes a $1.75B investment · Press report
8. Harness Hub directory: startups · Harness Hub directory
9. Axios Pro: Guild.ai raises $14M seed and $30M Series A (Mar 3 2026) · Press report
10. The AI Insider: Span secures $25M · Press report
11. Inc42: Prasanna Sankar raises $15M for Vorflux · Press report
12. Axios: Meta acquires AI coding startup Conductor · Press report
13. Factory: HumanLayer is joining Factory · Vendor-reported
14. Factory: Factory acquires Mentlio · Vendor-reported
15. Dealroom: Factory acquires Typedef · Press report
16. Harness: Harness acquires Augment Code assets · Vendor-reported
17. OpenAI: OpenAI to acquire Pierre · Vendor-reported

CC BY 4.0

14 PLAYBOOKS

What this means for you

Five readers, five short playbooks for December.

Executives

CARES ABOUT: ROI, COST PREDICTABILITY, VENDOR RISK

- Expect pricing and limits to change more often now that the leading harness vendor is public and the subsidy is shrinking.
- Ask vendors for cost per finished task, not seat prices.
- The software factory is arriving as a feature of tools you already pay for; check those before buying a separate platform.

CTOs & engineering leaders

CARES ABOUT: QUALITY, SECURITY, STANDARDISATION

- Run a harness egress review: what each agent sends to model providers and websites.
- Keep agent settings out of version control and broker credentials from outside the sandbox.
- Pin versions and run your own regression evals after every model or harness update.

Engineers using harnesses

CARES ABOUT: FLOW, LIMITS, RELIABILITY

- Pick the model for the score and the harness for the cost; measure both on your own tasks.
- Keep an open harness configured with an API key as a fallback for limits and outages.
- Feed the agent the first point of failure, not just the final error: it fixes the cause sooner.

Builders of harness tools

CARES ABOUT: GAPS, DEFENSIBILITY, DISTRIBUTION

- Labs now own review and approval; build across vendors, not inside one pipeline.
- The new gaps are silent-change detection, harness migration, cost per task and egress transparency.
- Plan for acquirers: the orchestration layer is consolidating.

Investors

CARES ABOUT: WHERE VALUE ACCRUES, TIMING

- A public Anthropic gives the category its first public-market price signal, and it was volatile within a week.
- Newcomer money still concentrates in security and memory; factory startups raise modest rounds against lab features.
- Exits so far are acqui-hires and asset deals with undisclosed terms.

CC BY 4.0

15 WHAT TO WATCH

October's seven predictions, one month in, and four new ones for the next quarter

October's predictions run to October 2027; the new ones will be scored in the February 2027 edition.

Tracker: October 2026 predictions after one month[1,2,3,4,5,6,7,8,9,10,11,12,13]

Prediction (October)	November evidence	Direction
Metered, token-based billing for heavy users of a flagship coding plan by Q2 2027	GPT-6.2 usage allowance tightened after a launch billing error was corrected; Claude Ship billed per token; Copilot code review billed to organisations by default	Moving toward
An independent model × harness × cost leaderboard cited by 3 of the top 5 harness vendors	Artificial Analysis added harness comparisons; a large vendor-run harness eval published costs	Early signs
Parallel orchestration built into 2 lab harnesses; 3 orchestrator startups shut down or pivot	Conductor acquired and its cloud closed; HumanLayer acquired; Antigravity shipped /swarm parallel agent teams and a model-switching Orchestrator	Moving toward
Sandboxing on by default in at least 2 major harnesses	Sandboxing recommended as the default by Thoughtworks; lab posts on sandbox design; not yet a default change	Early signs
AGENTS.md plus a portable skills format in all 5 most-used harnesses	Skills now ship inside MCP servers; Claude Code Action reads AGENTS.md from the base branch	Moving toward
An agent security or identity company raises $100M or more	Largest disclosed agent-security round this month was $37M	Not yet
An open-weight model in an open harness matches a lab pair on a major coding benchmark	Strong open releases, but benchmark claims are vendor-reported	Not yet

01 At least one more orchestration or software-factory startup is acquired by a lab, a large platform or a harness vendor by the end of February 2027. ●●● HIGH

02 A second lab or major platform, after Anthropic, ships policy-file-based automatic approval of pull requests by the end of February 2027. ●●○ MEDIUM

03 Another US state or federal authority takes legal action over the behaviour of an AI agent by the end of February 2027. ●●○ MEDIUM

CC BY 4.0

04 Anthropic, OpenAI or Google publishes a dated change log for model behaviour changes served to its coding harness by the end of February 2027. ●○○ LOW

SOURCES

1. Hacker News: OpenAI cuts GPT-6.2 usage limits · Community post
2. Claude: introducing Claude Ship · Vendor-reported
3. GitHub changelog: Copilot code review billing defaults to the organization · Vendor-reported
4. Artificial Analysis on X: harness comparisons · Analyst
5. Runta: FrontierHarness eval v1.1 · Vendor-reported
6. Axios: Meta acquires AI coding startup Conductor · Press report
7. Factory: HumanLayer is joining Factory · Vendor-reported
8. Google Antigravity: Gemini 4 Hydrogen and the Orchestrator · Vendor-reported
9. Thoughtworks Technology Radar Vol. 35: sandboxed execution for AI agents · Analyst
10. Claude Code v2.1.332 release (skills from MCP servers) · Vendor-reported
11. Claude Code Action v1.0.277 release (AGENTS.md from the base branch) · Vendor-reported
12. Unite.AI: Bay raises a $37M Series A · Press report
13. Thinking Machines: Inkling 2 Mini · Vendor-reported

16 METHODOLOGY

How we made this report

This second edition follows the method of Volume 1. Public research: arXiv papers, vendor announcements, release notes and press coverage from 1–28 November 2026. The conversation: Hacker News threads, Reddit posts and X posts from the same window, read and grouped into debates; Hacker News quotes were matched word for word against the comment text and X quotes against the post text. The directory: the Harness Hub catalogue of startups, open-source projects, events, news and papers, and the knowledge graph that links them, compared with the October edition.

Claims we could not confirm were left out, and vendor-reported figures are labelled. Reddit's comment text could not be retrieved this month, so Reddit appears through post titles and scores only. Directory totals grew mostly because we widened coverage (skills, plugins, protocol projects), so they describe what we track, not market growth. Data in this edition was read before the end of the month and will be refreshed in the month-end update. Posts about work by the Harness Hub team were excluded from the discussion analysis.

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

Licence: text and charts are CC BY 4.0. Cite as "Harness Hub, State of Harness, Vol. 2, November 2026". Spot an error or a missing harness? Use the submit page and we will correct it in the next edition.