NOW AVAILABLE The 512GB Mac Studio is on sale now, shipping February to March. See every configuration.

THE AI TOOLCHAIN NO. 089

Tail

THE DAILY RELEASE FIREHOSE

PUBLISHED NOVEMBER 10, 2026 · EVERY DAY

The daily firehose — everything the toolchain shipped today, already filtered.

▸ // HOW THIS ISSUE IS MADE SHOW

VIEW

ISSUE VIEW

scroll



SCROLL | EXPAND IN PLACE | PREVIEW PANE

INFORMATION DENSITY

everything



BRIEFED | **EVERYTHING**

Every tool in the issue, in full.

all 345 changes

Do you prefer **scroll**? 👍 👎

▸ // **CONTENTS** SHOW 73 of 76 tools

▾ // THE BRIEF HIDE SAME ISSUE, SAME PROMPT, TWO WRITERS: **Opus 5** | GPT-5.5

What stands out across today's releases, grouped by what it lets you do. Every tool named links to its entry below.

vim ? help

Three unrelated things stand out today. GPT Researcher now researches report sections in parallel, dropping a detailed report from roughly 9:11 to 3:27. browser-use 1.0 rebuilds its agent as a persistent Python REPL over raw CDP with typed structured output and secrets handling. Atomic Agent runs tasks overnight unattended under spend and time budgets with risk-scored approval rules.

BUILD

Get a long research report back in a coffee break instead of a lunch break

GPT Researcher parallelizes per-section research, cutting detailed report generation from roughly 9:11 to 3:27 and adding a lifecycle observer so you can watch where time goes rather than staring at a blank terminal. HyperResearch collapses semantic search into a single numpy matmul and adds a keyless Keenable web search provider, so a local research loop runs without provisioning an API key first.

GPT Researcher · HyperResearch

BUILD

Drive a real browser from an agent and get a typed object back, not a page of prose

browser-use 1.0 runs the agent as a persistent Python REPL over raw CDP with typed structured output, file upload/download, domain restrictions and secrets handling — so the agent's result parses instead of needing a second model call to extract it, and credentials never land in the transcript. playwright-mcp adds a background mode for extension-connected browsers and strict visible-only locator matching, which kills the class of failures where an agent clicks a hidden duplicate element.

browser-use · playwright-mcp

GOVERN

Let an agent work overnight without waking up to an unbounded bill

Atomic Agent's CLI and desktop both run unattended tasks against explicit spend and time budgets, with persistent approval rules and risk-scored smart approval deciding what proceeds without a human — the babysitting shift goes away, and a runaway loop stops at a ceiling rather than at your invoice. Copilot CLI adds Windows sandbox filesystem permission grants and tighter MCP tool-deny and consent handling for the same reason on the other side: narrowing what a tool call is allowed to touch.

> atomic-agent · GitHub Copilot CLI

Does **Opus 5** read better? 👍 👎

BUILD

◆ Model Training & Fine-Tuning

Thinking Machines Lab Tinker

Feedback ⓘ

changelog · Nov 10, 2026 | CHANGELOG ↗ | RANK

Tinker provides a managed API for training and sampling language models while users control their training algorithms and data.

> Tinker's SDK 0.38.0 introduces a new APIFuture return type that lets client methods be awaited in async code or used synchronously via .future(), alongside several related deprecations of older async and capability APIs.

└──▷ WHAT SHIPPED · 4 FEATURES most completely described first ? what's the number?

01 **APIFuture type for async/sync concurrency** NEW 95

All client methods now return the new APIFuture type, which can be awaited directly in async code or converted with .future() into a standard `concurrent.futures.Future` for use with `concurrent.futures.wait`, `as_completed`, and `add_done_callback`. Calling APIFuture.coroutine() lets the result be passed directly to coroutine-accepting APIs such as `asyncio.create_task`.

Run a training job from sync code and block until complete using the new APIFuture return type.

python

```
future =
training_client.create_lora_training_client(...).future()
concurrent.futures.wait([future])
client = future.result()
```

Schedule a sampling client creation as an asyncio task using APIFuture.coroutine().

python

```
task =
asyncio.create_task(training_client.create_sampling_client(...).
coroutine())
sampling_client = await task
```

— Runnable sync and async code examples included changelog-20261110-4fbee5f1

02 **Client-creation and RestClient/AdminClient methods return APIFuture**

80

BREAKING

Client-creation methods `create_lora_training_client`, `create_training_client_from_state`, `create_training_client_from_state_with_optimizer`, `create_sampling_client`, and `save_weights_and_get_sampling_client` now return an `APIFuture` of the client instead of the client directly (using the returned value directly still works but is deprecated). `RestClient` and `AdminClient` methods likewise now return an `APIFuture` instead of a `concurrent.futures.Future`; call .future() to continue using `Future` methods such as .done() and .add_done_callback().

— Named methods and migration path given, no runnable sample `changelog-20261110-4fbee5f1`

---

03 **Deprecated async methods and ServiceClient capability call**

DEPRECATED

60

The `_async` methods on `TrainingClient` are deprecated and will be removed in a subsequent release; users should migrate to the `APIFuture` `-based` equivalents. ServiceClient.get_server_capabilities() is deprecated in favor of RestClient.get_server_capabilities().

— Names deprecated and replacement methods, no example `changelog-20261110-4fbee5f1`

THINNER COVERAGE BELOW

04 **Immediate return for client info methods** IMPROVED

50

TrainingClient.get_info() and SamplingClient.get_base_model() now return immediately without making a network call.

— Clear behavior change but no usage example `changelog-20261110-4fbee5f1`

└──▷ BREAKING ON UPGRADE

! The `_async` methods on `TrainingClient` are deprecated and will be removed in a subsequent release; migrate to the `APIFuture` `-based` equivalents.

! Client-creation methods on `TrainingClient` and `save_weights_and_get_sampling_client` now return an `APIFuture` of the client rather than the client directly; using the returned value directly as the client still works but is deprecated.

! `RestClient` and `AdminClient` methods now return an `APIFuture` instead of a `concurrent.futures.Future`; call .future() to continue using `Future` methods such as .done() and .add_done_callback().

! ServiceClient.get_server_capabilities() is deprecated in favor of RestClient.get_server_capabilities().

WAS THIS USEFUL? 👍 👎

◆ AI Coding Agents

yetone magpie

Feedback ⓘ v0.1.3557 CODE ↗ RANK

Magpie is a macOS menu-bar app, TUI, and CLI that lists every AI coding agent on the machine and switches each one's model, such as Codex on DeepSeek or Claude Code on Kimi, by rewriting only that key in the agent's own config.

> This window magpie shipped gateway-level security and reliability work — mutual TLS for remote connections, Docker-option sandboxing for MCP registry servers, and Antigravity multi-account failover — alongside a `magpie data move` command, bulk memory management, Codex service-tier pinning, a webdav/s3 follow-server sync mode, and expanded Hermes/billion-context and agent-attribution routing.

└──▷ WHAT SHIPPED · 15 FEATURES most completely described first ? what's the number?

01 **Mutual TLS for the gateway** NEW 95

Enable mTLS on the gateway with `magpie mtls on` or via Settings → Network and sharing → Require a client certificate, so only devices holding a CA-issued client certificate can connect remotely. Issue per-device client certificates with `magpie mtls issue <name>`, exported as PEM files alongside the CA certificate for use by Remote magpie (`provider.ClientCert` , `ClientKey` , `ServerCA`) or curl (`--cert` , `--key` , `--cacert`).

Turn on mTLS so that the gateway refuses any remote connection that cannot present a valid client certificate.

```
$ magpie mtls on
```

Issue a client certificate for a new device so it can authenticate to the mTLS-protected gateway.

```
$ magpie mtls issue laptop-alice
```

— Runnable commands and exact fields named v0.1.3533

02 **Relocate config/cache with magpie data move** NEW 95

Move magpie's config and cache directories to another disk with `magpie data move <folder>` , which leaves a symlink (or Windows junction) at the original path so all existing agent configs, skill links, and MCP server commands continue to work without changes. Move files back to the default location with `magpie data move --home` ,

which copies the files home and removes the link. Inspect current file locations, move targets, and pending or failed move status with `magpie data` or `magpie data where`.

Relocate magpie's data to an external SSD so the main disk stays uncluttered — run while magpie is not serving the gateway for an immediate move.

```
$ magpie data move /Volumes/ExternalSSD
```

Check where magpie currently stores its files and whether a move is pending or previously failed.

```
$ magpie data where
```

Move config and cache back to the default on-disk location and remove the external-disk link.

```
$ magpie data move --home
```

— Runnable commands for move, inspect, restore v0.1.3529

03 **Hermes billion-context proxy routing** NEW 90

Route Hermes's magpie providers through a local proxy by calling HermesThrough(a, prefix) with a prefix such as `http://127.0.0.1:8787/bili/`, rewriting `base_url` in every bot config that has a magpie entry — bypassing the limitation that billion-context's own `HTTPS_PROXY` `-based` Hermes plugin does not reach magpie's `http://127.0.0.1` provider. Read the current `base_url` of magpie's provider in Hermes's connected config with HermesURL(a).

— Named functions and exact prefix given v0.1.3548

04 **WebDAV/S3 follow-server sync mode** NEW 90

Take a machine's configuration from the server without pushing local changes back using `magpie webdav follow on|off` or `magpie s3 follow on|off`; local changes made while following are replaced and a copy is kept in the sync folder. The `FollowServer` field (`json:"followServer"`) in the sync `Config` struct and the `Followed bool` field (`json:"followed"`) in `Notice` expose follow-server state and per-sync replacement events programmatically.

Pin a secondary machine to always receive configuration from the WebDAV server, so local edits never drift the shared setup.

```
$ magpie webdav follow on
```

Restore two-way sync on a machine that was previously set to follow the server.

```
$ magpie s3 follow off
```

— Runnable commands and config fields named `v0.1.3539`

05 **Docker run safety filtering for MCP registry** BREAKING

85

magpie now strips dangerous `docker run` options — including `--network host`, `--privileged`, `--pid host`, and host path bind mounts — from OCI package `runtimeArguments` supplied by MCP registry publishers, and skips any server whose required options cannot be made safe. Safe options — `-i`, `--rm`, `--init`, `--read-only`, `-e` / `--env`, `--entrypoint`, `-w` / `--workdir`, `-u` / `--user`, `--platform`, `--name`, `--pull`, and named or anonymous volumes — still pass through. MCP registry servers whose required `runtimeArguments` cannot be made safe are no longer offered or launchable via the market.

— Exact flags named, breaking change noted `v0.1.3547`

06 **Bulk memory status management** NEW

85

Bulk-act on multiple memories at once — keep, archive, restore, or delete — using the new `SetStatus` function backed by `/api/memory/status` and `/api/memory/remove` endpoints, with Shift-click range selection in the Memory tab UI.

Bulk-archive several memories by ID in one API call rather than editing each one individually.

```
$ curl -s -X POST http://localhost:<port>/api/memory/status \
  -H 'Content-Type: application/json' \
  -d '{"status":"archived","ids":["<id1>","<id2>","<id3>"]}'
```

— Named endpoints with runnable example `v0.1.3545`

07 **Full backlog memory distillation** IMPROVED

85

Read an entire backlog of conversations at once — not just the hourly cap — using the new `ReadAll` function, with live progress via Progress() and a stop button that calls `POST /api/memory/distill/stop`. The hourly auto-run cap `PerRun` is raised from 10

to 30 sessions, so a moderately busy day's sessions clear in one pass rather than queuing up.

Stop a long-running backlog distill mid-flight and see how many sessions remain unread.

```
$ curl -s -X POST http://localhost:
<port>/api/memory/distill/stop -d '{}'
```

— Named function, endpoint, and limit change v0.1.3531

08 **Codex service tier setting** NEW 80

Set a Codex service tier in magpie's settings via `codexServiceTierRow` to pin every Codex request to standard, Fast (`priority`), or Ultrafast (`ultrafast`), regardless of what Codex itself requests — works through both the `/v1/responses` provider route and Codex's signed-in ChatGPT backend route, with Ultrafast automatically downgraded to `priority` when no account on the plan supports it.

— Named setting and values, no CLI command v0.1.3537

09 **Agent-path routing for OpenBot** NEW 75

Route requests through gateway paths prefixed with `AgentPath` (`/agent/`) so agents like OpenBot that cannot supply their own key or User-Agent header are correctly identified — requests to `<gateway>/agent/openbot/v1/chat/completions` are served as `/v1/chat/completions` attributed to OpenBot rather than to the OpenCode process it wraps. Adds AgentURL(gw, id string) helper to construct the agent-specific gateway address.

— Exact path pattern and helper named v0.1.3544

10 **OpenCode 2.x config format support** IMPROVED 70

magpie now detects when the OpenCode on PATH is version 2.x and writes its config using OpenCode 2's schema — provider under `providers` and small model under `agents.title.model` — instead of the OpenCode 1 layout that caused a WARN per model on every load.

— Exact schema keys, automatic behavior only v0.1.3536

11 **Billion-context token savings display** NEW 65

The billion-context library card now queries the running proxy at `GET /__bili/overview` to display net and gross tokens saved across all sessions on the machine, falling back to a cached `bili-saved.json` when the proxy is not running.

— Named endpoint and cache file, no UI path given v0.1.3543

12 **Antigravity 2.30 config file support** IMPROVED 65

magpie now reads and writes the Antigravity gateway block to both `~/.gemini/config/config.json` (used by Antigravity 2.30 and later) and `~/.antigravity/settings.json` (used up to 2.29), so users on either app version stay connected.

— Exact file paths named, no command v0.1.3532

13 **Allowances tab drag-to-reorder** NEW 60

The tray panel's Allowances tab gains an Arrange mode letting you drag (or use Alt+arrows) to reorder subscriptions and hide them from the panel, without affecting routing, the Usage page, or the menu bar.

— UI path given, no API surface v0.1.3552

THINNER COVERAGE BELOW

14 **Antigravity multi-account failover** IMPROVED 50

magpie's gateway now cycles through all active Antigravity accounts when one is quota-exhausted, promoting the account with remaining quota to first place on the next turn.

— Behavior only, no flag or config key named v0.1.3552

15 **Settings hint cards extended to provider editor** IMPROVED 40

Settings hint cards in the GUI now also appear on fields in the provider editor and on Routing group rows (`.hint-row`), not only on the Settings page.

— Thin UI-only description v0.1.3546

NEW INTEGRATION TinyFish Search API

NEW INTEGRATION miniQ agent integration

NEW INTEGRATION mu (mu-agent) agent integration

└──▷ BREAKING ON UPGRADE

! MCP registry servers whose `runtimeArguments` include a required option that escapes the container (such as `--network host`, `--privileged`, a host path bind mount, `--pid host`, or a host env file) are no longer offered or launchable via the market.

v0.1.3557 CODE ↗ LATEST STABLE

└──▷ GET THIS VERSION ▾

Docker

```
$ docker pull ghcr.io/yetone/magpie:0.1.3557
```

```git
$ git clone --branch v0.1.3557 https://github.com/yetone/magpie.git
# already have the repo? check out this version:
$ git checkout v0.1.3557
```

published under the project's GitHub account · checked 10 Nov Wrong command?

v0.1.3556 CODE ↗ STABLE

magpie adds a miniQ integration, routing the miniQ desktop AI coworker and terminal client through magpie's gateway via settings.json or miniq configure.

└──▷ GET THIS VERSION ▾

```Docker
$ docker pull ghcr.io/yetone/magpie:0.1.3556
```

```git
$ git clone --branch v0.1.3556 https://github.com/yetone/magpie.git
# already have the repo? check out this version:
$ git checkout v0.1.3556
```

published under the project's GitHub account · checked 10 Nov Wrong command?

Configure miniQ's provider via `miniq configure --base-url` or `settings.json`'s `provider` object (fields: `baseUrl`, `apiKey`, `model`, `apiProtocol`, `reasoningEffort`) to point miniQ at magpie's gateway; magpie stashes and restores the original provider when it steps out.

magpie now supports miniQ as an agent integration, writing the gateway's `/v1` endpoint as miniQ's provider and updating the running daemon live via the `settings.update` RPC over the local WebSocket named in `daemon.json`.

miniQ's `settings.json` data folder is resolved from `$MINIQ_DATA_DIR`, then `miniq-data-dir.json` beside the default, then `%LOCALAPPDATA%\miniq` on Windows or `~/.local/share/miniq` on Linux and macOS.

v0.1.3550 CODE ↗ STABLE

magpie adds mu (mu-agent) as a supported agent, wiring config in `~/.mu/agent` alongside the user's own provider settings.

└──▷ GET THIS VERSION ▾

Docker

```
$ docker pull ghcr.io/yetone/magpie:0.1.3550
```

git

```
$ git clone --branch v0.1.3550 https://github.com/yetone/magpie.git
# already have the repo? check out this version:
$ git checkout v0.1.3550
```

published under the project's GitHub account · checked 10 Nov Wrong command?

WAS THIS USEFUL? 👍 👎

zot Feedback ⓘ v0.4.72 NOTES ↗

zot manages coding agent workflows from the command line.

zot adds `zot auth set <provider>` and `--stdin` for headless API-key login without starting the TUI, saving keys to `$ZOT_HOME/auth.json`.

└──▷ GET THIS VERSION ▾

```
$ git clone --branch v0.4.72 https://github.com/patriceckhart/zot.git
# already have the repo? check out this version:
$ git checkout v0.4.72
```

NEW SUPPORTED PLATFORM Claude Fable 5.2 provider support

WAS THIS USEFUL? 👍 👎

Letta Code

Feedback ⓘ v0.34.50 NOTES ↗ RANK

Letta Code is a terminal coding agent with persistent memory and identity across development sessions.

> Letta Code adds App Server registration of remote computers and workflow cost ceilings, alongside durable MCP connection management, cross-agent messaging on the local backend, WatchPR snapshot fetching via Letta Cloud, and listener-based tool execution controls.

└──▷ WHAT SHIPPED · 7 FEATURES most completely described first ? what's the number?

01 **App Server registration as named computer** NEW 85

Register the App Server as a named computer with the `--listen` flag so agents can route work to it. Run `letta server --listen` to register this machine as a routable computer so cloud agents can dispatch work to it via the App Server.

Register this machine as a routable computer so cloud agents can dispatch work to it via the App Server.

```
$ letta server --listen
```

— Names exact flag and runnable command v0.34.49

02 **USD cost ceiling for workflows** NEW 70

Cap workflow spending by setting `maxCostUsd` to enforce a USD cost ceiling before a run is allowed to proceed.

— Names config key but no file path or command v0.34.49

THINNER COVERAGE BELOW

03 **SendAgentMessage on local backend** NEW 50

Adds `SendAgentMessage` support on the local backend so agents can message each other without a cloud connection.

— Named API but no usage shown v0.34.50

04 **WatchPR snapshot fetching via Letta Cloud** NEW 50

Fetches WatchPR snapshots through Letta Cloud, enabling pull-request monitoring for cloud-connected agents.

— Clear capability, no concrete usage steps v0.34.50

05 **Listener controls for tool policy and cancellation** NEW

50

Adds per-conversation foreground yield control in `tool_policy` via the listener, and adds cancellation of external tools on capable controllers via the listener.

— Names config field but mechanism thin v0.34.50

06 **Ephemeral subagent conversations** NEW 35

Run fresh subagents in ephemeral conversations so each invocation starts from a clean state without persisting history.

— Behavior described but no config surface named v0.34.49

07 **Durable MCP connection event management** NEW 30

Adds durable MCP connection event management in the skills layer.

— Bare description, no mechanism or usage given v0.34.50

WAS THIS USEFUL? 👍 👎

Alibaba OpenCodeReview

Feedback ⓘ v1.13.7 NOTES ↗

OpenCodeReview is a secure, fast, efficient, battle-tested at Alibaba's scale. Hybrid architecture code review tool: deterministic pipelines + LLM Agent, precise line-level comments, built-in multi-language ruleset (NPE, thread-safety, XSS, SQL injection), OpenAI & Anthropic compatible.

OpenCodeReview adds session inspection by ID from any directory and gates `PLAN_TASK` on file size.

└─▷ GET THIS VERSION ▾

npm

```
$ npm install -g @alibaba-group/open-code-review@1.13.7
```

git

```
$ git clone --branch v1.13.7 https://github.com/alibaba/open-code-review.git
# already have the repo? check out this version:
$ git checkout v1.13.7
```

signed build provenance · checked 10 Nov Wrong command?

Inspect a saved session by ID from any working directory using the session inspection capability added to `session`.

Gate `PLAN_TASK` execution on file size to prevent oversized files from being processed.

NEW SUPPORTED PLATFORM Clojure and ClojureScript review rules

WAS THIS USEFUL? 👍 👎

All Hands AI OpenHands

Feedback ⓘ v1.38.0 NOTES ↗

RANK

OpenHands is an open-source software-development agent that writes code, runs commands, and browses the web.

> OpenHands' v1.38.0 release adds an automation-aware install mode with a new plugin-spec registration endpoint, lighter-weight launch links, and several UI refinements including conversation budget meters and an Enterprise license banner.

└──▷ WHAT SHIPPED · 7 FEATURES most completely described first ? what's the number?

01 **Automation-aware install and plugin spec registration** NEW — 80

Trigger automation-aware install by passing `mode=automation` to the launch flow, and register plugin specs via the new `POST /plugin-spec` endpoint.

– Names exact flag and endpoint; mechanism partly explained v1.38.0

THINNER COVERAGE BELOW

02 **Launch links without plugin requirement** NEW — 45

Create `/launch` links that carry only a message, with no plugin required.

– Names the surface but limited mechanism detail v1.38.0

03 **Plugin configuration editing in automation help panel** IMPROVED — 30

Edit executable plugin configuration directly inside the automation help panel.

– One-line addition with no further mechanism v1.38.0

04 **Agent profile presets for automations** NEW — 30

Run prompt and plugin presets from an agent profile in automations.

– Brief description, no config or API named v1.38.0

05 **Conversation budget meters in UI** NEW — 30

OpenHands now shows conversation budgets with semantic meter fills in the UI.

– Describes visual change but no exact UI path or config v1.38.0

06 **Enterprise license banner in Canvas** NEW — 30

OpenHands shows an Enterprise license banner in Canvas for licensed instances.

— Named feature but minimal operational detail v1.38.0

07 **Paper theme for Agent Canvas** NEW 15

OpenHands adds a Paper theme for Agent Canvas.

— Only a bare mention, no further detail provided v1.38.0

WAS THIS USEFUL? 👍 👎

The Open Engine Zeroshot

Feedback ⓘ v13.1.0 NOTES ↗

RANK

Zeroshot orchestrates executor-verifier agent loops that implement and validate software changes while logging workflow state to SQLite.

> Zeroshot v13.1.0 adds authenticated export downloads for retained hosted workspaces and checkpoints, streaming them into private tar files with atomic publication.

└──▷ WHAT SHIPPED · 1 FEATURE most completely described first ? what's the number?

01 **Export downloads for hosted workspaces** NEW 63

Download retained hosted workspaces and checkpoints from hosted targets that advertise export, streaming into private tar files and publishing only after a complete transfer, with cleanup on failure or cancellation.

— Explains mechanism (streaming, atomic publish, cleanup) but no CLI/API name given.

v13.1.0

NEW INTEGRATION OpenCode Go and OpenCode Zen providers

NEW INTEGRATION Plugin distribution for Codex/ChatGPT and Claude via repository marketplace

v13.0.0 NOTES ↗ STABLE

Zeroshot adds opt-in GitHub token forwarding via `--forward-gh-token`, co-author credits on delivered commits, and profile runtime exposure.

└──▷ GET THIS VERSION ▾

npm

```
$ npm install -g @the-open-engine-company/zeroshot@13.0.0
```

git

```
$ git clone --branch v13.0.0 https://github.com/the-open-engine/zeroshot.git
# already have the repo? check out this version:
$ git checkout v13.0.0
```

signed build provenance · checked 10 Nov Wrong command?

Control whether the shell `GH_TOKEN` is forwarded to named targets on `run`, `resume`, and `stack submit` with the new `--forward-gh-token` flag — previously the token was always sent silently.

Credit a co-author on delivered commits by setting `TargetRunRequest.commitCoAuthor` on submission or `RunResumeParams.commitCoAuthor` on resume, using the same `{name, email}` `CommitIdentity` type as `commitIdentity`.

Built-in profile templates now expose their authored runtime and derived connection requirements so clients can configure profiles without duplicating provider rules.

└─▷ BREAKING ON UPGRADE

! `GH_TOKEN` from the invoking shell is no longer forwarded automatically to named targets on `run`, `resume`, and `stack submit`. Add `--forward-gh-token` to restore the previous behaviour; without it, the ambient shell token is not sent.

WAS THIS USEFUL? 👍 👎

corecoder

Feedback ⓘ v0.10.0 NOTES ↗ RANK

Corecoder is a minimal AI coding agent (~1,000 lines of Python) inspired by Claude Code. Works with any LLM.

> Corecoder adds bounded dependency waves to agent teams, letting workers declare `depends_on` to sequence execution and receive upstream results automatically.

└──▷ **WHAT SHIPPED · 1 FEATURE** most completely described first ? what's the number?

01 **Dependency sequencing for agent team workers** NEW 72

Workers can declare `depends_on` in their definition so they run only after named dependencies succeed, receiving those dependencies' reports in their own prompt. Failed dependencies skip only their downstream branch rather than halting the entire team, and invalid dependency names or cycles are rejected before any worker starts.

Run the offline dependent-pipeline demo to see `depends_on` worker sequencing without needing an API key.

```
$ corecoder --demo
```

— Names config key and behavior, includes runnable demo command. v0.10.0

WAS THIS USEFUL? 👍 👎

Anthropic Claude Code

Feedback ⓘ v2.1.322 NOTES ↗

CODE ↗ RANK

Claude Code is Anthropic's terminal coding agent that plans, edits, and tests code in local repositories.

> Claude Code sped up Vertex AI inference, added a settings-lock flag for remote-control sessions, and switched its default Fable model to a larger, cheaper context option.

└──▷ WHAT SHIPPED · 3 FEATURES most completely described first ? what's the number?

01 **Settings file lock for remote-control** NEW 85

A new `--settings <file>` flag on `claude remote-control` applies a settings file to every session it starts; sessions will not start if the file changes until the daemon is restarted.

Lock all remote-control sessions to a shared settings file so policy changes take effect only after a deliberate restart of the daemon.

```
$ claude remote-control --settings /etc/claude/team-settings.json
```

— Exact flag, command, and restart behavior given v2.1.321

02 **Claude Fable 5.2 as default model** IMPROVED 70

Claude Code now defaults to Claude Fable 5.2 (`claude-fable-5-2`), offering 1M context at $8/$40 per Mtok with $0.20/Mtok cache reads.

— Names model id, context size, and pricing precisely v2.1.321

THINNER COVERAGE BELOW

03 **Fast mode for Vertex AI regions** IMPROVED 50

Claude Code now uses fast mode on Vertex AI in the `global` , `us` , and `eu` regions, automatically falling back to standard speed if Vertex refuses the request.

— Names regions and fallback behavior but no config or metrics v2.1.322

WAS THIS USEFUL? 👍 👎

Earendil Works Pi

Feedback ⓘ v1.5.1 NOTES ↗ RANK

Pi is an AI coding agent toolkit that provides a unified LLM API, agent loop, TUI interface and CLI.

> Pi v1.5.1 adds shell-based provider authentication, new models.json compatibility controls, a GitHub Copilot classifier, a new Qwen China provider and Claude model, plus several breaking changes to tool-call handling, grep search, and default model selection.

└──▷ WHAT SHIPPED · 11 FEATURES most completely described first ? what's the number?

01 New compatibility options in models.json NEW

90

`models.json` gains `compat.promptCacheKeyMode` (`"auto"`, `"enabled"`, or `"disabled"`) to control prompt cache key behavior per model, `compat.supportsMessageThreads` to control Anthropic message-thread support per model (on by default for the `anthropic` provider on `api.anthropic.com`), and `compat.requiresNonNullAssistantContent` to handle OpenAI-compatible endpoints that require an empty string instead of `null` in assistant messages.

Enable prompt cache keys for a proxy that supports them, or disable them for endpoints that reject the header, without changing global behavior.

json

```
{
  "models": [
    {
      "provider": "my-proxy",
      "compat": {
        "promptCacheKeyMode": "enabled"
      }
    }
  ]
}
```

— Three config keys fully named with defaults and example v1.5.1

02 Omitted empty tools array for OpenAI-compatible models BREAKING

85

OpenAI-compatible models no longer send `"tools": []` when all tools are disabled in a session that has used tools — the field is omitted instead. Proxies that require `tools` once history has tool calls can set `compat.requiresToolsWithToolHistory: true` for that model in `models.json`.

— Clear before/after behavior with named config workaround v1.5.1

03 **Shell auth commands for providers** NEW 80

`pi auth login <provider>` and `pi auth logout <provider>` let you log in and out of providers from the shell without starting a session, with `--api-key` to save an API key or `--no-browser` for device-code sign-in.

Save an API key for a provider from the shell without opening Pi, so CI or scripts can authenticate non-interactively.

```
$ pi auth login anthropic --api-key
```

— Runnable command with flags and example shown v1.5.1

04 **Scoped Markdown transformers by message type** IMPROVED 75

pi.registerMarkdownTransformer() now accepts an optional `{ messageTypes }` argument to target specific Markdown surfaces for a transformer, including `custom`, `summary`, and `changelog`.

Opt a custom Markdown transformer into summary surfaces so it also processes branch and compaction summaries.

```
$ pi.registerMarkdownTransformer(myTransformer, { messageTypes:
['summary'] });
```

— Named API argument with concrete example v1.5.1

05 **New model and provider additions** NEW 75

Adds the `qwen-token-plan-individual-cn` provider for Qwen Token Plan Individual on the China endpoint, authenticated with `QWEN_TOKEN_PLAN_CN_API_KEY`, and the `claude-fable-5-2` model to the `anthropic` provider at 8/40 USD per million input/output tokens, 1M context, 128K output, with per-message effort and server-side Claude Opus 5 fallback.

— Provider and env var named but no usage shown Detailed pricing and specs but no usage example

v1.5.1

06 **Extension update reporting in pi update** IMPROVED 70

`pi update --extensions` reports each updated package with its npm version or git commit change, with distinct messages for unchanged, skipped, and warned updates.

— Named flag with behavior detail, no example run v1.5.1

07 **Deprecated GrepOperations.readFile callback** DEPRECATED 70

Custom `GrepOperations.readFile` callbacks are no longer called for context lines; the field is now optional and deprecated. To search a remote system, override grep's execute() instead.

— Names field, deprecation path, and replacement method v1.5.1

THINNER COVERAGE BELOW

08 **Mermaid diagram rendering improvements** IMPROVED 50

Mermaid diagrams now support pie charts, author-defined styles, clickable labels on terminals with hyperlink support, and rendering inside expanded branch and compaction summaries.

— Lists features but no mechanism or example v1.5.1

09 **New default Anthropic model: claude-opus-5-5** BREAKING 50

The default model for the `anthropic` provider is now `claude-opus-5-5` instead of `claude-opus-4-8`.

— States the change but no migration guidance v1.5.1

10 **GitHub Copilot classifier via codemode** NEW 45

Pi can call GitHub Copilot's Microsoft-Decision-1 classifier for typed decisions and probabilities via models.classify() in codemode.

— Names the API but no usage example given v1.5.1

11 **Syntax highlighting for Dart, Nix, Groovy, Gradle** NEW 45

Adds syntax highlighting for Dart, Nix, Groovy, and Gradle files in read and write views.

— Simple named addition with no further detail v1.5.1

└──▷ BREAKING ON UPGRADE

! Custom `GrepOperations.readFile` callbacks are no longer called for context lines; the field is accepted as optional and deprecated. To search a remote system, override grep's execute() instead.

! OpenAI-compatible models no longer send `"tools": []` when all tools are disabled in a session that has used tools — the field is omitted. If your proxy requires `tools` once the history has tool calls, set `compat.requiresToolsWithToolHistory: true` for that model in `models.json`.

! The default model for the `anthropic` provider is now `claude-opus-5-5` instead of `claude-opus-4-8`.

WAS THIS USEFUL? 👍 👎

wrongstack

Feedback ⓘ v1.5.5 NOTES ↗ RANK

WrongStack is an AI coding agent that reads code, edits files, runs commands, and reasons through bugs across CLI, TUI, and browser interfaces.

> WrongStack v1.5.5 lets users configure model, mode, and effort per ACP agent and per session across all its interfaces, and fixes a write-authorization bug affecting symlinked or short-path project roots.

└──▷ **WHAT SHIPPED · 2 FEATURES** most completely described first ? what's the number?

01 **Per-agent, per-session ACP runtime settings** NEW 62

Model, mode, and effort options can now be saved per ACP agent and per session from the CLI, WebUI, VibeUI, HQ, and the desktop app. Pickers and monitors now badge ACP agents to reflect these settings.

— Names all surfaces but no exact UI steps or keys v1.5.5

THINNER COVERAGE BELOW

02 **Fixed ACP write authorization for symlinked paths** IMPROVED 45

Edits inside a symlinked project root or a Windows short path no longer incorrectly fail ACP write authorization as outside-project writes.

— Describes the bug and fix but no reproduction steps v1.5.5

WAS THIS USEFUL? 👍 👎

Block Goose

Feedback ⓘ v1.58.0 NOTES ↗ CODE ↗ RANK

Goose is an open-source AI agent with a desktop app and CLI for coding and other automated tasks.

> Goose expanded its provider ecosystem with a new local Lemonade provider plus two more declarative routing providers, and added context window visibility, an opt-in relevance filter, and a quieter default for thinking blocks in the desktop app.

└──▷ WHAT SHIPPED · 5 FEATURES most completely described first ? what's the number?

01 **Lemonade local AI provider** NEW 90

Goose now supports a Lemonade local AI provider, configured via `LEMONADE_HOST` and optional `LEMONADE_API_KEY`, connecting by default to `http://localhost:13305` with dynamic model discovery via `GET /v1/models`.

Connect Goose to a local Lemonade runtime running on a non-default host, without requiring an API key.

```
$ export LEMONADE_HOST=http://192.168.1.50:13305
```

— Named env vars, default host, and discovery endpoint with a runnable example v1.58.0

THINNER COVERAGE BELOW

02 **AIHubMix and RouterPlex declarative providers** NEW 35

Goose adds an AIHubMix declarative provider for routing requests through the AIHubMix inference service, and a RouterPlex declarative provider for routing requests through the RouterPlex inference service.

— Names two providers but no config detail or mechanism given v1.58.0

03 **Context window usage report** NEW 35

Goose adds a context window report and viewer so practitioners can inspect how the context window is being used during a session.

— Describes the capability but no exact UI path or format v1.58.0

04 **Opt-in context relevance filtering** NEW 35

Goose adds an opt-in context relevance operation to trim or filter context based on relevance before sending it to the model.

— Explains purpose but no config key or activation detail v1.58.0

05 **Collapsed thinking blocks by default** IMPROVED 25

The desktop app now collapses thinking blocks by default, with a setting to override the behavior.

— States the change but no setting name or location `v1.58.0`

WAS THIS USEFUL? 👍 👎

GitHub Copilot CLI

Feedback ⓘ v1.0.108 NOTES ↗ RANK

GitHub Copilot CLI is a terminal-based coding agent that can plan, write, and test code in local repositories.

> GitHub Copilot CLI v1.0.108 adds Windows sandbox filesystem permission grants, new `/index`, `/limits unset`, and `/changelog <version>` commands, and tightens MCP tool-deny and consent handling.

└──▷ WHAT SHIPPED · 10 FEATURES most completely described first ? what's the number?

01 Windows sandbox filesystem enumerate grants NEW

80

Add `sandbox.userPolicy.filesystem.enumeratePaths` in `settings.json` to grant Windows sandbox sessions enumerate-only directory permissions, without allowing reads or writes to those paths.

Grant a Windows sandbox session enumerate-only access to a directory without allowing reads or writes.

json

```
{
  "sandbox": {
    "userPolicy": {
      "filesystem": {
        "enumeratePaths": ["C:\\Projects\\secrets"]
      }
    }
  }
}
```

— Exact config key and file with a working example v1.0.108

02 MCP deny-rule and consent handling improvements IMPROVED

80

MCP deny rules now block a tool by its advertised name (for example server(weather.lookup)) in addition to its sanitized name, even with `--allow-all-tools`; interactive CLI sessions now pass local MCP server environment values literally, using `${NAME}` to reference a host environment variable; `/mcp enable` now respects repository settings that disable an MCP server; and Interactive Autopilot now waits for your response to MCP input and consent requests instead of automatically declining them.

— Multiple named flags and syntax, no direct example v1.0.108

03 **Clear session AI credit caps** NEW 65

`/limits unset` clears session AI credit caps for subsequent requests mid-session.

Remove a session AI credit cap mid-session so subsequent requests are no longer rate-limited by the cap.

```
$ /limits unset
```

— Exact command with runnable usage example v1.0.108

04 **Retrieve release notes by version** NEW 65

`/changelog <version>` retrieves release notes for any version, including those not bundled with the installed CLI.

Pull up release notes for a specific older version without needing that build installed.

```
$ /changelog v1.0.100
```

— Exact command with runnable usage example v1.0.108

THINNER COVERAGE BELOW

05 **/index commands and managed tgrep indexing** NEW 55

Adds `/index` commands and enables managed shared tgrep indexing by default for eligible repositories.

— Command named but no example or eligibility detail v1.0.108

06 **Session lifecycle events in JSON output** IMPROVED 50

`-p --output-format json` output now includes session start, resume, and model-change events.

— Exact flag named, no example of resulting output v1.0.108

07 **First Pass reasoning visibility toggle** NEW 45

Show First Pass reasoning in Fusion routing-profile sessions by toggling visibility with `Ctrl+T`.

— Named shortcut but limited mechanism detail v1.0.108

08 **Config help lists Claude instruction files** IMPROVED 45

`copilot help config` now names the `.claude/rules` instruction files and the `.claude/settings.json` repository settings the CLI reads.

— Names exact files but describes only help text v1.0.108

09 **Default-enabled managed plugin assignments** IMPROVED 35

Managed plugin assignments with state `defaultEnabled` now install and activate plugins by default while preserving user overrides.

— Thin description of plugin state behaviour v1.0.108

10 **Headless session resume conflict error** IMPROVED 35

Headless runs now report an error when resuming a session already in use by another Copilot process, keeping its history intact.

— Behaviour described, no command or flag named v1.0.108

WAS THIS USEFUL? 👍 👎

swttch

Feedback ⓘ v0.36.1 NOTES ↗ RANK

Swttch is a claude Code GUI plugin for JetBrains IDEs. (ex - Claude Code with GUI)

> Swttch adds Transcript view modes with read-aloud playback, matching Claude Desktop behaviour, plus a keyboard shortcut to cycle through view states.

└──▷ **WHAT SHIPPED · 1 FEATURE** most completely described first ? what's the number?

01 **Transcript view modes and read-aloud playback** NEW 70

Swttch now supports Transcript view modes and read-aloud playback, bringing the experience in line with Claude Desktop. Users can cycle through Transcript view states with `Ctrl+Alt+I`, advancing one step per key release.

Cycle through Transcript view modes during a session without leaving the keyboard.

📍 Press Ctrl+Alt+I to advance to the next Transcript view mode; release and press again to continue cycling.

— Names exact shortcut and behaviour but not full mechanism of read-aloud v0.36.1

WAS THIS USEFUL? 👍 👎

Alibaba Qwen Code

Feedback ⓘ desktop-v0.27.1 NOTES ↗

RANK

Qwen Code is an open-source AI coding agent that lives in your terminal.

> Qwen Code's two releases round out managed-agent sessions with Mem0-backed memory, MCP installs and archiving, polish a string of web shell UI surfaces, and add opt-in configuration for LSP, bot message forwarding, background execution and PR observation.

└──▷ **WHAT SHIPPED · 11 FEATURES** most completely described first ? what's the number?

01 **Opt-in bot message forwarding in qqbot** NEW 75

A new `forwardBotMessages` option lets qqbot channels relay messages authored by other bots to the agent. Enable it by setting `forwardBotMessages: true` in the qqbot channel configuration.

Forward messages from bot accounts in a qqbot channel — useful when you want agent responses triggered by another bot to be relayed.

yaml

```yaml
# In your qqbot channel configuration, set:
forwardBotMessages: true
```

— Named config flag with a runnable usage example desktop-v0.27.1

THINNER COVERAGE BELOW

02 **Managed-agent memory, MCP installs, archiving and diagnostics** NEW 55

Managed-agent sessions gain opt-in personal memory backed by Mem0, self-service installation of MCP servers from within a session, a session archive capability, and a Session resource with agent-trace diagnostics.

— Names Mem0, MCP and Session resource but gives no setup steps v0.27.1 desktop-v0.27.1

03 **Web shell UI enhancements across Scheduled Tasks, sidebar and artifacts** NEW 55

The web shell gains Channel delivery targets in the Scheduled Tasks view, sidebar multi-select with bulk pin and archive, workspace collapse controls with drag sorting, artifact cards relabeled Preview, and a strengthened UI-language prompt in the details panel.

— Lists five UI surfaces but no navigation detail given v0.27.1 desktop-v0.27.1

04 **Bound PR observation and state-transition detection** NEW

45

A PR watch store in the daemon/serve layer observes bound PRs and detects state transitions, shipped off by default and toggled via an off switch.

— Describes default-off toggle but not the exact flag name v0.27.1 desktop-v0.27.1

05 **Opt-in LSP activation via user setting** NEW

45

Qwen Code lets users enable the Language Server Protocol through a new user-scoped setting paired with an environment variable (`feat(cli): enable LSP through a user-scoped setting and env pair`).

— Names the setting/env mechanism but not the exact key desktop-v0.27.1

06 **Mod-submitted prompts in OpenTUI** NEW

40

Prompts submitted by a Mod can now be run directly in the OpenTUI terminal.

— Names OpenTUI surface but lacks setup steps v0.27.1

07 **Role guidance for bound sessions in Channels** IMPROVED

30

Bound sessions in Channels now receive role guidance without requiring operators to author templates.

— Describes a behavior change with no mechanism specifics desktop-v0.27.1

08 **Active shell named in agent startup context** IMPROVED

30

The startup context now names the active shell so agents can see the current shell environment.

— States the change but no config or flag given desktop-v0.27.1

09 **Replay projection watermark in acp-bridge** NEW

30

The acp-bridge now exposes the replay projection watermark.

— Names the bridge but not how to read the watermark desktop-v0.27.1

10 **Daemon prompt timing and session sorting** NEW

25

The daemon now exposes the last user prompt time and supports session sorting.

— Bare description, no field or API names given desktop-v0.27.1

11 **Configurable default background execution behavior** NEW

25

Qwen Code adds a configurable agent setting to control the default background execution behavior.

— Mentioned only in summaries, no setting name given `v0.27.1` `desktop-v0.27.1`

WAS THIS USEFUL? 👍 👎

Sourcegraph Amp

Feedback ⓘ Portals on Your Runner (2026-11-09)

BLOG ↗ RANK

Amp is Sourcegraph's agentic coding tool for the terminal and editor, running multi-step edits with subagents and shared team threads.

> Amp adds Portals support for local runners, letting you open dev servers and websites from the thread Portal tab remotely, including shared access for team members.

└──▷ **WHAT SHIPPED · 1 FEATURE** most completely described first ? what's the number?

01 **Portals support for local runners** NEW 83

Start a local runner with Portals enabled using `amp --no-tui --remote-control-portals` to open dev servers and websites from the thread Portal tab from any location. Adding `--share` (`amp --no-tui --remote-control-portals --share`) starts a shared runner so every workspace member can access its Portals, useful for demoing a build or sharing a local preview server with the team.

Run a local runner with Portals enabled so you can open a dev server from the thread Portal tab, even when away from your machine.

```
$ amp --no-tui --remote-control-portals
```

Start a shared runner so every workspace member can access its Portals — useful for demoing a build or sharing a local preview server with the team.

```
$ amp --no-tui --remote-control-portals --share
```

— Names exact flags and commands; mechanism and use cases both given

Portals on Your Runner (2026-11-09)

WAS THIS USEFUL? 👍 👎

Cotool

Feedback ⓘ v0.79.0 NOTES ↗

Cotool is an AI security operations platform for investigating alerts, enriching threat intelligence, and automating response workflows.

Cotool adds per-action approval controls for agent tools (via Slack or Microsoft Teams), MCP detection authoring actions, and a Read-only + Custom tool mode for agents.

Cotool adds per-action approval controls for agent tools, with review in Cotool and optional approval requests through Slack or Microsoft Teams.

Cotool adds MCP actions that draft code detections with the Detection Authoring Agent for review before publishing.

Agents now support a Read-only + Custom tool mode so agents can keep every read-only action while adding selected write actions.

Agent model selection now offers per-chat and per-agent reasoning controls, plus Fast and GPT-6.1 Sol Ultrafast options for token-based plans.

Alert collaboration now sends analyst comments to response agents, mirrors Slack thread replies in Activity, and keeps Slack escalation status synchronized.

+1 more

NEW DETECTION Zendesk built-in detections

NEW DETECTION Socket supply-chain detections

NEW INTEGRATION Zendesk integration

NEW INTEGRATION Iru integration

NEW INTEGRATION Socket and Sigil managed log ingestion

WAS THIS USEFUL? 👍 👎

◆ Coding Agent Add-ons

Paradev

Feedback ⓘ desktop-v2.0.8 NOTES ↗ RANK

Paradev is a desktop workspace for running CLI coding agents like Claude Code in parallel, with terminals, diff viewer, and browser previews.

> Paradev's latest desktop releases add Claude Fable 5.2 model support, CLI-driven cloud workspace networking (port forwarding and host reach), git worktree adoption, and a batch of UI refinements across chat/terminal sessions, browser preview panes, and pull request views.

└──▷ WHAT SHIPPED · 12 FEATURES most completely described first ? what's the number?

01 Cloud workspace networking: port forwarding and host reach NEW

80

Cloud workspaces can now forward workspace ports to your local machine and list active forwards with the `superset ports` command, and can be enabled to reach a host via a new switch in the desktop app or CLI, backed by the `v2_hosts.allow_cloud_workspaces` database field.

List all ports forwarded from a cloud workspace to your local machine during a remote dev session.

```
$ superset ports
```

Allow your cloud workspaces to reach a local host — useful when your team needs remote access to a host you have open.

📍 In the desktop app, go to Settings › Hosts, find your host, and toggle Allow cloud workspaces on.

— Names exact CLI command with a runnable example Names exact database field and gives UI steps

desktop-v2.0.7 desktop-v2.0.6

02 Git worktree adoption via CLI NEW

80

Adopt existing git worktrees into Paradev with the `superset worktrees list` and `superset worktrees open` commands, without re-cloning.

Bring an existing git worktree into Paradev without re-cloning, useful when you already have multiple worktrees checked out.

```
$ superset worktrees list
superset worktrees open
```

— Exact CLI commands with usage example `desktop-v2.0.7`

THINNER COVERAGE BELOW

03 **Chat/terminal session mode setting** IMPROVED 55

The Chat interface setting moved out of Experimental and into the main settings panel, and now only picks the default session mode rather than being the sole option — any agent session can be opened in chat or terminal mode.

— Describes evolution across two releases, no exact path given `desktop-v2.0.7` `desktop-v2.0.6`

04 **Multi-org GitHub connections** NEW 50

Connect multiple GitHub organizations to a single Paradev organization via Settings > Integrations > GitHub, selecting Connect another organization and authorizing each additional org.

Link more than one GitHub org to your Paradev org to consolidate PR review and code scanning across multiple organizations.

📍 In the desktop app, go to Settings › Integrations › GitHub and select Connect another organization, then authorize each additional GitHub organization.

— Gives exact UI navigation steps `desktop-v2.0.6`

05 **Browser preview pane improvements** IMPROVED 50

Hidden browser panes are now unloaded by memory budget — keeping the pages Chrome would keep — to reduce memory use; every pane now shares one browser session and extension system; and new page versions are pushed to open viewers in real time.

— Explains mechanism for three pane behaviors, no config exposed `desktop-v2.0.7`

06 **PR list and code view improvements** IMPROVED 45

The PR Code tab now shows a file list matching the Changes pane with a tree view, and PR list rows now appear before their check results so PRs open faster.

— Names affected UI areas, no exact navigation given `desktop-v2.0.7` `desktop-v2.0.6`

07 **Tab and pane drag and keyboard navigation** IMPROVED 35

Tabs and panes can now be dragged between the main area and the sidebar, and keyboard controls let you navigate right sidebar tabs.

— Names two UI behaviors without exact shortcuts `desktop-v2.0.7`

---

08 **Approval reasoning shown in Full Access mode** IMPROVED

35

The chat view now shows the reason Claude requests approval when operating in Full Access mode, surfacing the agent's rationale before you approve an action.

— Explains behavior but no UI path or example `desktop-v2.0.8`

---

09 **Chat history sync and search** NEW

30

Chat history from cloud workspaces now syncs locally and can be searched across chats.

— States capability without mechanism or UI path `desktop-v2.0.7`

---

10 **Claude Fable 5.2 model support** NEW

30

Paradev now supports Claude Fable 5.2 as a selectable model for agent sessions.

— Names the model but gives no behavior detail `desktop-v2.0.8`

---

11 **Cloud workspace early chat access** IMPROVED

30

Cloud workspaces now open directly to their chats while the underlying box is still waking up.

— Explains behavior but lacks mechanism detail `desktop-v2.0.8`

---

12 **Individually dismissible sidebar cards** IMPROVED

25

Every sidebar card can now be dismissed individually.

— Bare description of a small UI change `desktop-v2.0.8`

WAS THIS USEFUL? 👍 👎

Diagram Design

Feedback

changes since 2026-10-16

CODE ↗

Diagram Design is a Claude Code skill that generates editorial HTML and SVG diagrams and imports draw.io or Mermaid sources.

Diagram Design: Diagram Design adds native Qoder plugin support via `dd0f582`.

Diagram Design adds `skills/diagram-design/scripts/plantuml_extract.py` to extract a normalized intermediate representation from PlantUML `.puml`, `.plantuml`, `.pu`, and fenced Markdown blocks, supporting sequence and class diagrams, with options `--diagram N|all`, `--json`, `--max-rows N`, and `--out PATH`.

WAS THIS USEFUL? 👍 👎

Token Optimizer

Feedback ⓘ v5.13.69 NOTES ↗ RANK

Token Optimizer is a coding-agent plugin that trims wasted context, checkpoints work across sessions and compactions, and reports token spend on a live dashboard.

> Token Optimizer shipped a bundle of Cursor integration features — re-read warnings, checkpoints, quality scoring, coaching, and a beta fleet audit — alongside expanded dashboard history ranges and a Windows install fix.

└──▷ **WHAT SHIPPED · 3 FEATURES** most completely described first ? what's the number?

01 **Windows install support for Cursor hooks** IMPROVED 80

Cursor support can now be installed on Windows, and `install.sh --cursor` honors `--home`; this was previously broken by a POSIX-only file call.

Install Cursor hooks on a Windows machine pointing to a custom home directory.

```
$ install.sh --cursor --home C:\Users\analyst\.token-optimizer
```

— Exact flags and a runnable install command provided v5.13.69

02 **Cursor integration: checkpoints, scoring, coaching, fleet audit** NEW 65

Token Optimizer adds Cursor re-read warnings, checkpoints, quality scoring, coaching, and fleet audit (beta); counts use Cursor data where available, with estimates labelled and read blocking off by default.

— Five named capabilities but no flags or config surfaces given v5.13.69

03 **Dashboard history range selector** IMPROVED 65

Dashboard now offers 90-day, 180-day, and All history ranges, remembers the last choice, and loads full history on demand when served locally.

— Explains behavior but only a UI path, no exact navigation v5.13.69

WAS THIS USEFUL? 👍 👎

AGENT

◆ AI Agent Frameworks

assistant-ui

Feedback ⓘ @assistant-ui/react@0.15.34 NOTES ↗

RANK

Assistant-ui provides React components and state management for chat interfaces with streaming, tool calls, and attachments.

> assistant-ui shipped multi-tab run tracking so stale or in-progress AI replies from another tab are surfaced and resumable instead of silently vanishing.

└──▷ WHAT SHIPPED · 1 FEATURE most completely described first ? what's the number?

01 **Multi-tab stale and in-progress run tracking** NEW 95

New `ActionBarPrimitive.Resume` starts a fresh branch beside a reply whose run went stale, the same way Regenerate does, with its own run lease, while `ActionBarPrimitive.RefreshRun` re-reads a run on demand when a reply from another tab may still be in progress. Messages now carry `runState` (`'stale'` or `'in-progress'`), exposed by `ErrorPrimitive.Root` as the `data-run-state` attribute with `role='status'` while in progress; `ActionBarPrimitive.Root` gains `hideWhenRunning` to hide the action bar while another tab writes the reply. `MessageFormatItem` gains `unstable_status` and `GenericThreadHistoryAdapter` gains `unstable_onReload` to surface these states in the AI SDK runtime, and a reply whose run misses a heartbeat past `staleAfter` (default 30 seconds, configurable on `useCloudThreadListAdapter`) now loads as an interrupted error instead of disappearing.

Show a Resume button beside a stale interrupted reply so users can continue the conversation from that branch.

tsx

```tsx
import { ActionBarPrimitive } from "@assistant-ui/react";

export function MessageActionBar() {
  return (
    <ActionBarPrimitive.Root hideWhenRunning>
      <ActionBarPrimitive.Resume />
      <ActionBarPrimitive.RefreshRun />
    </ActionBarPrimitive.Root>
  );
}
```

Tune how long a Cloud-backed thread waits for a heartbeat before treating a reply as interrupted.

```tsx
import { useCloudThreadListAdapter } from "assistant-cloud";

const adapter = useCloudThreadListAdapter({
  staleAfter: 60,  // seconds before a heartbeat-less run is
marked stale
});
```

— Names every API surface with a runnable config and usage example.

`@assistant-ui/react@0.15.34`

WAS THIS USEFUL? 👍 👎

cotal

Feedback ⓘ v0.154.0 NOTES ↗ RANK

Cotal is the open standard for agent coordination

> This window's biggest addition is endpoint delegation via persona `endpoints`, letting agents grant spawned helpers scoped access to specific commands, alongside daemon setup that now hosts a Manager per space and a new permission-scoped endpoint discovery mode for the auth service. Other releases added structured JSON receipts for `deliver pending` and `up`, workflow-status and watch-roster visibility improvements, and several breaking changes including stricter persona `endpoints` validation and removal of the `terminal` extension kind.

└──▷ **WHAT SHIPPED · 16 FEATURES** most completely described first ? what's the number?

01 **Endpoint delegation via persona `endpoints`** NEW 90

Agents can delegate specific endpoint commands to a spawned helper by listing them under `endpoints` in the helper's persona — each entry names an endpoint and its commands, e.g. `{ endpoint: com.example.orders, commands: [list] }`, with an optional `instanceId` to restrict to one instance; this works in a persona file, in the inline persona on `cotal_spawn`, in `cotal_persona`, and in a cotal-lang inline `spawn`. Operators can mint endpoint commands for their own seat with `cotal spawn --foreground`, and `@cotal-ai/core` now exports `parsePersonaEndpoints`, `lowerPersonaEndpoints`, `copyPersonaEndpoints`, and the `PersonaEndpoint` type, with `LaunchOpts` gaining `endpoints`, forwarded to the session as `COTAL_ENDPOINTS`. A persona file whose `endpoints` key is not a valid list now fails to load instead of being silently ignored, and the manager cluster contract moves to revision 47 to add `endpoints` to the `spawn` persona record and `define-persona`.

Grant a spawned helper access to specific order-service commands so it can call only those endpoints and no others.

text

```
cotal_spawn {
  persona: {
    endpoints: [
      { endpoint: com.example.orders, commands: [list] }
    ]
  }
}
```

— Full API surface and config example given, no CLI command shown v0.154.0

02 **`spawn --dry-run` fails fast on unresolvable harness** IMPROVED

86

Running `spawn --dry-run` now reports `DEPENDENCY_UNAVAILABLE` with exit code 2 when an unknown flag, persona pin, or preference makes the effective harness unresolvable, instead of reporting success with a warning.

Catch an unresolvable harness before actually launching, so CI fails fast with a clear exit code rather than a misleading success.

```
$ cotal spawn --dry-run --server wss://broker.example.com
```

— Exact error code, exit status, and runnable example given v0.152.0

03 **Machine-daemon hosts a Manager per space** NEW

83

Daemon setup now hosts a stock Manager per space, resumes it after restart, and reports launch-readiness separately from machine-service approval, with `manager.instanceId` carrying the pinned Manager's identity in setup output. Use `--no-manager` on `cotal daemon setup` to skip hosting a Manager while still connecting the machine daemon (required for `--wait`), and `cotal update` now waits for hosted Manager readiness before returning.

Set up a machine daemon for connection-only use — no Manager is hosted — when the space already has a Manager running elsewhere.

```
$ cotal daemon setup --no-manager
```

— Names flags and output field but no full workflow walkthrough v0.154.0

04 **JSON receipts for `deliver pending` and `up`** IMPROVED

83

`cotal deliver pending --json` now returns a single JSON receipt including candidate IDs and not-found refusals, with results read from a `result` field instead of the old first-line object plus trailing text. Interrupted `cotal up --json` runs now also return a JSON receipt naming the members still running and the recovery commands.

Retrieve pending deliveries as a machine-readable receipt, including candidate IDs and any not-found refusals, for use in CI pipelines or scripted workflows.

```
$ cotal deliver pending --json
```

— Exact command and field name given, runnable example included v0.153.1

05 **Permission-scoped endpoint discovery in auth service** NEW

83

Enable permission-scoped endpoint discovery in the auth service by passing `--registry` to the builtin auth-service preset, or by enabling it per entry in multi-space auth manifests via `registryService`; the registry is disabled by default and stops with its auth plane.

— Names flag and config key, no example given v0.151.0

06 **Idle-wake notice on Claude seats** NEW

78

Cotal now surfaces an idle-wake notice on Claude seats that showed no development-channels prompt at startup: `cotal spawn` prints a `STARTUP_NOTICE` warning, the `cotal_spawn` result carries the notice, and `cotal ps` marks the seat `no idle wake`.

— Names exact warning strings and commands, lacks runnable example v0.151.0

07 **Terminal extension kind removed from core** BREAKING

73

`@cotal-ai/core` no longer exports the `terminal` extension kind (`TerminalLayout`, `Tab`, and `Pane`); `@cotal-ai/tmux` and `@cotal-ai/cmux` no longer export or register `tmuxTerminalProvider` and `cmuxTerminalProvider`. Affected imports must be removed before upgrading.

— Names exact removed exports and migration step v0.151.0

08 **`cotal space forget` for offline spaces** IMPROVED

70

Offline registered spaces now name `cotal space forget` as the command to unregister them, replacing the hidden legacy `cotal meshes rm` suggestion.

Unregister an offline registered space using the now-surfaced command, instead of the deprecated legacy alias.

```
$ cotal space forget
```

— Exact replacement command named with example v0.153.1

09 **Kick-authority exports for eviction sweeps** NEW

68

`@cotal-ai/core` now exports `KickTarget`, `kickConnection`, `accountHoldsConnection`, and a `kick` function so the multi-space delivery host's eviction sweep can delegate kick authority without holding a broker-wide `$SYS` connection between kicks.

— Names exact exports and mechanism, no usage example v0.151.0

10 **Workflow commands require regular source files** BREAKING

65

Workflow commands now require a regular source file and refuse devices, FIFOs, and sockets — pipe and process substitution inputs must be saved to a regular file first.

— Clear migration note but no exact command listed v0.152.0

11 **`--no-persona` flag for agent replacement** NEW 60

Replacing an agent with `--no-persona` now keeps explicit grants and identity-carried grants while excluding persona defaults, on agent replacement commands.

— Names flag and behavior but no example or command context v0.154.0

12 **Checkpoint schema display in workflow status** NEW 60

An open checkpoint's declared schema is now shown in workflow status views, recorded with the question and included in both text and JSON output.

— States output formats but no command or flag shown v0.152.0

THINNER COVERAGE BELOW

13 **Broker URL validation and TLS enforcement** IMPROVED 55

Commands now reject invalid broker URLs and require TLS for secure schemes when connecting via `--server`, including on registered and raw connections.

— Names flag and scope, no example or error detail v0.152.0

14 **Chat sessions and endpoints in watch roster** NEW 55

The watch roster now shows chat sessions and service endpoints alongside agents, with each participant's kind, reported state, and identity shown in both plain and terminal views.

— Describes feature scope but no command to invoke it v0.152.0

15 **Inherited-value marker in channel views** IMPROVED 50

Channel list and detail views now mark inherited replay, window, and delivery settings with `inherited` for values without a saved override, including built-in policy values.

— Describes UI change with no navigation path given v0.152.0

16 **Shell completion for `--daemon-home`** IMPROVED 35

Shell path completion is now enabled for the daemon's `--daemon-home` option.

— Single-line description, no example or scope detail v0.153.1

└──▷ BREAKING ON UPGRADE

! A persona file whose `endpoints` key is not a valid list now fails to load; previously the key was silently ignored.

! The manager cluster contract moves to revision 47, adding `endpoints` to the `spawn` persona record and to `define-persona`.

! The `cotal deliver pending --json` receipt format changes: results must now be read from `result` instead of the old first-line object and trailing text.

! Workflow commands now require a regular source file and refuse devices, FIFOs and sockets — pipe and process substitution inputs must be saved to a regular file first.

! `@cotal-ai/core` no longer exports the `terminal` extension kind (`TerminalLayout`, `Tab`, and `Pane`); `@cotal-ai/tmux` and `@cotal-ai/cmux` no longer export or register `tmuxTerminalProvider` and `cmuxTerminalProvider`. Remove the affected imports before upgrading.

WAS THIS USEFUL? 👍 👎

HeyGen HyperFrames

Feedback ⓘ v0.9.28 NOTES ↗ RANK

HyperFrames is an open-source HTML-to-video renderer that runs in AI-agent workflows.

> HyperFrames extended its motion blur support across both reporting and rendering, added reverse-any-video and style-sampling skills with a simpler install path, and improved CLI and telemetry reporting including better agent-runtime detection.

└──▷ **WHAT SHIPPED · 4 FEATURES** most completely described first ? what's the number?

01 **Reverse-any-video and style-sampling skills added to core set**

NEW 85

The `reverse-any-video` and `style-sampling` skills now install with the core skill set via `npx hyperframes skills update`, so agents have them without an extra install step. Styles can also now be routed outside the templates listing via the Skills system.

– Names both skills, the install command, and the Skills system routing change. v0.9.27

02 **Motion blur outcome fields in piped render results** NEW

65

Piped renders now include motion blur outcome data — tier, movers, and failed frames — in the `@hf-result` line, so a host app can read and report those results.

– Names exact fields and output line, but no usage example given. v0.9.28

THINNER COVERAGE BELOW

03 **CLI and telemetry reporting improvements** IMPROVED 55

CLI commands now report which skills are installed, and the `reverse` and `templates` commands report their own runs. Telemetry now adds agent markers that other detectors recognise (verified upstream) and detects more agent runtimes.

– Names affected commands but lacks concrete output format or new runtimes. v0.9.27

04 **Motion blur honoured in Producer distributed and range renders**

IMPROVED 40

Distributed and range renders in Producer now honour page motion blur.

– States the change but no mechanism or configuration detail. v0.9.27

WAS THIS USEFUL? 👍 👎

sortie

Feedback ⓘ v1.32.0 NOTES ↗ RANK

sortie turns tracker tickets into autonomous agent sessions.

> Sortie's v1.32.0 release adds a live issue-comment feed that reaches agents mid-run, token-level spend breakdowns and fuller time-range drill-down in Analytics, and consolidates Copilot CLI and OpenCode onto a new `agent-client-protocol` agent kind with added capabilities.

└──▷ **WHAT SHIPPED · 5 FEATURES** most completely described first ? what's the number?

01 Live issue-comment feed into agent turns

BREAKING

87

Agents now receive new issue comments before each turn — only from users with write access — so human guidance mid-run reaches the agent without a restart. Disable it with `comment_feed.enabled: false` in the workflow config, or place the feed into a custom prompt with the `comment_feed` template variable. The feed is enabled by default, so upgrading changes agent prompts unless explicitly disabled.

Opt out of the live comment feed for a workflow where you control the full agent prompt and do not want mid-run comments injected automatically.

yaml

```yaml
comment_feed:
  enabled: false
```

— Exact config key, template variable, and opt-out example given. v1.32.0

02 Deprecation of copilot-cli and opencode agent kinds

DEPRECATED

80

Copilot CLI and OpenCode now run through the `agent-client-protocol` agent kind; the `copilot-cli` and `opencode` kinds are deprecated and will be removed in a later release. Workflows using those kinds log a warning on upgrade, and workflows that rely on `no_custom_instructions` should stay on `copilot-cli` for now.

Migrate a Copilot CLI workflow from the deprecated `copilot-cli` kind to `agent-client-protocol` to get environment-variable passing and reasoning-level control.

```yaml
agent:
  kind: agent-client-protocol
  command: copilot --acp --allow-all-tools
  max_turns: 15
  max_concurrent_agents: 4
```

— Names deprecated kinds plus a concrete migration example. v1.32.0

THINNER COVERAGE BELOW

03 **Token-level spend breakdown in History and Analytics** NEW

55

History and Analytics now show spend as tokens broken into input, output, cache write, and cache read (including prompt-cache share), or as cost — cost remains the default view and you switch between the two one at a time.

— Names the breakdown categories but no exact UI path. v1.32.0

04 **Full time-range coverage and drill-down in Analytics chart**

IMPROVED

55

The Analytics chart now covers every time range including Last 24 hours, places spend at the time each agent turn ended, and lets you select a bar or stretch of bars to see cost, turns, errors, and the sessions behind them.

— Clear chart behaviour described, no named config surface. v1.32.0

05 **Environment variables, reasoning level and session options for agent-client-protocol**

NEW

50

The `agent-client-protocol` agent kind now supports passing environment variables to the agent, setting a reasoning level, and choosing the agent's own session options such as a named Copilot CLI agent.

Migrate a Copilot CLI workflow from the deprecated `copilot-cli` kind to `agent-client-protocol` to get environment-variable passing and reasoning-level control.

```yaml
agent:
  kind: agent-client-protocol
  command: copilot --acp --allow-all-tools
  max_turns: 15
  max_concurrent_agents: 4
```

— Lists new capabilities but no example isolating each. v1.32.0

┗──▷ BREAKING ON UPGRADE

! Agent prompts change on upgrade because `comment_feed` is enabled by default; set `comment_feed.enabled: false` to restore previous behaviour.

! Workflows using the `copilot-cli` or `opencode` agent kinds will log a warning on upgrade; a later release removes those kinds entirely. Workflows that rely on `no_custom_instructions` should stay on `copilot-cli` for now.

WAS THIS USEFUL?  

atomic-agent

Feedback ⓘ v0.7.0 NOTES ↗ RANK

Atomic Agent is a local-first AI agent. Runs open-weight models on your own machine via llama.cpp.

> Atomic Agent's CLI and desktop app both shipped overnight unattended task runs with spend and time budgets, a persistent approval-rules system with risk-scored smart approval, a plugin framework for Claude Code/Codex/Cursor extensions, and session handoff from those tools, alongside Fusion worker controls and on-screen click guidance.

└──▷ **WHAT SHIPPED · 12 FEATURES** most completely described first ? what's the number?

01 **Overnight unattended task runs** NEW 95

CLI adds `atomic-agent overnight "<task>"` to run long unattended tasks, with `--budget <usd>` to cap spend, `--hours <h>` to cap time (default 8, max 24), `--dir <path>` to set the folder, `--notify telegram` to send a report, and `--detach` to background the run and print the task id and log path; `--no-approval` allows all actions or the default deny-with-log behaviour logs each denied action; quiet hours are set with `ATOMIC_AGENT_TASKS_REPORT_QUIET_HOURS=23:00-08:00`, and running tasks can be checked with `/status` in Telegram. The desktop app exposes the same run via the composer's + menu (8-hour default, up to 24), showing a card with time and spend, with paired Telegram delivering the report on completion.

Start a long background coding task capped at $5 and 8 hours, with Telegram notification on completion — then check on it in the morning.

```
$ atomic-agent overnight "Refactor the auth module to use JWTs, add tests, and open a PR" --budget 5 --hours 8 --dir ~/projects/myapp --notify telegram --detach
```

— Full CLI flags plus desktop equivalent, with a runnable example v0.7.0

desktop-v0.0.8

02 **Approval rules and smart approval** NEW 93

Approval rules persist in `approval-rules.json` in the agent folder (allow, ask, deny) or `.atomic-agent/approval-rules.json` in a project (ask and deny only); save an Always allow rule with `ctrl+n` in the TUI or the `w` key in the CLI; `agent.smartApproval` (optionally with `agent.smartApprovalModel`) enables risk-scored approvals that auto-allow low-risk actions and prompt only for the rest. In desktop, rules are saved via the approval card's Always allow button, listed and deletable under Settings › Privacy, and every auto-approved action in a finished turn can

be reviewed via an `Auto-approved N actions` disclosure showing whether a saved rule or earlier chat approval covered it.

Cap task spend and enable risk-scored approvals so low-risk actions run unattended while dangerous commands still prompt.

```yaml
agent.task.maxCostUsd: 10
agent.smartApproval: true
agent.smartApprovalModel: gemma-4-26b
agent.task.resumeInterrupted: true
agent.task.maxResumes: 3
```

— Named config keys, files and UI paths across both surfaces `v0.7.0` `desktop-v0.0.8`

---

03 **Plugin system for Claude Code, Codex, and Cursor** NEW

93

Install plugins with `atomic-agent plugin install <folder>` (skills appear as `plugin:skill`, MCP servers join the agent, commands become TUI slash commands); browse a marketplace with `atomic-agent plugin marketplace add <owner/repo>`, then `plugin install <name>`, `plugin update`, `plugin list`, `plugin enable`, `plugin disable`, and `plugin remove`. Hooks use Claude's `hooks.json` format loaded from `~/.atomic-agent/hooks.json` or a plugin, with `PreToolUse` / `PostToolUse` matchers for tool names like `Bash`, `Read`, `Write`, `Edit`, and `mcp__server__tool`; plugin code execution requires `--accept-code` or `plugin allow-code <name>`, with the `/plugins` TUI view (`ctrl+g x`) toggling plugins, parts, and code consent. In desktop, plugins are managed (toggle parts, allow/block code, remove) via Settings › Plugins, with the same `atomic-agent plugin install` available from the terminal.

Install a plugin from the terminal when the Settings › Plugins UI is not enough.

```
$ atomic-agent plugin install <plugin-name>
```

— Enumerates every subcommand, config path and matcher named `v0.7.0` `desktop-v0.0.8`

---

04 **Visual screen guidance and window capture** NEW

87

The `os.screen.point` tool (macOS, desktop app), controlled by `tools.screen.point`, shows on-screen arrows, boxes, and numbered steps; `os.screen.capture` captures a specific app's window (even when obscured) and reads its text via `app`, `window`, and `text: true` parameters. The desktop app overlays arrows, boxes, and numbered steps on any app to guide click targets on macOS, with markers clearing when the pointer reaches them, on Esc, or on timeout, and lets you

attach a screenshot of another app's window via the button next to + or File › Attach App Window....

— Named tool and params plus exact desktop UI paths `v0.7.0` `desktop-v0.0.8`

05 **Session handoff from Claude Code and Codex** NEW 86

Hand off a stalled Claude Code or Codex session with `atag handoff --from claude-code` (or `codex`), using `--session <id>`, `--list`, `--note "<text>"`, and `--no-open`; long sessions are trimmed to fit the local model window via `--context <tokens>`. In desktop, the same continuation is reachable via Command palette › `Continue from Claude Code / Codex` or Settings › Import › Continue a task, reopening it as a new chat with its history.

Hand off a mid-task Claude Code session to atomic-agent to continue with a local model, trimming the history to fit a 32k context window.

```
$ atag handoff --from claude-code --context 32000 --note "Focus on the failing tests in src/auth"
```

— Exact flags and desktop menu path, with a runnable example `v0.7.0` `desktop-v0.0.8`

06 **Fusion worker controls** NEW 82

Cap a single Fusion turn with `llm.runMode.fusion.turnBudgetUsd` and `turnTimeBudgetMs`, and assign models by role using `llm.runMode.fusion.roles` (plan, build, review, explore, utility), opting into a task tree (`taskTree.enabled`), worktree isolation (`isolation: "worktree"`), a green-checklist judge (`judge`), automatic replanning (`autoReplan`), and stuck-worker escalation (`escalation`).

— Config keys named but no worked example given `v0.7.0`

07 **Task progress and step visibility** NEW 78

In the TUI, finished-turn display is controlled with `/expand`, `/expand all`, `/collapse`, `/steps on`, and the `tui.showSteps` config key. In desktop, a live plan bar shows `Plan N/M` above the composer and folds to a chip, per-chat cost and plan limit with reset times appear in a chip beside the context ring, and `/status` prints the current turn's status in the composer.

Check what the agent is doing mid-turn without leaving the composer.

```
$ /status
```

— Named commands and UI elements across TUI and desktop `v0.7.0` `desktop-v0.0.8`

08 **Goal checklist in TUI** NEW 70

The TUI gains a goal checklist with `/goal`, with optional checks such as `check: npm test`; manage it with `/goal status`, `/goal pause`, `/goal resume`, and `/goal clear`.

— Exact subcommands named but no worked example v0.7.0

09 **Prompt compaction for small local models** NEW 65

`agent.prompt.compact` (`auto` by default, also `on` or `off`) tunes prompt size to fit more task context in 64K-or-smaller context windows.

— Named config key with default but no worked example v0.7.0

10 **Local fallback no longer automatic** BREAKING 65

`llm.fallback.appendLocal` is now off by default, so a working setup that relied on automatic fallback to a local model when a cloud provider fails will no longer switch automatically.

— Exact config key and before/after behaviour named v0.7.0

11 **Model selection UX in desktop** NEW 62

Desktop automatically downloads and switches to the best local model your hardware can run, prompting to Switch now or Later while the current model keeps working, and marks the first three models as Cheap, Balanced, and Best after adding OpenRouter, AI/ML API, or Gemini.

— UI behaviour described but no exact navigation path given desktop-v0.0.8

12 **Reply and markdown rendering controls** IMPROVED 60

Reply post-processing (closing unclosed fences, removing filler lines) can be disabled with `ATOMIC_AGENT_REPLY_POLISH=0`. Desktop renders nested lists, footnotes, and safer links in Markdown, and raw HTML now shows as text instead of rendering.

— Env var named but desktop rendering changes stated only in prose v0.7.0 desktop-v0.0.8

└──▷ **BREAKING ON UPGRADE**

! `llm.fallback.appendLocal` is now off by default, so a working setup that relied on automatic fallback to a local model when a cloud provider fails will no longer switch automatically.

WAS THIS USEFUL? 👍 👎

lexmount Moli

Feedback ⓘ v1.1.23 NOTES ↗ RANK

Moli is a best headless browser for AI agents. Lite, Fast, High-Compatibility. Built in Rust

> Moli's latest release deepens native WebAudio engine support and adds lower-level performance and DOM fidelity, including Resource Timing initiator URL snapshots and persisted response status on Document identities.

└──▷ **WHAT SHIPPED · 3 FEATURES** most completely described first ? what's the number?

01 **Native WebAudio engine node implementations** NEW 68

Moli now implements extensive native WebAudio engine functionality: scheduled audio sources, `DelayNode` and `StereoPannerNode`, channel routing nodes and port controls, native `AudioBuffer` storage and channel bindings, `WaveShaperNode` controls and backend guards, `ConvolverNode` impulse response controls, `IIRFilterNode` and analytic frequency responses, and `PeriodicWave` with native oscillator controls.

— Names every node type added but gives no usage or API example. v1.1.23

THINNER COVERAGE BELOW

02 **Resource Timing initiatorUrl snapshots** NEW 41

Moli now exposes native Resource Timing `initiatorUrl` snapshots in performance data.

— Names the field but no mechanism or usage path. v1.1.23

03 **Response status retained on Document identities** IMPROVED 38

Moli now retains source response status on `Document` identities in the DOM.

— States the change but no detail on how it's surfaced or used. v1.1.23

WAS THIS USEFUL? 👍 👎

agentlas-os

Feedback ⓘ os-v1.2.96 NOTES ↗ RANK

Agentlas-os is the Open-Source AI Agent OS inspired by OpenAI Dots, Meta Muse, OpenClaw, Hermes & Mem0. Autonomous 3-tier headhunting, Receipt & Verifier loops, scoped memory, and BYOM for Claude Code, Codex & Cursor.

> Agentlas OS expanded its code map to parse imports across nine languages with language-specific alias resolution, sharpened impact-analysis labeling, locked down the routing benchmark runner behind ownership and SHA-256 checks, and tuned Context Map's result ranking and deduplication.

└──▷ **WHAT SHIPPED · 4 FEATURES** most completely described first ? what's the number?

01 Hardened routing benchmark runner distribution BREAKING

90

The routing benchmark runner (`hephaestus network bench`) is now excluded from new installs and offered only where it is already present; on POSIX, the installed copy must be owned by the current user, must not sit in a world-writable directory, and is verified against its exact SHA-256 before use.

— Exact command and concrete security checks named, ready to verify os-v1.2.96

02 Code map reads imports across 9 languages NEW

70

The code map now reads project-internal imports for JavaScript/TypeScript, Dart, C, Go, Ruby, Python, Java/Kotlin and Rust, and understands `tsconfig` path aliases, Dart `package:` imports, Go modules, Rust `crate::`, C headers and Python `from .` `import module`.

— Names many languages and import syntaxes but no command to invoke os-v1.2.96

THINNER COVERAGE BELOW

03 Impact and verify dependency labeling IMPROVED

50

Impact and verify now label files whose language lacks import support as not looked at, separate direct dependents from the transitive chain, and always name direct importers even when the change is too large to list in full.

— Behavior change described but no invocation point given os-v1.2.96

04 Context Map ranking and dedup improvements IMPROVED

40

Context Map now ranks results more accurately for plain-language queries and excludes byte-identical plugin mirror copies and vendored libraries from results.

— No metric or example quantifies the ranking improvement os-v1.2.96

WAS THIS USEFUL? 👍 👎

CopilotKit OpenBot

Feedback ⓘ v1.0.0 NOTES ↗ RANK

OpenBot runs self-hosted AI coworkers that each get their own browser, files, and gated tools, with every action decided and recorded.

> OpenBot's v1.0.0 release adds multi-account app connections with per-account approval rules, clearer Composio key-setup guidance with 1Password links, and richer Desktop diagnostics for setup failures and resource shortfalls, alongside smaller UI and reliability improvements across chat, learning, and conversations.

└──▷ **WHAT SHIPPED · 10 FEATURES** most completely described first ? what's the number?

01 **Desktop diagnostics for setup failures and resource shortfalls**

IMPROVED 95

Back end setup failures on OpenBot Desktop now report the exit code, error output, and duration in the log, with the card distinguishing a mid-setup error, a silent quit, and a launch failure. Docker or Podman resource shortfalls now name what is short with actual numbers and where to raise the limit — e.g. 'Docker has 0 GB of free disk. OpenBot needs 40 GB. Raise the disk limit in Docker Desktop > Settings > Resources' — or give the `podman machine set` command.

— Includes exact error fields, example message, and a runnable command. v1.0.0

02 **Multi-account connections with per-account approval rules** NEW

90

OpenBot now supports connecting multiple accounts of the same app (e.g. Gmail), with per-account approval rules — prompts name the account ("Research Bot wants to send email from … on Gmail") and Always allow here saves the rule for that account only. Existing per-app Always allow rules stop applying once a second account is connected, requiring new per-account rules to be saved. Two database migrations run on first start: `0084_brokered_accounts` creates a new `brokered_accounts` table and `0085_approval_rule_account` adds a nullable `account` column to `approval_rules`. Rolling back to 1.0.0-rc.1 or 0.3.4 after connecting multiple accounts is unsafe (a rule for one account can apply to every account, and Disconnect can leave an extra account connected at Arcade) — restore a pre-upgrade backup instead.

Connect a second Gmail account after the upgrade so a Bot can send from either address.

📍 In Connections, open Gmail's app page and press Connect another account, sign in with the second address, then press Make default next to whichever address the Bot should use when none is named in a request.

— Names migrations, table/column, rollback risk, and shows connect steps. v1.0.0

03 **Composio key-setup guidance with 1Password links** NEW

60

Apps that require an API key now say so before you connect on Composio, with guidance on where to get the key and a Get a key from 1Password link surfaced for 33 of the most-used key apps.

— Names Composio and 1Password integration but no exact UI path or command. v1.0.0

THINNER COVERAGE BELOW

04 **Desktop update download fallback source** IMPROVED

50

App updates for OpenBot Desktop now download from `downloads.copilotkit.ai` first, falling back to GitHub if that download fails, with the log recording which source was used.

— Names the domain and fallback order, no reader action needed. v1.0.0

05 **Bots report accurate app connection status** IMPROVED

45

Bots are now told which apps they hold and their connection status from the first message, so they never list a disconnected app as reachable or call a missing tool.

— Describes behavior change but no mechanism or config surface. v1.0.0

06 **Teach a Bot learning lifecycle UI** NEW

45

Teach a Bot now shows the full lifecycle of a recording through Learning — Learn now, Learning from your recordings, Ready to review with Review, and Skill learned with Try it — and sends a notification when the skill is ready.

— Names each lifecycle state but no config or command. v1.0.0

07 **Built-in component descriptions reset on upgrade** BREAKING

45

On first browser or Desktop open after upgrade, every unedited built-in component description (Notice, Checklist, Record, Headline figures, Data table, Bar chart, Donut chart, Progress against target) is overwritten with this release's wording; admin-saved or published/unpublished descriptions are preserved.

— Names affected components but no migration or opt-out step. v1.0.0

08 **Chart and table rows link to source records** IMPROVED

40

Bar, progress row, headline figure, and donut chart legend rows now link to the source record or search URL, opening in a new tab.

— Names components affected but only a UI click, no command. v1.0.0

09 **New conversation shortcut defaults** IMPROVED 40

Opening a new conversation with `+` or Shift+N now starts with To: empty and the Bot menu open.

— Exact shortcuts given but the change itself is minor. v1.0.0

10 **Markdown-first default answer format** IMPROVED 30

Bots now answer in Markdown by default; notice cards, charts, and tables are reserved for alerts that need action or when structure genuinely aids the answer.

— States the change but no example or toggle named. v1.0.0

└──▷ BREAKING ON UPGRADE

! Rolling back to 1.0.0-rc.1 or 0.3.4 after users have connected more than one account of an app is not safe: a rule saved for one account applies to every account, and Disconnect can leave an extra account connected at Arcade — restore the pre-upgrade backup instead.

! Two database migrations run on first start: `0084_brokered_accounts` creates a new `brokered_accounts` table and `0085_approval_rule_account` adds a nullable `account` column to `approval_rules`.

! Existing per-app Always allow rules stop applying once a user connects a second account of that app; Bots will ask for approval again for each account until new per-account rules are saved.

! The Helm chart's bundled `intelligence` subchart is updated to 0.16.0-rc.10 from rc.9.

! On first browser or Desktop open after upgrade, every unedited built-in component description (Notice, Checklist, Record, Headline figures, Data table, Bar chart, Donut chart, Progress against target) is overwritten with this release's wording; admin-saved or published/unpublished descriptions are preserved.

WAS THIS USEFUL? 👍 👎

◆ Agent Memory

claude-mem

Feedback ⓘ v15.15.1 NOTES ↗ RANK

claude-mem gives coding agents persistent memory by capturing session activity, compressing it with a model, and injecting relevant context into later sessions.

> claude-mem shipped auto-clearing of near-full context windows, raw tool-evidence search, daily self-updates, and a faster hook-only execution bundle, alongside expanded secret redaction, a consolidated session-start rules file, Korean UI support, and two new openai-compatible provider presets.

└──▷ WHAT SHIPPED · 11 FEATURES most completely described first ? what's the number?

01 Auto-clear context window at one-third full NEW

95

Auto-clear the context window when a Claude Code conversation reaches one-third full by setting `CLAUDE_MEM_AUTO_CLEAR_ENABLED=true`; Claude saves work state, runs `/clear`, sends `continue`, and the fresh conversation resumes where it left off — disabled by default, and skipped if no save exists. The setting can be persisted in `~/.claude-mem/.env`.

Persist the auto-clear setting so every session in a project picks it up without re-exporting the variable.

```ini
CLAUDE_MEM_AUTO_CLEAR_ENABLED=true
```

— Exact flag, config file path and full mechanism given v15.13.0

02 Daily self-update with doctor status check NEW

90

claude-mem enables automatic daily self-updates by default; check status with `npx claude-mem doctor` (shows an Auto-update row), or disable by setting `CLAUDE_MEM_AUTO_UPDATE=false` or `DISABLE_AUTOUPDATER`.

Verify auto-update is working and see when the next update check will run.

```
$ npx claude-mem doctor
```

— Exact runnable command and both disable flags given v15.14.0

03 Token redaction and Sensitive filter in Viewer NEW

85

claude-mem now redacts Grafana service account tokens (`glsa_`), Grafana Cloud tokens (`glc_`), and Postman API keys (`PMAK-`) in both the worker and the Cowork shim, with per-token opt-out via `CLAUDE_MEM_REDACT_DISABLED_BUILTINS` . The Viewer session page now shows a dedicated `Sensitive` filter for observations of type `sensitive` , instead of grouping them under `Other` .

— Names exact token prefixes and the opt-out variable v15.15.1

04 **SessionStart message restructuring** IMPROVED 85

Session-start instructions are now consolidated into a single file at `~/.claude-mem/RULES-AND-TOOLS.md` , referenced by one `Access Claude-Mem Rules + Tools` line in the SessionStart header instead of inline fields for format, legend, mode, fetch, and work-state rules; startup context now consists of the timeline alone, with the SessionStart message being the header lines followed by the timeline ending on the newest entry, while work-state and memory-use instructions have moved to the MCP server's instructions.

— Names the exact file path and replaced header line Explains restructuring but gives no config to act on

v15.15.0 v15.14.0

05 **Raw tool evidence search** NEW 85

Enable raw tool evidence search — looking through stored tool input and output when observation search finds nothing — by setting `CLAUDE_MEM_TOOL_USES_SEARCH_ENABLED=true` ; it returns `#T` IDs for `get_tool_uses` .

— Exact env var and ID format for the new search path v15.14.0

06 **Removal of savings-display environment variables** BREAKING 70

The `CLAUDE_MEM_CONTEXT_SHOW_SAVINGS_AMOUNT` and `CLAUDE_MEM_CONTEXT_SHOW_SAVINGS_PERCENT` environment variables and their viewer toggle are removed; the savings line no longer appears in the context header, though existing settings files that still carry them load without error.

— Names both removed variables, no migration step needed v15.15.0

07 **My Work focus-areas overview** NEW 60

My Work now opens on a focus-areas overview: each area gets a card with artwork, its goal and live request counts; artwork display can be set to full, compact or off.

— Describes new layout and a display setting, no exact path v15.14.0

THINNER COVERAGE BELOW

08 **Hook-only bundle performance improvement** IMPROVED 55

Hooks now run from a small hook-only bundle instead of the full worker bundle, cutting single-hook time and memory in half and reducing CPU usage for a dozen parallel hooks to about a quarter.

— Concrete before/after numbers but nothing to configure v15.14.0

09 **Korean UI support** NEW 40

Adds Korean UI support; the viewer, My Work, setup and Observation TV can now be shown in Korean.

— Names affected surfaces but no setting shown v15.14.0

10 **Rust smart navigation** NEW 35

Smart navigation now shows Rust traits, implementations and data definitions.

— Bare description, no usage path given v15.14.0

11 **New openai-compatible provider presets** NEW 35

Adds the Standard Compute preset and the PopNow preset for the openai-compatible provider.

— Names presets but gives no setup detail v15.13.0

└──▷ **BREAKING ON UPGRADE**

! The `CLAUDE_MEM_CONTEXT_SHOW_SAVINGS_AMOUNT` and `CLAUDE_MEM_CONTEXT_SHOW_SAVINGS_PERCENT` environment variables and their viewer toggle are removed; the savings line no longer appears in the context header.

WAS THIS USEFUL? 👍 👎

Volcengine OpenViking

Feedback ⓘ v0.5.4 NOTES ↗ RANK

OpenViking is a context database that manages agent memory, knowledge retrieval, and skills.

> OpenViking v0.5.4 adds Streamable HTTP MCP server integration, a memory-consolidation command, and distributed MySQL-backed Watch workloads, alongside several breaking changes to storage locking, group membership, snapshot restore, SDKs, CacheFS/Redis defaults, and reranking visibility.

└──▷ **WHAT SHIPPED · 12 FEATURES** most completely described first ? what's the number?

01 Streamable HTTP MCP server registration and tool mounting NEW

95

Register Streamable HTTP MCP servers, discover and invoke their tools, and mount tools as readable Markdown nodes using `ov mcps add`, `ov mcps find`, `ov mcps mount`, and `ov mcps invoke`, available via REST, CLI, and SDK interfaces.

Register a live external MCP server, mount one of its tools as a readable Markdown node, then read that node — useful for surfacing real-time data (e.g. weather) inside an agent context.

```
$ ov mcps add -p viking://agent/tools --wait '{"mcpServers":
{"weather":{"url":"https://mcp.example.com/mcp"}}}'
ov mcps mount viking://agent/tools/weather/current.json
viking://resources/live/weather-beijing.md --title "Beijing live
weather" --args '{"city":"Beijing"}' --wait
ov read viking://resources/live/weather-beijing.md
```

— Names all four commands and shows a runnable mounting example v0.5.4

02 Group member removal endpoint changed BREAKING

90

The group member removal endpoint changed from `DELETE .../members/{user_id}` to `DELETE .../members` with request body `{"user_ids":["alice","bob"]}`. Direct callers of the old HTTP path must migrate; the CLI and Python/Go SDKs adapt automatically.

— Exact old and new endpoint paths and body shape given v0.5.4

03 CacheFS and Redis default configuration changes BREAKING

90

CacheFS removes `max_file_size_bytes`, `traversal_mode`, and `bypass_prefixes`; caching now covers filesystem query results. Redis defaults change to a 10-second

connection timeout, no command timeout, and a 300-second cache TTL, so deployments relying on the old defaults must set explicit values.

— Names removed keys and exact new default values v0.5.4

04 **Snapshot restore requires full commit IDs** BREAKING 85

`POST /api/v1/snapshot/restore` single-file mode now requires `restore_commit_id` and `current_commit_id` as full 40-character commit IDs; version UUIDs are no longer accepted.

— Exact endpoint and required field names and format v0.5.4

05 **Memory consolidation command** NEW 80

Consolidate existing memories for the current user with the new `ov compile memory` command, optionally passing `--instruction` to control how memories are merged (e.g. retaining distinct per-entity facts while merging duplicate observations).

Deduplicate and consolidate a user's accumulated memories before a new project phase, passing a custom consolidation instruction.

```
$ ov compile memory --instruction "Retain distinct per-entity facts; merge duplicate observations"
```

— Exact command and flag with a working example v0.5.4

06 **Noul reranking candidate visibility change** BREAKING 80

Noul reranking now isolates candidates by default; set `candidates_visible=true` to restore the previous shared-candidate visibility behavior. Choice mode is unaffected by this change.

— Exact flag and scope of the behavior change named v0.5.4

07 **Lock lease protocol change requires coordinated upgrade** 75

BREAKING

The lock lease protocol changed, so old and new deployments must not be mixed on shared storage: all lock participants (including queue workers) must be stopped before deploying. SQLite automatically adds a `holders` column, and persisted queue handoffs must be retained during the upgrade.

— Gives concrete migration steps and schema change v0.5.4

08 **Optional Kafka reporting and inventory export** NEW 65

Enable optional Kafka events and inventory export by installing `openviking[reporting]` and explicitly enabling reporting; this preserves deployment, workspace, user, and model-cost ownership data without exposing source text, prompts, or API keys.

— Names install flag but no endpoint or schema detail v0.5.4

09 **Distributed MySQL-backed Watch workloads** NEW 60

Distribute Watch workloads across nodes with optional MySQL storage, paginated listing, and automatic pause on connector credential revocation; reauthorization resumes only the Watches that were paused due to credential revocation.

— Explains mechanism but no named config keys or commands v0.5.4

THINNER COVERAGE BELOW

10 **Rule Search entry point in agent integrations** NEW 50

A new Rule Search entry point was added across multiple agent integrations, letting agents query applicable rules before taking an action; it activates by updating the plugin or reconnecting the MCP client.

— Describes behaviour but no concrete command or endpoint v0.5.4

11 **Agent Evolution and legacy Account settings SDK methods removed** DEPRECATED 50

Agent Evolution and old Account settings methods are removed from the Python and Go SDKs; callers must migrate to the Account configuration HTTP API, though deprecated server-side HTTP adapters remain in this release.

— Names affected SDKs but not the replacement endpoint path v0.5.4

12 **GLOBAL identities and data-owner access policies** NEW 45

Configure ROOT-managed GLOBAL identities and data-owner read access policies, alongside account-level extraction toggles, generation language settings, and parser credentials.

— Lists settings but no command, endpoint, or config path v0.5.4

BREAKING ON UPGRADE

! Do not mix old and new deployments on shared storage: the lock lease protocol changed. Stop all lock participants (including queue workers) before deploying. SQLite adds the `holders` column automatically; retain persisted queue handoffs.

! Group member removal endpoint changed from `DELETE .../members/{user_id}` to `DELETE .../members` with request body `{"user_ids":["alice","bob"]}`. Direct callers of the old HTTP path must migrate; the CLI and Python/Go SDKs now adapt automatically.

- ! `POST /api/v1/snapshot/restore` single-file mode now requires `restore_commit_id` and `current_commit_id` (full 40-character commit IDs); version UUIDs are no longer accepted.
- ! Agent Evolution and old Account settings methods are removed from the Python and Go SDKs; migrate to the Account configuration HTTP API. Deprecated server-side HTTP adapters remain in this release.
- ! CacheFS removes `max_file_size_bytes`, `traversal_mode`, and `bypass_prefixes`; caching now covers filesystem query results. Redis defaults change to a 10-second connection timeout, no command timeout, and a 300-second cache TTL — deployments relying on the old defaults must set explicit values.
- ! Noul reranking now isolates candidates by default; set `candidates_visible=true` to retain the previous shared-candidate behavior. Choice mode is unaffected.

WAS THIS USEFUL? 👍 👎

◆ Browser & Computer Use

browser-use

Feedback ⓘ 1.0.0 NOTES ↗ RANK

The browser-use Python library automates web browsers for AI agents.

> browser-use 1.0 rewrites the agent core as a persistent Python REPL over raw CDP, shipping typed structured output, secrets handling, file upload/download, domain restrictions, web search, unified authentication, CLI/MCP tooling and official Docker images — alongside sweeping breaking changes to browser connection defaults, LLM configuration, run limits, result/error APIs and package extras.

└──▷ WHAT SHIPPED · 14 FEATURES most completely described first ? what's the number?

01 Browser connection modes and defaults

BREAKING

90

Browser() now connects to a Browser Use Cloud browser by default; use Browser('chromium') for a local browser, Browser('chrome') to attach to an installed Chrome, or Browser('chrome', profile='Work') to pick a profile. Browser(local=True) is removed in favor of Browser('chromium'), and Browser('chrome') now only attaches to a browser whose live `DevToolsActivePort` is found in your own profile dirs, not whatever answers on `127.0.0.1:9222`.

— Explains before/after behavior with exact constructor forms. 1.0.0

02 LLM provider configuration changes

BREAKING

90

Route the agent through any compatible provider with a Pydantic AI model string, e.g. `llm='groq:llama-3.3-70b-versatile'`, on Agent(...). The `llm=` argument is no longer required — omitting it uses `bu-fast` on the Browser Use gateway, which requires `BROWSER_USE_API_KEY`. LangChain chat models now raise `TypeError`; Azure reads `AZURE_OPENAI_API_KEY` (not `AZURE_OPENAI_KEY`) and requires `OPENAI_API_VERSION`. 0.x gateway models (`bu-2-0`, `bu-1-0`, `bu-latest`, ...) are rejected, and `BROWSER_USE_LLM_URL` is no longer read — use ChatBrowserUse(base_url=...) instead.

— Enumerates every model string, env var and before/after case. 1.0.0

03 **Result and error reporting changes** BREAKING 90

is_successful() now returns `None` when the agent gave no verdict — use `r.is_done() and r.success is not False` for the previous behavior. errors() now returns only the error that ended the run, with recovered cell errors moved to `result.messages`. ActionResult(include_in_memory=...) now warns and is ignored: every tool result stays in the conversation.

— Gives exact migration code and behavior for each API. 1.0.0

04 **Secrets handling with type_secret()** NEW 85

Pass secrets into a run with `sensitive_data=` typed via type_secret(), which redacts values from cell output, checkpoints, tool results, and the final result.

Keep a login session out of logs and results when the task needs a password.

python

```python
from browser_use import Agent, Browser
from browser_use import type_secret

result = Agent(
    'Log in to example.com with username admin and password
SECRET and export the report',
    browser=Browser('chromium'),
    sensitive_data={'SECRET': type_secret('hunter2')},
).run_sync()
print(result.final_result())
```

— Named API with a runnable example of redaction. 1.0.0

05 **File uploads and downloads via workspace** NEW 85

Set a `workspace=` path on Agent(...) to handle file uploads (the agent reads files you name in the task) and downloads (files land in `workspace/downloads` and are listed in `result.downloads`).

Collect files the agent downloads during a run, with upload inputs pre-staged in a workspace directory.

```python
import pathlib
from browser_use import Agent, Browser

workspace = pathlib.Path('/tmp/my-agent-workspace')
workspace.mkdir(exist_ok=True)
(workspace / 'query.csv').write_text('id,name\n1,Alice')

result = Agent(
    'Upload query.csv to example.com/upload, run the report, and
download the result',
    browser=Browser('chromium'),
    workspace=workspace,
).run_sync()

for f in result.downloads:
    print(f)
```

— Named config key with a runnable example of both directions. 1.0.0

06 **Domain restrictions and IP blocking** BREAKING 80

Restrict or block domains per run with `allowed_domains`, `prohibited_domains`, and `block_ip_addresses` on Agent(...); `prohibited_domains` now also blocks subdomains, and setting any domain rule also blocks all `file:` URLs. The old `BROWSER_USE_ALLOWED_DOMAINS` environment variable is no longer read — Browser() raises while it is set.

— Names every param and env var plus the behavior change. 1.0.0

07 **Package extras and migration guardrails** DEPRECATED 80

Extras `aws`, `oci`, `video`, `examples`, `eval`, `core`, `openai-tools`, and `cli-oci` are removed; pip warns and installs without them. Every 0.x public import now raises an `ImportError` naming the replacement and linking `MIGRATING.md`. Removed 0.x options (`use_vision=True`, `flash_mode`, `max_failures`, etc.) warn once and are ignored; unrecognized options raise `TypeError`.

— Lists every removed extra and option by name. 1.0.0

08 **CLI entry point removals and MCP deprecation** BREAKING 75

`browser-use init` was removed and the `bu` and `browser` console scripts are gone — use `browser-use` or `browseruse` instead. `browser-use --mcp` now prints a

deprecation notice and serves only the `--cli-mcp` tools (`browser_exec` , `browser_screenshot` , `browser_login`).

— Names removed commands and the replacement tool set. 1.0.0

09 **Typed structured output from runs** NEW 70

Pass `output_model_schema=MyModel` to Agent(...) and read `result.structured_output` to get typed, schema-validated results from a run instead of free-form text.

Return structured data from a web task — useful when downstream code needs typed fields rather than free-form text.

python

```
from pydantic import BaseModel
from browser_use import Agent

class RepoStats(BaseModel):
    stars: int
    forks: int

result = Agent(
    'Find the star and fork counts of the browser-use GitHub
repo',
    output_model_schema=RepoStats,
).run_sync()
print(result.structured_output)  # RepoStats(stars=..,
forks=...)
```

— Names the exact params with a runnable example. 1.0.0

10 **CLI and MCP tooling additions** NEW 65

Register the agent as a skill with `browser-use skill install` ; run the MCP server with `browser-use --cli-mcp` ; pipe Python to browser-harness with `browser-use` .

— Names three commands but no example output shown. 1.0.0

11 **Official Docker images** NEW 65

Pull official Docker images at `browseruse/browseruse:1.0.0` (Docker Hub) or `ghcr.io/browser-use/browser-use:1.0.0` (GHCR), smoke-tested on amd64 and arm64.

— Exact image tags given but no run command shown. 1.0.0

12 **Unified authentication via BROWSER_USE_API_KEY** NEW 60

Authenticate with a single `BROWSER_USE_API_KEY` environment variable (or `browser-use login`) covering model calls, cloud browsers, web search, CLI, and the MCP server.

— Names env var and command but no walkthrough. 1.0.0

13 **Run step and time limit defaults** BREAKING 60

run() now defaults to `max_steps=100` (down from 500) and adds a `timeout=1700` s limit.

— States exact old and new defaults, no usage example. 1.0.0

THINNER COVERAGE BELOW

14 **Web search within agent runs** NEW 55

Enable web search within a run with `web_search=True` on Agent(...); search charges are tracked per search and added to `result.cost_usd`.

— Flag and cost field named but no example shown. 1.0.0

└──▷ ALSO FROM THESE RELEASES

└──▷ BREAKING ON UPGRADE

! Browser() now connects to a Browser Use Cloud browser by default; use Browser('chromium') for a local browser.

! Browser(local=True) is now Browser('chromium'); Browser('chrome') attaches only to a browser whose live `DevToolsActivePort` is in your own profile dirs, not whatever answers on 127.0.0.1:9222.

! run() now defaults to `max_steps=100` (was 500) and adds a `timeout=1700` s limit.

! The `llm=` argument is no longer required; omitting it uses `bu-fast` on the Browser Use gateway, which requires `BROWSER_USE_API_KEY`.

! LangChain chat models now raise `TypeError`; Azure reads `AZURE_OPENAI_API_KEY` (not `AZURE_OPENAI_KEY`) and requires `OPENAI_API_VERSION`.

! 0.x gateway models (`bu-2-0`, `bu-1-0`, `bu-latest`, ...) are rejected; `BROWSER_USE_LLM_URL` is not read — use ChatBrowserUse(base_url=...).

! `BROWSER_USE_ALLOWED_DOMAINS` is not read; Browser() raises while that variable is set.

! `browser-use init` was removed; the `bu` and `browser` console scripts are gone — use `browser-use` or `browseruse` instead.

! `browser-use --mcp` now prints a deprecation notice and serves only the `--cli-mcp` tools (`browser_exec`, `browser_screenshot`, `browser_login`).

! Extras `aws`, `oci`, `video`, `examples`, `eval`, `core`, `openai-tools`, and `cli-oci` are removed; pip warns and installs without them.

! is_successful() returns `None` when the agent gave no verdict; use `r.is_done() and r.success is not False` for the previous behavior.

! errors() now returns only the error that ended the run; recovered cell errors are in `result.messages`.

! ActionResult(include_in_memory=...) warns and is ignored: every tool result stays in the conversation.

! Every 0.x public import raises an `ImportError` naming the replacement and linking `MIGRATING.md`.

! Removed 0.x options (`use_vision=True`, `flash_mode`, `max_failures`, etc.) warn once and are ignored; unrecognized options raise `TypeError`.

! `prohibited_domains=['example.com']` now also blocks subdomains; any domain rule set also blocks all `file:` URLs.

WAS THIS USEFUL? 👍 👎

playwright-mcp

Feedback ⓘ v0.0.85 NOTES ↗ RANK

Playwright MCP server

> playwright-mcp v0.0.85 adds a background mode for extension-connected browsers, a tab-group opt-out flag, strict visible-only locator matching, and accessibility-tree refinements for snapshots and custom form elements.

└──▷ **WHAT SHIPPED · 5 FEATURES** most completely described first ? what's the number?

01 Background mode for extension connections NEW

86

A new `--background` CLI flag (also settable via `background: true` in the config file or `PLAYWRIGHT_MCP_BACKGROUND=1`) keeps the browser window unfocused and new tabs inactive while connected via `--extension`, requiring extension 0.5.0 or later and Chrome 158.0.8127.2 or later.

Run the MCP server with an existing Chrome extension without stealing focus from the app you are working in.

```
$ playwright-mcp --extension --background
```

— Names flag, config key, env var and version requirements v0.0.85

02 Visible-only locator restriction NEW

86

Setting `browser.contextOptions.locateOnlyVisible: true` in the config file restricts all locators — including `target` values in tools such as `browser_click` and code run via `browser_run_code_unsafe` — to visible elements only, avoiding accidental clicks on hidden duplicate elements.

Avoid accidental clicks on hidden duplicate elements by restricting all locators to visible elements only.

```yaml
browser:
  contextOptions:
    locateOnlyVisible: true
```

— Named config key, affected tools, and concrete use case v0.0.85

THINNER COVERAGE BELOW

03 Form-associated custom element labeling IMPROVED

57

Element tools now recognize form-associated custom elements through getByLabel() targets, honoring both `for` attributes and nested labels.

— Names mechanism and API but lacks a usage example v0.0.85

04 **Improved element naming in browser_snapshot** IMPROVED

55

`browser_snapshot` now retains identifying names on containers (dialogs, landmarks) and table headers when their heading or caption text also appears among their children.

— Describes exact behavior change but no action steps v0.0.85

05 **Tab-group opt-out for extension tabs** NEW

53

The `--no-tab-group` CLI flag keeps automation tabs in their existing tab groups instead of moving them into the Playwright tab group when running with `--extension`.

Keep automation tabs in their existing tab groups rather than moving them into the Playwright group.

```
$ playwright-mcp --extension --no-tab-group
```

— Named flag with clear purpose but no deeper mechanism v0.0.85

WAS THIS USEFUL? 👍 👎

◆ Applied AI Agents

HyperResearch

Feedback ⓘ v0.13.1 NOTES ↗ RANK

HyperResearch collects and synthesizes web research into a persistent searchable wiki for AI agents.

> HyperResearch v0.13.1 adds a keyless Keenable web search provider, tightens install safety with a required `--create` flag, and speeds up semantic search with a single numpy matmul, alongside sync and cleanup fixes for note id edits, asset removal, and import diagnostics.

└──▷ WHAT SHIPPED · 6 FEATURES most completely described first ? what's the number?

01 Keenable web search provider in vault config NEW

90

Adds a Keenable web provider — keyless web search with page text and page fetch — enabled by setting `[web] provider = "keenable"` in vault config; setting `KEENABLE_API_KEY` lifts per-IP rate limits.

Set the Keenable provider in your vault config so all web searches use it by default.

```ini
[web]
provider = "keenable"
```

— Exact config key and env var given, fully actionable v0.13.1

02 Explicit --create flag required for install BREAKING

90

`install <path>` and `install --steps-only` now refuse a path that does not yet exist, failing with `PATH_NOT_FOUND` instead of silently creating an empty vault; scripts that install into a self-created directory must pass `--create`.

Install into a new directory in a CI script without being refused by the missing-path check.

```
$ hyperresearch install ./research-vault --create
```

— Exact flag, error code, and migration note given v0.13.1

03 Faster semantic search scoring IMPROVED

75

`hpr search --semantic` now scores all candidate notes in a single numpy matmul via `np.frombuffer`, replacing one cosine() call per vector; measured sub-second up to

~50k notes.

— Mechanism and benchmark given, command named `v0.13.1`

---

04 **Sync handles note id renames** IMPROVED 65

Sync now moves a note's database row to its new id when the `id:` frontmatter key is edited, keeping claims, embeddings, assets, `sources`, and `escalations` rows in place.

— Behavior and affected rows named, no example `v0.13.1`

---

05 **Precise asset cleanup in note rm** IMPROVED 65

`note rm` now removes asset files by consulting the note's `assets` rows anywhere under `research/assets/`, rather than deleting the whole `research/assets/<id>/` directory.

— Before/after behavior and path named, no example `v0.13.1`

---

06 **Skipped and not-indexed keys in import output** NEW 60

`import` adds `skipped` and `not_indexed` keys to its output, listing files skipped with `DESTINATION_EXISTS` or not indexed with `NOTE_NOT_INDEXED` respectively.

— Exact keys and error codes named, no usage example `v0.13.1`

╰──▷ BREAKING ON UPGRADE

! `hyperresearch install <path>` and `install --steps-only` now refuse a path that does not exist, returning `PATH_NOT_FOUND`; scripts that install into a directory they create themselves must add `--create`.

WAS THIS USEFUL? 👍 👎

GPT Researcher

Feedback ⓘ v3.8.0 NOTES ↗ RANK

GPT Researcher runs autonomous research workflows over web and local sources and produces cited reports.

> GPT Researcher v3.8.0 cuts detailed report generation time from roughly 9:11 to 3:27 by parallelizing section research, adds a research lifecycle observer, prebuilt multi-arch Docker images and a Render deployment blueprint, and ships a redesigned Next.js UI — alongside several breaking changes to the default writer model, Docker secret handling, and failure behaviour of `generate_report`.

└──▷ **WHAT SHIPPED · 11 FEATURES** most completely described first ? what's the number?

01 Docker images no longer bundle root .env file

BREAKING

95

The root `.env` file is no longer copied into Docker images; `docker compose` now loads it at runtime (requires Compose 2.24+), and `docker run` requires `--env-file .env`. Locally built images that were previously pushed may have contained secrets, so any keys that were in `.env` should be rotated.

— Exact commands, version requirement, and remediation step given v3.8.0

02 Runtime overrides for document path and report source

NEW

85

Override document path and report source at runtime with the `DOC_PATH` and `REPORT_SOURCE` environment variables, letting the same image target different data sets without editing config files.

Point the researcher at a local document folder and fix the report source without editing config files — useful when running the same image against different data sets.

```
$ export DOC_PATH=/data/classified-docs
export REPORT_SOURCE=local
python cli.py "Key findings in the Q1 audit" --report_type
detailed_report
```

— Names both env vars and includes a runnable example v3.8.0

03 Prebuilt Docker images and Render deployment blueprint

NEW

85

Adds new deployment options: prebuilt multi-arch Docker images for `linux/amd64` and `linux/arm64` published to `ghcr.io/assafelovic/gpt-researcher` and

`ghcr.io/assafelovic/gpt-researcher-frontend-nextjs`, tagged by version and `latest`, plus a `render.yaml` blueprint added to the repository for deploying to Render.

— Names exact registries, architectures, tags and blueprint file v3.8.0

04 **Default writer model changed to gpt-6-sol** BREAKING 80

The default writer model `SMART_LLM` is now `openai:gpt-6-sol` (previously `openai:gpt-5.4`); set `SMART_LLM` explicitly to keep the previous model.

— Names exact config key, old and new values, and mitigation v3.8.0

05 **Multi-agent writer output and reasoning controls** NEW 65

The multi-agent writer now reads `SMART_TOKEN_LIMIT` and `REASONING_EFFORT` environment variables to control output size and reasoning depth.

— Names both env vars without a usage example v3.8.0

THINNER COVERAGE BELOW

06 **generate_report raises on total LLM failure** BREAKING 55

`generate_report` now raises an exception when every LLM attempt fails, instead of silently returning an empty report.

— Names the function and the before/after behaviour change v3.8.0

07 **Research lifecycle observer for progress events** NEW 55

Adds an optional research lifecycle observer for integrations that need structured progress events, enabled by passing an observer to GPTResearcher.

— Names the API surface but gives no code example v3.8.0

08 **Faster detailed reports via parallel section research** IMPROVED 50

Detailed reports now finish in roughly 3:27 by researching sections in parallel, down from roughly 9:11.

— Concrete before/after timings but nothing to configure v3.8.0

09 **Redesigned Next.js UI with classic fallback** IMPROVED 45

A new paper-design Next.js UI becomes the default interface; the previous UI remains available at `/classic`.

— Exact fallback path given, mechanism of redesign unspecified v3.8.0

10 **MiniMax regional provider and Flash model support** NEW 30

MiniMax regional providers and Flash model defaults are now supported.

— Bare mention with no setup detail v3.8.0

11 **Keenable and Sciverse retriever plugins** NEW 30

Retriever plugins now include community-contributed Keenable and Sciverse plugins.

— Named but no usage or configuration detail given v3.8.0

└──▷ ALSO FROM THESE RELEASES

Cap how long a stalled LLM call can run before it fails with a clear error — useful in CI or batch pipelines where a hanging call would block indefinitely.

```
$ export LLM_TIMEOUT=120
python cli.py "Summarise recent CVEs in OpenSSL" --report_type
research_report
```

└──▷ BREAKING ON UPGRADE

! The default writer model `SMART_LLM` is now `openai:gpt-6-sol` (was `openai:gpt-5.4`); set `SMART_LLM` explicitly to keep the previous model.

! The root `.env` file is no longer copied into Docker images; `docker compose` loads it at runtime (requires Compose 2.24+), and `docker run` requires `--env-file .env`. Locally built images that were pushed may have contained secrets — rotate any keys that were in your `.env`.

! `generate_report` now raises an exception when every LLM attempt fails instead of returning an empty report.

WAS THIS USEFUL? 👍 👎

◆ MCP Servers & Tooling

godot-mcp

Feedback ⓘ v0.38.0 NOTES ↗ RANK

Godot-MCP — Model Context Protocol (MCP) integration for the Godot Engine. dev.

> Godot-MCP's screenshot tool can now save captures directly into the project via an optional path parameter.

└──▷ WHAT SHIPPED · 1 FEATURE most completely described first ? what's the number?

01 **Optional save path for editor screenshots** NEW 60

The screenshot capture tool accepts an optional `savePath` parameter, writing the captured image to disk at the specified location within the project instead of only returning it in-memory.

— Names the exact parameter but no further mechanism detail v0.38.0

WAS THIS USEFUL? 👍 👎

magic

Feedback ⓘ v24.4.19 NOTES ↗ RANK

Magic is an instant SECURE Full Stack Apps and AI Agents

> Magic's latest release focuses on Chat Ops improvements, adding an attachment workflow and clearer AI tool contracts, alongside tighter sandbox enforcement for slots.

└──▷ **WHAT SHIPPED · 3 FEATURES** most completely described first ? what's the number?

01 **Slot-level whitelist pinning** NEW 45

Adds a slot-level declaration that forces a whitelist to pin any slot that uses it, tightening sandbox enforcement.

– Explains mechanism but no concrete syntax or config key named v24.4.19

02 **Attach menu in Chat Ops** NEW 40

Chat Ops now includes an Attach menu that lets users choose between Upload and Analyse actions when adding content to a conversation.

– Names UI feature but no exact navigation path or mechanism v24.4.19

03 **Required AI tool parameters in Chat Ops** IMPROVED 35

Chat Ops AI tool parameters now list mandatory arguments as required, improving agent-facing API contracts.

– States the change but no schema or example given v24.4.19

WAS THIS USEFUL? 👍 👎

tabularis

Feedback ⓘ v0.32.0 NOTES ↗ RANK

Tabularis is an open-source desktop SQL workspace with 3 built-in database drivers and 29 shipped driver plugins, including SQL Server, DuckDB, ClickHouse and Redis. Built-in MCP server for Claude, Cursor and Devin, SQL notebooks and visual EXPLAIN.

> Tabularis v0.32.0 adds table transfer between same-driver connections, plan snapshot comparison in the Explain view, conditional row/cell formatting and clipboard-based filtering in the data grid, and PostgreSQL-specific query timeouts and SQL diagnostics, alongside smaller notebook, JSON viewer and text viewer improvements.

└──▷ **WHAT SHIPPED · 10 FEATURES** most completely described first ? what's the number?

01 **Data grid formatting and filtering** NEW 75

Adds conditional formatting to color rows and cells in the data grid, letting users define a rule (e.g. value thresholds) and a highlight color via the column menu's Conditional Formatting option to flag matching rows at a glance, and adds a grid context menu option to filter a column by clipboard value, supporting `equals clipboard` or IN clipboard matching.

Highlight slow or anomalous rows at a glance without writing extra queries.

📍 In the data grid, open the column menu, select Conditional Formatting, define a rule (e.g. value › 1000), choose a highlight color, and confirm to color matching rows and cells.

— Mechanism and UI steps given via example Named operators but no usage walkthrough

v0.32.0

02 **Plan snapshot comparison in Explain view** NEW 70

Lets users save query plan snapshots and compare them against a baseline in the Explain view, enabling before/after comparison of plans after changes such as adding an index.

Compare how a query plan changed after adding an index — save a baseline snapshot, then run the query again and diff the two plans side by side.

📍 In the Explain view, run your query, then choose Save Snapshot to capture the current plan. After schema changes, run the query again and select Compare Against Baseline to diff the two plans.

— Clear mechanism and runnable steps, but limited scope detail v0.32.0

03 **Explorer sidebar table transfer and schema actions** NEW

65

Adds explorer sidebar actions to export a table to SQL and import it into a same-driver connection, plus new `New Database` and `New Schema` actions.

Move a table from one database to another that uses the same driver without leaving Tabularis.

> 📍 In the explorer sidebar, right-click a table, choose `Export to SQL`, select a same-driver connection as the destination, and confirm the import.

— Named UI actions with steps, but no mechanism beyond same-driver limit v0.32.0

THINNER COVERAGE BELOW

04 **PostgreSQL query timeout and SQL diagnostics** NEW

55

Adds PostgreSQL-specific capabilities: a per-connection query timeout with server-side cancel, and opt-in metadata-aware SQL diagnostics for PostgreSQL `SELECT` statements in the editor.

— Mechanism named but no config key or exact setting location v0.32.0

05 **Notebook toolbar export and clear outputs actions** NEW

55

Adds notebook toolbar actions to export a SQL cell's source to a `.sql` file and to clear all cell outputs via a new `Clear Outputs` action.

— Named actions but no usage steps shown v0.32.0

06 **Multi-database plugin database selection in connection dialog**

IMPROVED

40

Lets multi-database plugins accept a database on the Databases tab in the connection dialog.

— Thin description, names the dialog tab only v0.32.0

07 **Query history duration filter** NEW

35

Adds ability to filter query history entries by minimum recorded duration.

— One-line addition with no further detail v0.32.0

08 **Text viewer character, line and byte counts** NEW

30

Shows character, line, and UTF-8 byte counts in the text viewer.

— Names the metrics but no further mechanism v0.32.0

09 **SQL editor status bar cursor and selection info** NEW

25

Shows cursor position and selection size in the SQL editor status bar.

— Bare description of a small UI addition v0.32.0

10 **JSON viewer alphabetical key sorting** IMPROVED 25

Sorts object keys alphabetically in read-only JSON trees in the JSON viewer.

— Minimal one-line description v0.32.0

└──▷ ALSO FROM THESE RELEASES







WAS THIS USEFUL? 👍 👎

context7

Feedback ⓘ `@upstash/context7-mcp@4.6.0` NOTES ↗ RANK

Context7 maintains up-to-date documentation and context for AI agents through an MCP server.

> Context7 shipped two new research-preview MCP tools: `query-code` for searching a library's source code at the exact documented version, and `get-upgrade-guide` for cross-version upgrade diffs.

└──▷ **WHAT SHIPPED · 2 FEATURES** most completely described first ? what's the number?

01 **Source-code search via query-code tool** NEW 91

The new `query-code` MCP tool and `ctx7 code` command let users search and read a library's source code at the exact version its documentation describes, using ripgrep-style subcommands `grep`, `definition`, `references`, `outline`, `read`, and `list`. Enable it via `--research-preview`, `CONTEXT7_RESEARCH_PREVIEW=1`, or `?research_preview=1` on the MCP URL.

— Names tool, command, six subcommands, and three enable methods

`@upstash/context7-mcp@4.6.0` `ctx7@0.11.0`

THINNER COVERAGE BELOW

02 **Cross-version upgrade guide tool** NEW 55

The new `get-upgrade-guide` research preview MCP tool returns everything that changed between two versions of a library in one response — release notes, guides, and a public API diff.

— Describes what it returns but no enable flag given `@upstash/context7-mcp@4.6.0`

WAS THIS USEFUL? 👍 👎

mcpvault

Feedback ⓘ CVE-2026-57442 CVE ↗ STABLE

Mcpvault is a lightweight Model Context Protocol (MCP) server for safe Obsidian vault access

`CVE-2026-57442`: PathFilter restricted directories (.git/.obsidian/node_modules) only denied at vault root, not nested (affects `< 0.11.5`)

▾ **1 more release in this issue** · 2026-06-17 → 2026-06-18

> CVE-2026-57441 CVE ↗ STABLE
>
> `CVE-2026-57441`: PathFilter restricted-directory deny-list bypass via case and trailing dot/space equivalence (affects `< 0.11.4`)

WAS THIS USEFUL? 👍 👎

◆ AI Agent Communication

AG-UI

Feedback ⓘ release/2026-11-09 NOTES ↗ RANK

AG-UI: the Agent-User Interaction Protocol. Bring Agents into Frontend Applications.

> AG-UI's ag_ui_strands package adds fine-grained subagent visibility controls with lifecycle event emission for agents-as-tools, and introduces an opt-out for forced A2UI tool choice across both its TypeScript and Python packages.

└──▷ WHAT SHIPPED · 2 FEATURES most completely described first ? what's the number?

01 **Subagent visibility control in ag_ui_strands** NEW 88

Set `StrandsAgentConfig.subagent_visibility` to `inline` (default), `attributed`, or `hidden` in `ag_ui_strands` to control how subagent events surface in agents-as-tools. With `attributed`, sub-agents emit `SUBAGENT_STARTED`, `SUBAGENT_FINISHED`, and `SUBAGENT_ERROR` lifecycle events and carry `subagentRunId` on text, reasoning, and tool call events, letting the UI track each sub-agent's progress and errors separately.

Track each sub-agent's lifecycle separately when using agents-as-tools, so the UI can display per-subagent progress and errors.

python

```python
from ag_ui_strands import StrandsAgentConfig

config = StrandsAgentConfig(
    subagent_visibility="attributed"
)
# Sub-agents will now emit SUBAGENT_STARTED, SUBAGENT_FINISHED,
# and SUBAGENT_ERROR events, each carrying a subagentRunId.
```

Hide sub-agent activity entirely from the UI when you want agents-as-tools to appear as opaque steps to end users.

python

```python
from ag_ui_strands import StrandsAgentConfig

config = StrandsAgentConfig(
    subagent_visibility="hidden"
)
```

— Names config key, all three values and event names, with runnable code

release/2026-11-09

THINNER COVERAGE BELOW

02 **Opt-out of forced A2UI tool choice** IMPROVED 45

Opt out of forced A2UI tool choice in both `@ag-ui/aws-strands` (TypeScript) and `ag_ui_strands` (Python), while forced choice is preserved when config is null.

— Names both packages but no exact flag or config key given release/2026-11-09

WAS THIS USEFUL? 👍 👎

DEPLOY

◆ ML Platforms & SDKs

Hugging Face

Feedback ⓘ changelog · Nov 10, 2026 CHANGELOG ↗

RANK

Hugging Face is an open platform for building, sharing, and deploying machine-learning models, datasets, and applications.

> Hugging Face replaced its main search bar with a keyboard-driven command palette for quickly navigating to models, datasets, and Spaces.

└──▷ WHAT SHIPPED · 1 FEATURE most completely described first ? what's the number?

01 **Command palette replaces search bar** NEW 75

The main search bar is replaced by a command palette opened with `⌘K` (macOS) or `Ctrl+K` (Windows/Linux). Users navigate results with arrow keys and press Enter to jump directly to any model, dataset, or Space, or use quick actions to create a new one.

Jump to any model, dataset, or Space without leaving the keyboard.

> 📍 Press ⌘K (macOS) or Ctrl+K (Windows/Linux) to open the command palette, type the name of the model or dataset, navigate with the arrow keys, and press Enter to go directly to it.

— Names exact shortcut and interaction, but limited mechanism detail beyond that.

changelog-20261110-fc3f0dc2

WAS THIS USEFUL? 👍 👎

OpenAI TypeScript SDK

Feedback ⓘ v7.55.0 NOTES ↗ RANK

The OpenAI TypeScript SDK provides typed JavaScript and TypeScript clients for OpenAI APIs and streaming responses.

> The OpenAI TypeScript SDK added typed async function tools and recognized the gpt-6-rosalind model in the beta Responses API.

└──▷ **WHAT SHIPPED · 2 FEATURES** most completely described first ? what's the number?

01 **gpt-6-rosalind model recognition** IMPROVED 40

The SDK now recognizes `gpt-6-rosalind` as a valid model in the beta Responses API.

— Names model and API but minimal mechanism detail v7.54.0

02 **Typed async function tools** NEW 35

Adds typed async function tools to the API, enabling type-safe tool definitions for use with OpenAI models.

— Describes capability but no concrete API names or usage v7.55.0

WAS THIS USEFUL? 👍 👎

Hugging Face Datasets

Feedback ⓘ 5.2.0 NOTES ↗ RANK

Hugging Face Datasets datasets is a Python library for loading, streaming, and processing datasets for machine learning.

> Hugging Face Datasets 5.2.0 adds fast streaming decode for media folder datasets and extends 3D mesh support to `.gltf` files.

└──▷ WHAT SHIPPED · 2 FEATURES most completely described first ? what's the number?

01 Fast streaming decode for media folder datasets IMPROVED

70

Image, audio, video, and other media folder datasets can now be streamed with fast decode, whereas previously these required a full download before iteration. Enable via load_dataset('path/to/imagefolder', streaming=True) and iterate directly.

Stream a large image folder dataset without downloading it first, now benefiting from fast decode.

python

```
from datasets import load_dataset

dataset = load_dataset('path/to/imagefolder', streaming=True)
for example in dataset['train']:
    print(example['image'])
    break
```

— Names mechanism and before/after, with runnable command 5.2.0

02 .gltf support in 3D Mesh feature NEW

60

The 3D Mesh feature type and `meshfolder` dataset builder now accept `.gltf` files alongside the existing formats. Load such a dataset with load_dataset('path/to/meshfolder').

Load a 3D mesh dataset that includes `.gltf` files using the meshfolder builder.

python

```
from datasets import load_dataset

dataset = load_dataset('path/to/meshfolder')
print(dataset['train'][0])
```

— Names builder and format with runnable command, thin on mechanism 5.2.0

WAS THIS USEFUL? 👍 👎

AWS Amazon Bedrock AgentCore Feedback ⓘ

Web Search: Search the web from your code with the InvokeWebSearch API (2026-11-09)

BLOG ↗ RANK

Amazon Bedrock AgentCore provides managed agent runtimes, memory, identity, gateways, observability, and evaluation services.

> Amazon Bedrock AgentCore added a direct Web Search API for agents and application code, removing the need for a gateway to search the web.

└──▷ **WHAT SHIPPED · 1 FEATURE** most completely described first ? what's the number?

01 **InvokeWebSearch API for web search** NEW 70

Applications and agents can search the web directly from code using the new `InvokeWebSearch` API, accessible via the AWS SDKs or AWS CLI across six AWS Regions, with no gateway setup required.

— Names exact API and access paths but lacks params or limits detail

Web Search: Search the web from your co…

WAS THIS USEFUL? 👍 👎

Anthropic TypeScript SDK

Feedback ⓘ sdk-v1.7.0

NOTES ↗ RANK

The Anthropic TypeScript SDK provides typed JavaScript and TypeScript clients for Claude APIs, streaming, and tool use.

> The Anthropic TypeScript SDK's v1.7.0 release adds organization-level filtering for usage and cost reporting, a new webhook event, a new sandbox error type, and two additional model identifiers.

└──▷ **WHAT SHIPPED · 4 FEATURES** most completely described first ? what's the number?

01 **New user profile webhook event** NEW 50

Organization webhook endpoints can now subscribe to the `user_profile.program_status_updated` event.

— Exact event name given but no payload/schema detail sdk-v1.7.0

02 **New Claude model identifiers** NEW 45

Adds support for new model identifiers `claude-fable-5-2` and `claude-mythos-5-2`.

— Exact model IDs given, ready to use, but no further detail sdk-v1.7.0

03 **Parent organization filtering for usage/cost reports** NEW 40

Usage and cost reports can now be filtered and grouped by parent organization via the API.

— Names the capability but no endpoint or param given sdk-v1.7.0

04 **Sandbox egress restricted error** NEW 40

Admin API write methods can now return a new `sandbox_egress_restricted` error, which the SDK handles.

— Names the exact error but no handling mechanism shown sdk-v1.7.0

WAS THIS USEFUL? 👍 👎

Anthropic Python SDK

Feedback ⓘ v2.7.0 NOTES ↗ RANK

The Anthropic Python SDK provides typed Python clients for Claude APIs, streaming responses, and tool use.

> Anthropic Python SDK v2.7.0 adds parent-organization filtering for usage and cost reports, a new webhook event, a new Admin API error, and two new model identifiers.

└──▷ **WHAT SHIPPED · 4 FEATURES** most completely described first ? what's the number?

01 **New Claude model identifiers** NEW 55

Two new model identifiers, `claude-fable-5-2` and `claude-mythos-5-2`, can now be used in API calls.

— Exact model IDs given, ready to use directly v2.7.0

02 **New webhook event for program status updates** NEW 50

Organization webhook endpoints can now handle a new `user_profile.program_status_updated` event.

— Named event but no payload or setup detail v2.7.0

03 **New sandbox egress error from Admin API** NEW 45

Admin API write methods can now return a new `sandbox_egress_restricted` error.

— Named error but no cause or handling guidance v2.7.0

04 **Parent organization filtering for usage and cost reports** NEW 42

Usage and cost report endpoints in the API can now filter and group results by parent organization.

— No endpoint path or parameter name given v2.7.0

WAS THIS USEFUL? 👍 👎

OpenAI Python SDK

Feedback ⓘ v3.48.0 NOTES ↗ RANK

The OpenAI Python SDK provides typed Python clients for OpenAI model APIs, streaming, and asynchronous requests.

> OpenAI Python SDK v3.48.0 adds streaming support for agent sessions and recognizes a new beta model identifier in the Responses API.

└──▷ **WHAT SHIPPED · 2 FEATURES** most completely described first ? what's the number?

01 **gpt-6-rosalind model support in beta Responses API** NEW

45

The SDK recognizes `gpt-6-rosalind` as a valid model identifier for beta Responses API calls.

— Names exact model identifier but lacks usage example v3.48.0

02 **Stream options for agent sessions** NEW

40

The SDK now supports streaming agent session output by passing stream options when creating agent sessions.

— Describes capability but no exact parameter name or code shown v3.48.0

WAS THIS USEFUL? 👍 👎

◆ Inference Providers & GPU Clouds

Crusoe Cloud

Feedback ⓘ updated · Nov 10, 2026 NOTES ↗ RANK

Crusoe Cloud provides GPU infrastructure and managed inference for AI training and deployment.

> Crusoe Cloud shipped CMK v2 as the new cluster default with audit logging and auto-repair, plus enterprise identity controls for Google Workspace SSO enforcement and Okta SCIM provisioning.

└──▷ WHAT SHIPPED · 3 FEATURES most completely described first ? what's the number?

01 **CMK v2 clusters with audit logs and auto-repair** NEW

65

New clusters now default to CMK v2, which adds control plane audit logs, in-place edits to most node pool settings, and automatic node repair via AutoClusters. This is unavailable in the `us-southcentral1-a` and `us-east2-a` regions.

— Names mechanism and region exceptions but no config path or command snapshot-20261110

THINNER COVERAGE BELOW

02 **Google Workspace SSO enforcement** NEW

50

Organizations can enforce Google Workspace as their SSO provider, configured in the Crusoe Cloud console under IAM > User Authentication, after the account team enables SSO.

Set up Google Workspace SSO enforcement for your organization after your account team has enabled it.

> 📍 In the Crusoe Cloud console, go to IAM › User Authentication and configure Google Workspace as your organization's SSO provider.

— Gives exact UI navigation path to configure it snapshot-20261110

03 **Okta SCIM user provisioning** NEW

45

A new guide covers provisioning Crusoe Cloud users from Okta using SCIM, including connecting a SCIM app, mapping organization roles, and rotating SCIM API keys.

— Describes steps covered but no exact commands or keys named snapshot-20261110

WAS THIS USEFUL? 👍 👎

Eigen Labs Darkbloom

Feedback ⓘ v0.10.11 NOTES ↗

CODE ↗ RANK

Darkbloom is a private Inference Network on Idle Macs

> Darkbloom's v0.10.11 release adds a local-first serving mode that pauses network sharing while you use your Mac locally, wires in OpenCode 2 and Hermes Agent, adds new routing and pairing CLI commands, raises default token and context limits, and tightens request error handling with typed errors and early 503s.

└──▷ **WHAT SHIPPED · 9 FEATURES** most completely described first ? what's the number?

01 Local-first serving pauses network sharing BREAKING

95

Configure local-first serving with `me_first_idle_minutes` (1–120) and opt out with `me_first = false` under `[local]` in `~/.config/darkbloom/provider.toml`, then apply with `darkbloom restart`. Macs serving with `--local-endpoint` (including the desktop app's Alongside the network) now pause Darkbloom while you use their local API and resume 2 minutes after your last request.

Keep a Mac sharing Darkbloom network requests at all times, disabling the new local-first pause behaviour.

toml

```toml
# In ~/.config/darkbloom/provider.toml
[local]
me_first = false
```

— Full config path, key, values and restart command all given. v0.10.11

02 Typed 400 error for invalid reasoning_effort BREAKING

85

Invalid `reasoning_effort` values now return a typed 400 error (`invalid_reasoning_effort`) with the field name in `param` and the accepted names; accepted values are `none`, `minimal`, `low`, `medium`, `high`, `xhigh`, and `max`. This refusal now applies on every reasoning model, including Qwen 3.5/3.6, Bonsai 2, Gemma 4, GPT-OSS and Nemotron Lightning, which previously accepted it silently.

— Error code, param field and accepted values all named. v0.10.11

03 OpenCode and Hermes agent integrations NEW

80

Set up OpenCode 2 or Hermes Agent to use this Mac's models with `darkbloom --opencode` or `darkbloom --hermes` (also `darkbloom start --opencode` / `--hermes`); Hermes requires v0.20.2 or later.

Point OpenCode 2 at this Mac's local models right after installing Darkbloom.

```
$ darkbloom --opencode
```

— Exact flags and version requirement given, runnable directly. v0.10.11

04 **Routing and pairing commands** NEW 80

Save request routing preference with `darkbloom route this-mac`, `my-macs`, or `my-macs-first`; save an account key for the paid network with `darkbloom route key`. Re-run the serving-mode setup interactively with `darkbloom start --choose`. Leave a two-Mac pair from either Mac in one command with `darkbloom cluster leave`.

Route your own requests to this Mac first, fall back to your other Macs, and never touch the paid network.

```
$ darkbloom route my-macs
```

— All subcommands named verbatim and directly runnable. v0.10.11

05 **Higher default max_tokens ceiling** IMPROVED 75

Requests with no `max_tokens` can now produce up to 32,768 tokens (previously capped at 8,192, or 4,096 on MiMo V2.6, Nemotron Super and GLM-5.3-Flash); oversized `max_tokens` is cut to the model's window and completes with `finish_reason: length`.

— Exact before/after numbers and affected models named. v0.10.11

06 **Gemma 4 full 262k context window** IMPROVED 70

Gemma 4 pairs in tensor-parallel mode now serve the model's full 262,144-token context window (prompts up to 262,143 tokens, answers up to 32,768 tokens).

— Exact token limits named, no user action required. v0.10.11

07 **Stable local API port and token** IMPROVED 65

The local API port is now stable and shared across all local-serving modes; the key is stored at `~/.darkbloom/local_token`.

— Exact token file path given, port number unspecified. v0.10.11

THINNER COVERAGE BELOW

08 **Early 503 for oversized requests** IMPROVED 55

Requests that exceed the Mac's available memory now receive a 503 before any SSE header, with a plain error message, instead of HTTP 200 followed by an in-stream error.

— Clear before/after but no field or endpoint named. v0.10.11

09 **darkbloom doctor diagnostic command** NEW 40

A new `darkbloom doctor` diagnostic command reports per-Mac serving status.

— Only named in the release summary, no further detail given. v0.10.11

└──▷ BREAKING ON UPGRADE

! Macs serving with `--local-endpoint` (including the desktop app's Alongside the network) now pause Darkbloom while you use their local API and resume 2 minutes after your last request; to restore the previous always-sharing behaviour, add `me_first = false` under `[local]` in `~/.config/darkbloom/provider.toml` and run `darkbloom restart`.

! An unknown `reasoning_effort` value is now refused with a typed 400 (`invalid_reasoning_effort`) on every reasoning model, including Qwen 3.5/3.6, Bonsai 2, Gemma 4, GPT-OSS and Nemotron Lightning, which previously accepted it silently.

WAS THIS USEFUL?  

AWS Amazon Bedrock

Feedback ⓘ

Added Claude Fable 5.2 and Claude Mythos 5.2 (2026-11-09) BLOG ↗ RANK

Amazon Bedrock provides managed foundation-model inference, customization, knowledge bases, and APIs for generative AI applications.

> Amazon Bedrock added two new Claude models with expanded cross-Region access, raised the output token burndown rate for the Mythos 5 line, and extended compaction, computer use, and tool-use flexibility across several Claude models.

└──▷ WHAT SHIPPED · 5 FEATURES most completely described first ? what's the number?

01 **Claude Fable 5.2 and Mythos 5.2 added** NEW 65

Amazon Bedrock adds Claude Fable 5.2 and Claude Mythos 5.2, accessible on the `bedrock-runtime` endpoint across US, AWS GovCloud (US), and Global cross-Region inference.

— Names endpoint and regions but no usage example

Added Claude Fable 5.2 and Claude Mytho…

02 **Claude Mythos 5.1 gains computer use and toolsets** NEW 65

Claude Mythos 5.1 now supports computer use, client toolsets, compaction, and the `aws_review` data retention mode.

— Names four distinct capabilities including a data-retention mode

Added Claude Fable 5.2 and Claude Mytho…

03 **Compaction extended to Claude Fable 5 and 5.1** NEW 60

Compaction is now supported on Claude Fable 5 and Claude Fable 5.1, with both compaction beta values listed on supported model cards and the `compact-2026-01-12` supported-model list expanded.

— Names the supported-model list version but no usage steps

Added Claude Fable 5.2 and Claude Mytho…

THINNER COVERAGE BELOW

04 **2.5x output token burndown rate for Mythos line** IMPROVED

45

Claude Mythos 5, Claude Mythos 5.1, and Claude Mythos 5.2 now carry a 2.5x output token burndown rate.

— States the change but no mechanism or config detail

Added Claude Fable 5.2 and Claude Mytho…

05 **Mid-conversation tool changes for Sonnet 5.5** NEW

40

Claude Sonnet 5.5 now supports mid-conversation tool changes.

— Bare statement with no mechanism or example Added Claude Fable 5.2 and Claude Mytho…

WAS THIS USEFUL?  

Nebius AI Cloud

Feedback ⓘ November 2–8, 2026 (2026-11-09)

BLOG ↗ RANK

Nebius AI Cloud provides GPU compute, managed Kubernetes, distributed training infrastructure, and serverless AI workloads.

> Nebius AI Cloud focused this window on AI-agent tooling, adding an MCP server for documentation and CLI access alongside new agent skills for planning and troubleshooting resources.

└──▷ WHAT SHIPPED · 2 FEATURES most completely described first ? what's the number?

01 **MCP server and AI documentation access** NEW 50

Nebius AI Cloud adds an MCP server that lets coding agents look up services and run Nebius AI Cloud CLI commands directly. It also adds AI documentation access via Markdown pages, `llms.txt`, and a docs MCP server for use with AI tools.

— Names MCP server and llms.txt but no exact commands given product docs

02 **Agent skills for resource planning and troubleshooting** NEW 30

Nebius AI Cloud adds agent skills — guidance for AI agents to plan, provision, and troubleshoot Nebius AI Cloud resources.

— Thin description with no concrete mechanism or surface named product docs

NEW INTEGRATION Dropbox as Object Storage transfer source and destination

NEW SUPPORTED PLATFORM Managed Service for MLflow in eu-west1

November 2–8, 2026 (2026-11-09) BLOG ↗ LATEST STABLE

Nebius AI Cloud adds an MCP server for coding agents, resource cloning in the web console, compute quota tiers, and uncapped spot pricing with per-platform minimums.

Connect a coding agent (Claude Code, Codex, or Cursor) to Nebius AI Cloud via the Nebius MCP server (preview) to run CLI commands, provision GPUs, run Serverless AI workloads, and control costs.

Clone Serverless AI jobs, endpoints, and Devlabs in the web console — the creation page opens with the original resource's settings prefilled.

Compute now has default and full quota tiers: new accounts start with default quotas and may be raised to full quotas automatically based on usage history.

Spot prices for preemptible GPU platforms are no longer capped and can rise above on-demand rates; each platform now has a minimum price, and you can set a max price in a pricing policy to limit costs and influence allocation priority.

└──▷ **BREAKING ON UPGRADE**

! Spot prices for preemptible GPU platforms are no longer capped — prices can now exceed regular on-demand rates, which may break cost assumptions in existing workloads that relied on the previous price ceiling.

WAS THIS USEFUL? 👍 👎

◆ AI / LLM Gateways

SMG Project Shepherd Model Gateway

Feedback ⓘ

v1.15.0 NOTES ↗

Shepherd Model Gateway is an engine-agnostic LLM gateway in Rust with full OpenAI & Anthropic API compatibility, gRPC pipeline, and KV cache routing.

SMG v1.15.0 adds load streaming from every gRPC worker to the router, hybrid KV cache-group routing, `smg.tokenizer` in the Python package, and major tokenizer and media-preprocessing speed gains.

└──▷ GET THIS VERSION ▾

pipx

```
$ pipx install smg==1.15.0
# already installed? upgrade to this version:
$ pipx install --force smg==1.15.0
```

Cargo

```
$ cargo install smg --version 1.15.0
# already installed? upgrade to this version:
$ cargo install smg --version 1.15.0 --force
```

git

```
$ git clone --branch v1.15.0 https://github.com/smg-project/smg.git
# already have the repo? check out this version:
$ git checkout v1.15.0
```

named in the project's README · checked 10 Nov Wrong command?

UPDATED INTEGRATION Keep the Python servicer for vLLM, SGLang or TokenSpeed by passing `--servicer-impl python` or setting `SMG_<ENGINE>_SERVICER_IMPL=python`; the Rust servicer is now the default for those engines.

Score a worker by its resident context tokens with `--power-of-two-load-metric resident_tokens:<request_tokens>` for engines that report resident token counts.

NEW SUPPORTED PLATFORM UPDATED SUPPORTED PLATFORM Control per-request media counts with `--mm-per-request-image-limit` or the environment variables `SMG_IMAGE_MAX_COUNT`, `SMG_VIDEO_MAX_COUNT`, and `SMG_AUDIO_MAX_COUNT` now that the gateway's own model-spec bounds are removed for ten families.

Replace `--least-load-max-waiting-requests` with `--worker-overload-waiting-requests` (applied per worker under every policy) and `--worker-overload-shed` to control refusal.

`SubscribeKvEvents` gains a `loads_only` request field so a servicer streams per-rank load records to the router without KV events being enabled; older servicers ignore the field and continue being polled.

+6 more

NEW SUPPORTED PLATFORM tekken.json vocabulary loading

UPDATED INTEGRATION TensorRT-LLM servicer now declares GetLoads and reports token-work, throughput and hit rate

└─▷ **BREAKING ON UPGRADE**

- ! The Rust gRPC servicer is now the default for vLLM, SGLang, and TokenSpeed; a launch the Rust servicer does not serve is refused by name with no fallback. Use `--servicer-impl python` or `SMG_<ENGINE>_SERVICER_IMPL=python` to keep the Python servicer.
- ! `--least-load-max-waiting-requests` and the `least_load` policy's `max_waiting_requests` config key are retired; a gateway or policy configuration that still sets either does not start. Replace with `--worker-overload-waiting-requests` and `--worker-overload-shed`.
- ! Video loader rules (`fps`, `max_frames`, `min_frames`, `temporal_patch_size`, `video_backend`) passed as launch flags on vLLM workers of the Qwen2-VL, Qwen3-VL, or GLM-5 line now cause the worker to refuse to start.
- ! Per-request media counts for Qwen2-VL, Qwen3-VL, Qwen3-Omni, Qwen3-ASR, Inkling, Muse Spark, Kimi K3, Gemma 4, Nemotron Omni, and DeepSeek V4.1 are no longer bounded by the gateway's own model specs.
- ! Load records now stream to the router from every gRPC worker under every policy; `least_load` and `power_of_two` placements can differ from v1.14.0 under the same traffic.
- ! Prompt token counts and prefix-cache keys change for DeepSeek V3.2 and V4 developer messages, system-final conversations, the tools block, the reasoning-effort prompt, Kimi K2 tool declarations, gpt-oss on the gRPC Harmony path, and GLM-5.3 reasoning efforts.
- ! The out-of-tree routing policy trait's completion hook no longer takes a `success` flag; the parameter must be dropped.
- ! The Rust library `llm-tokenizer` steps to 2.0.0: the three DeepSeek renderers now share one `DsEncodingError` type.
- ! The Helm chart's gRPC worker startup probe now asks the servicer's gRPC health service and the chart now requires the native JPEG decoder (`router.mmTurbojpeg:`

`require`).

- ! The launcher's `--eviction-interval-secs` now defaults to 120 s; smg_rs.Router(...) now starts with cache-aware routing, port 30000, and a 1800 s wait for the first worker.
- ! A ZMQ worker is no longer pinned to the SGLang wire without `--backend`.
- ! An encode policy outside `random`, `round_robin`, and `consistent_hashing`, and a discovery selector entry without `key=value`, are now refused by name instead of silently dropped.
- ! A Symphony tool call the parser could not read whole now reads as `incomplete` instead of arriving as a well-formed call.

WAS THIS USEFUL? 👍 👎

Mozilla AI Otari

Feedback ⓘ v0.41.0 NOTES ↗ RANK

Mozilla AI Otari is an open-source AI gateway you run yourself. One endpoint for 40+ providers, with virtual keys, budgets, and usage tracking.

> Mozilla AI Otari's v0.41.0 release ships context-aware pricing by service tier, request options and region, validates price rules against models.dev, adds model-kind filtering to the catalog API and dashboard, and renames the Feedback feature to Contact as a breaking change.

└──▷ WHAT SHIPPED · 8 FEATURES most completely described first ? what's the number?

01 **Context-aware pricing by tier, options, and region** IMPROVED

70

Otari now prices completions at the service tier they are served at, images and videos at the options they request, and Anthropic requests at the region they ran in.

— Describes mechanism across three request types, no config key named v0.41.0

02 **End user identifier in budget refusal responses** IMPROVED

65

Otari's budget refusals now carry `end_user` in the response body so callers can identify which end user triggered a spend limit.

— Names exact response field, no endpoint path given v0.41.0

03 **Feedback feature renamed to Contact, delivered via port**

65

BREAKING

The Feedback feature is renamed to Contact, now delivered through a port, and is off by default — any configuration or integration referencing `Feedback` must be updated.

— Clear migration note but no exact config key named v0.41.0

04 **Price rule validation against models.dev** NEW

60

Otari now refuses a price rule whose condition is not listed on models.dev, preventing misconfigured pricing from silently passing.

— Clear safeguard behavior, no exact surface named v0.41.0

05 **Dashboard catalog pricing, rules, filters, and route previews**

IMPROVED

60

Otari's dashboard now shows each catalog price in the billing unit it uses, shows the pricing rules each offering carries, filters the catalog by model kind, and shows the

sample request for each route an offering is served on. The dashboard model catalog also now compares a model's cheapest price within one unit.

— Lists several UI additions but no exact navigation path v0.41.0

THINNER COVERAGE BELOW

06 **Model-kind filter in catalog API** NEW 50

Otari's catalog API now lists a provider's offered models filtered by kind.

— Names catalog API but no endpoint path or param v0.41.0

07 **Tool configuration controls provider cost tier** IMPROVED 45

Otari's tool configuration now exclusively controls which provider cost tier a tool uses.

— States exclusivity but no config key named v0.41.0

08 **Account avatars with picker** NEW 35

Otari now gives every account an avatar with a picker to change it.

— Thin description with no further mechanism v0.41.0

└──▷ BREAKING ON UPGRADE

! The Feedback feature is renamed to Contact, now delivered through a port, and is off by default — any configuration or integration referencing Feedback must be updated.

WAS THIS USEFUL? 👍 👎

◆ Local LLM Hardware

Beelink GTR9 Pro

Feedback ⓘ price change · Nov 10, 2026

STORE ↗

Beelink GTR9 Pro is a Mac Studio-styled mini PC on AMD's Ryzen AI Max+ 395, with dual 10GbE ports aimed at running it as a local AI compute hub.

> GTR9 Pro · 128 GB / 2 TB gtr9-pro-128
>
> **$4,199** ~~$4,349~~ ▼ $150 · 3%
>
> list price at www.bee-link.com · Nov 10, 2026

WAS THIS USEFUL? 👍 👎

◆ Local LLM Runtimes

Ollama

Feedback ⓘ v0.40.16 NOTES ↗ CODE ↗ RANK

Ollama downloads, serves, and chats with local language models through a CLI and REST API.

> Ollama's v0.40.16 release fixes truncation handling in the OpenAI-compatible responses endpoint, adds the ripple-preview 65B coding model, and extends CUDA support to Windows on ARM.

└──▷ WHAT SHIPPED · 5 FEATURES most completely described first ? what's the number?

01 Truncation field respected in /v1/responses IMPROVED

90

The `/v1/responses` endpoint now respects the `truncation` field, mapping `'disabled'` to no truncation and `'auto'` to truncation, and defaults to no truncation when omitted — inputs no longer silently shorten when `truncation` is set to `'disabled'`.

Call the OpenAI-compatible responses endpoint with truncation explicitly disabled so long prompts are never silently shortened.

```
$ curl http://localhost:11434/v1/responses -d '{"model":
"ripple-preview", "input": "Refactor this codebase...",
"truncation": "disabled"}'
```

— Exact endpoint, field values, and a runnable curl example v0.40.16

02 Claude Code routing via ollama launch claude IMPROVED

90

`ollama launch claude` now always routes Claude Code to the selected Ollama model, overriding any inherited cloud provider env vars such as `CLAUDE_CODE_USE_BEDROCK` or `CLAUDE_CODE_USE_VERTEX`.

Route Claude Code through a locally selected Ollama model, ensuring cloud provider settings in the environment do not redirect requests elsewhere.

```
$ ollama launch claude
```

— Exact command, override env vars, and example provided v0.40.16

03 ripple-preview agentic coding model NEW

75

Run the new ripple-preview 65B dense agentic coding model locally with `ollama run ripple-preview`.

Run the new ripple-preview agentic coding model for local code generation tasks.

```
$ ollama run ripple-preview
```

— Model size and runnable command both given v0.40.16

04 **Empty array response from /v1/models** IMPROVED 65

`/v1/models` now returns an empty array `[]` instead of `null` when no models are installed, keeping it consistent with the OpenAI API spec.

— Named endpoint and exact behavior change, no example given v0.40.16

THINNER COVERAGE BELOW

05 **CUDA 13.5 support on Windows ARM** NEW 40

Adds CUDA 13.5 support for Windows on ARM (ARM64), including SM 103.

— Names platform and version but no usage steps v0.40.16

WAS THIS USEFUL? 👍 👎

◆ Inference Serving & Optimization

XGrammar

Feedback ⓘ v0.2.9 NOTES ↗ RANK

XGrammar constrains language-model output to grammars and schemas with optimized token-mask generation.

> XGrammar's v0.2.9 release adds structural tag grammar enhancements including Lark syntax and token boundary markers, opt-in JIT grammar compilation, Gemma 4 tool calling support, and jump-forward token generation for faster constrained decoding.

└──▷ WHAT SHIPPED · 4 FEATURES most completely described first ? what's the number?

01 **Structural tag grammar enhancements** NEW 70

`get_model_structural_tag` in Python now exposes opt-in `token_markers` to control token boundary marking in structural tag grammars, and structural tag grammar formats now support Lark syntax.

— Names the API and syntax addition, no usage example shown v0.2.9

THINNER COVERAGE BELOW

02 **Opt-in JIT grammar compilation** NEW 45

Grammar compilation can now be deferred until first use via opt-in JIT compilation, reducing upfront latency.

— Mechanism described but no flag or config key given v0.2.9

03 **Gemma 4 tool calling support** NEW 45

XGrammar now supports Gemma 4 tool calling using the `<|"|>` string delimiter.

— Names the model and delimiter but no usage walkthrough v0.2.9

04 **Jump-forward token generation** IMPROVED 35

XGrammar now adds jump-forward tokens while preserving accepted boundaries, speeding up constrained decoding.

— Mechanism described, no interface or benchmark given v0.2.9

WAS THIS USEFUL? 👍 👎

Niko1221 Strata

Feedback ⓘ v0.2.13 NOTES ↗ RANK

Strata is a local inference engine with one-click install for Windows and Linux, providing an OpenAI/Anthropic-compatible API on localhost for models like Qwen and DeepSeek.

> Strata's v0.2.13 release adds Intel Arc Linux support with manual GGUF loading, new engine-management CLI commands, per-card DFlash2 tuning defaults, and groundwork for Qwen4 models, alongside a faster startup and a full settings reference.

└──▷ **WHAT SHIPPED · 10 FEATURES** most completely described first ? what's the number?

01 **Engine shape query and rollback commands** NEW 85

Strata adds `strata --geometry` to query the engine's shape and confirm which engine variant is running (for example `flash_next` on CUDA 12, CUDA 13, HIP rdna3, and Spark engines), and `strata components rollback` to roll back the engine to the previous version while Control stays on the new version.

Check which engine shape is active after updating, to verify the correct engine variant was installed before loading a model.

```
$ strata --geometry
```

Roll back to the previous engine after a problematic update while keeping Control at the new version.

```
$ strata components rollback
```

— Names runnable commands and exact engine variants v0.2.13

02 **hipBLASLt table default change for RDNA cards** BREAKING 80

The hipBLASLt table is now off by default for Windows RX 9070, R9700 (gfx1201), and RX 7900 (gfx1100) cards; set the environment variable `STRATA_HIPBLASLT_TABLE=on` to restore the previous behavior.

Re-enable the hipBLASLt table on an RX 9070 or RX 7900 card if you notice a performance regression after upgrading, since it is now off by default.

```
$ STRATA_HIPBLASLT_TABLE=on strata
```

— Names affected cards, env var, and before/after v0.2.13

03 **Intel Arc support on Linux** NEW 70

Update Strata on Intel Arc (Linux) with `strata update` in the Terminal, then pick `Use your own GGUF…` to load a model from a folder of .gguf files, since the app cannot yet download models for Arc cards.

— Exact command and UI path, notes current limitation v0.2.13

04 **Removal of legacy tuning environment variables** BREAKING 60

The environment variables `STRATA_METAL_ROUTE`, `STRATA_METAL_PARTIAL`, `STRATA_METAL_WPF_MIB`, `STRATA_MMVQ_PF`, `STRATA_MMVQ_SG`, `STRATA_GR_Q32_PL`, `STRATA_DF2_GATE`, and `STRATA_DF2_ROW_MS` have been removed; any scripts or configs that set them will silently have no effect.

— Lists all removed vars but no migration guidance v0.2.13

05 **Per-card DFlash2 window and confidence defaults** IMPROVED 60

Strata now sets per-card DFlash2 window and confidence defaults: Strix Halo moves to window 5–8; Blackwell cc 12.0 with 48 GB or more moves to window 6–8 and confidence 0.35–0.55 for `IQ3_S`.

— Precise per-card numbers but nothing for reader to do v0.2.13

THINNER COVERAGE BELOW

06 **Qwen4 model groundwork** NEW 50

Strata adds Qwen4 groundwork: a model of a new shape gets its own engine download (`strata-q4-<shape>`) alongside the existing Strata engine, and the signed update feed can now carry a model list.

— Describes mechanism but is preparatory, not usable yet v0.2.13

07 **Faster app startup** IMPROVED 50

App start screen time drops from 248.5 ms to 143.4 ms and the main file shrinks from 3.12 MB to 0.76 MB, making startup noticeably faster.

— Concrete before/after metrics but nothing actionable v0.2.13

08 **External drive data folder support on Linux** NEW 45

Strata now accepts a data folder on an external drive (`/run/media/…`, `/media/…`, `/mnt`) on Linux.

— Names paths but no setup steps given v0.2.13

09 **Settings reference documentation** NEW 40

Strata adds `docs/SETTINGS.md`, a new reference page listing every available setting.

— Names the file but no content summary v0.2.13

10 **Engine defaults aligned with launcher settings** IMPROVED 40

Settings the launcher previously wrote are now the engine's own defaults, so a plain engine start matches the app and Terminal behavior without manual configuration.

— Explains behavior change but no specific keys named v0.2.13

└──▷ **BREAKING ON UPGRADE**

! The environment variables `STRATA_METAL_ROUTE`, `STRATA_METAL_PARTIAL`, `STRATA_METAL_WPF_MIB`, `STRATA_MMVQ_PF`, `STRATA_MMVQ_SG`, `STRATA_GR_Q32_PL`, `STRATA_DF2_GATE`, and `STRATA_DF2_ROW_MS` have been removed; any scripts or configs that set them will silently have no effect.

! The hipBLASLt table is now off by default for Windows RX 9070 / R9700 (gfx1201) and RX 7900 (gfx1100); set `STRATA_HIPBLASLT_TABLE=on` to restore the previous behavior.

WAS THIS USEFUL? 👍 👎

InternLM LMDeploy Feedback ⓘ CVE-2026-123818 CVE ↗

LMDeploy compresses, deploys, and serves language and vision-language models with its TurboMind engine, offering quantization and an OpenAI-compatible API.

.
.
.
.

`CVE-2026-123818`: SSRF bypass in lmdeploy v0.12.3 (CVSS 7.5, affects `>= 0.12.3, < 0.15.0`)

WAS THIS USEFUL? 👍 👎

DATA

◆ Document Parsing

Docling

Feedback ⓘ v2.150.0 NOTES ↗ RANK

Docling parses PDFs, Office files, and HTML into structured documents with layout, tables, and reading order preserved for downstream RAG pipelines.

> Docling extends its document parsing coverage to structured financial reports with new XBRL and iXBRL support.

└──▷ WHAT SHIPPED · 1 FEATURE most completely described first ? what's the number?

01 **XBRL and iXBRL financial document parsing** NEW SUPPORTED PLATFORM

NEW 40

Docling now parses XBRL (eXtensible Business Reporting Language) and iXBRL (Inline eXtensible Business Reporting Language) documents, enabling extraction from structured financial reports.

— Names formats but no mechanism, API, or usage detail given product docs

v2.150.0 NOTES ↗ LATEST STABLE

Docling adds iXBRL parsing with remote taxonomy resolution via `source_uri`, extending existing XBRL financial report support.

└──▷ GET THIS VERSION ▾

pip

```
$ pip install docling==2.150.0
# already installed? upgrade to this version:
$ pip install --upgrade docling==2.150.0
```

git

```
$ git clone --branch v2.150.0 https://github.com/docling-
project/docling.git
# already have the repo? check out this version:
$ git checkout v2.150.0
```

PyPI attestation · checked 10 Nov Wrong command?

Parse iXBRL (Inline eXtensible Business Reporting Language) documents with remote taxonomy resolution using the `source_uri` parameter.

WAS THIS USEFUL? 👍 👎

◆ Vector DBs & RAG

askimo

Feedback ⓘ v1.5.5 NOTES ↗ RANK

Askimo is an AI desktop app for chat, RAG, Skills, MCP tools, and agents, supporting multiple LLMs (Anthropic, OpenAI, VertexAI, vLLM, Nvidia NIM, Gemini, OpenRouter).

> Askimo v1.5.5 adds local LLM execution and Gemini support, granular tool-approval policies, and persistent MCP server tracking per chat session, alongside chat organization, enterprise resource segmentation, and several configuration refinements.

└──▷ WHAT SHIPPED · 9 FEATURES most completely described first ? what's the number?

01 **Tool execution approval policies** NEW 65

Tool execution granularity can now be controlled with `requiresApproval` helpers and the `ALWAYS_AUTO` policy in tool configuration.

– Names config helpers and policy but no usage flow v1.5.5

02 **Chat session organization and history controls** NEW 60

Chat sessions can now be labeled and organized, with sidebar filtering to locate and manage chat transcripts, and soft-delete for chat turns enabling recoverable removal within the chat history module.

– Describes three concrete UI behaviors in chat module v1.5.5

THINNER COVERAGE BELOW

03 **Local LLM execution and Gemini support** NEW 50

Askimo now supports running large language models locally within the desktop environment, and adds Gemini support via the `google-genai` SDK.

– Names SDK but no setup detail v1.5.5

04 **MCP server management and session persistence** IMPROVED 45

The MCP registry is extended with new templates and utilities for managing and integrating external processing servers, and Askimo now tracks and persists which MCP servers are enabled for specific chat sessions to maintain context across sessions.

– Two MCP additions grouped, no concrete command shown v1.5.5

05 **TEAM_MANAGED resource collections** NEW 45

Adds `TEAM_MANAGED` resource collections across the stack for enterprise-level resource segmentation.

— Names concrete resource type, lacks usage example v1.5.5

06 **Git-ignore based project filtering** NEW 35

Adds git-ignore specification support for project structure filtering utilities.

— Names gitignore mechanism, no further detail v1.5.5

07 **Advanced and voice settings sections** NEW 30

Enables advanced and voice settings sections in desktop and shared configuration modules.

— Names config modules but not setting contents v1.5.5

08 **RAG citation source configuration** IMPROVED 25

RAG citation sources are now configurable via advanced settings.

— Thin description, no setting name given v1.5.5

09 **Attachment support for inputs** NEW 20

Adds attachment support for various inputs across the platform.

— Generic description without specifics v1.5.5

WAS THIS USEFUL? 👍 👎

liliang-cn CortexDB

Feedback ⓘ v2.149.0 NOTES ↗ RANK

CortexDB is a local-first AI agent memory with a knowledge graph, ontology constraints, and traceable reasoning—all in one SQLite file.

> CortexDB's v2.149.0 release adds real-time recall streaming to the live view and propagates ontology-declared types into extraction and capture workflows.

└──▷ **WHAT SHIPPED · 2 FEATURES** most completely described first ? what's the number?

01 **Ontology-type awareness for extractors and capture mod** IMPROVED

55

Extractors now receive the active ontology's declared types so extraction output aligns with the governed type vocabulary, and the `cortexdb-live` Claude Code mod passes those same ontology types to its capture step so captured memories use governed type names.

– Names the mod and mechanism but no exact types or config shown v2.149.0

02 **Real-time recall streaming in live view** NEW

45

The live view now streams every agent recall to the open page as it happens, letting observers watch hits appear in real time instead of polling or refreshing.

– Explains behaviour but no config, flag, or limit named v2.149.0

WAS THIS USEFUL? 👍 👎

knowhere

Feedback ⓘ v1.2.50 NOTES ↗ RANK

Knowhere extracts, parses, and outputs structured chunks ready for AI Agents and RAG.

> Knowhere shipped two capabilities in this window: smarter Excel table parsing with automatic header detection and queryable flattened columns, and automatic OCR recovery for corrupt-text PDF pages during progressive ingestion.

└──▷ **WHAT SHIPPED · 2 FEATURES** most completely described first ? what's the number?

01 **OCR recovery for corrupt-text PDF pages** NEW 65

During progressive PDF ingestion, knowhere now runs OCR on pages whose native text is detected as corrupt, making the recovered text available to search and later enrichment. Recovery operates within a bounded budget; pages that exceed the budget retain their page image but may lack searchable text. Existing parse snapshots do not pick up this OCR-based recovery automatically and require reprocessing.

– Explains mechanism, budget behavior, and migration note; no exact commands given

v1.2.49

02 **Excel table parsing: headers, regions, flattened columns**

IMPROVED 62

Excel tables now automatically detect headers and captions, and split and merge sheet regions during ingestion. Parsed tables also expose flattened column names accessible through `corpus.query_table`, enabling structured queries against spreadsheet data.

– Names a concrete API (`corpus.query_table`) but lacks usage example v1.2.50

└──▷ **BREAKING ON UPGRADE**

! Existing parse snapshots do not pick up OCR-based text recovery automatically and require reprocessing.

WAS THIS USEFUL? 👍 👎

fastembed-rs

Feedback ⓘ v7.7.0 NOTES ↗ RANK

fastembed-rs is a Rust library for generating vector embeddings and reranking locally.

> fastembed-rs added Metal GPU acceleration for BERT-family text embedding models, speeding up local inference on Apple Silicon.

└──▷ **WHAT SHIPPED · 1 FEATURE** most completely described first ? what's the number?

01 **Metal acceleration for BERT embeddings** IMPROVED 45

BERT-family text embedding models now run on Apple Metal with fused epilogues, accelerating local embedding inference on Apple Silicon hardware.

— Names mechanism (fused epilogues, Metal) but no API or flag v7.7.0

WAS THIS USEFUL? 👍 👎

◆ Data Labeling

HumanSignal Label Studio Feedback ⓘ CVE-2026-22033

CVE ↗ STABLE

Label Studio is an open source data labeling platform for text, images, audio, video and time series, used to build training and evaluation datasets for AI models.

`CVE-2026-22033` : [DCODX-AI] - Full account takeover chaining Stored XSS + IDOR in User Profile via custom_hotkeys field (CVSS 5.4, affects `1.21.0`)

▾ 9 more releases in this issue · 2023-11-08 → 2026-01-12

> `CVE-2025-47783` CVE ↗ STABLE
>
> `CVE-2025-47783` : Cross-Site Scripting (Reflected) via the label_config parameter. (CVSS 6.1, affects `< 1.18.0`)

> `CVE-2025-25297` CVE ↗ STABLE
>
> `CVE-2025-25297` : [XBOW-024-151] Server-Side Request Forgery in Label Studio S3 Storage Endpoint (CVSS 8.6, affects `<1.16.0`)

> `CVE-2025-25296` CVE ↗ STABLE
>
> `CVE-2025-25296` : [XBOW-024-119] Cross-Site Scripting (XSS) via GET request to `/projects/upload-example` endpoint (CVSS 5.4, affects `< 1.16.0`)

> `CVE-2025-25295` CVE ↗ STABLE
>
> `CVE-2025-25295` : [XBOW-025-020] Path Traversal Vulnerability in Label Studio via image Field (affects `<1.0.10`)

> `CVE-2024-26152` CVE ↗ STABLE
> .
> .
> .
> .
>
> `CVE-2024-26152` : XSS Vulnerability if `<Choices>` or `<Labels>` are used in labeling config (CVSS 4.7, affects `< 1.11.0`)

> `CVE-2023-47116` CVE ↗ STABLE
>
> `CVE-2023-47116` : Label Studio SSRF on Import Bypassing `SSRF_PROTECTION_ENABLED` Protections (CVSS 5.3, affects `< 1.11.0`)

CVE-2024-23633 CVE ↗ STABLE

`CVE-2024-23633` : XSS Vulnerability on Data Import (CVSS 4.7, affects `< 1.10.1`)

CVE-2023-47117 CVE ↗ STABLE

`CVE-2023-47117` : Object Relational Mapper Leak Vulnerability in Filtering Task (CVSS 7.5, affects `<= 1.9.2`)

CVE-2023-43791 CVE ↗ STABLE

`CVE-2023-43791` : Hardcoded Django `SECRET_KEY` that can be Abused to Forge Session Tokens (CVSS 9.8, affects `<= 1.8.1`)

WAS THIS USEFUL? 👍 👎

EVALUATE

◆ AI Observability & Evals

LangChain LangSmith

Feedback ⓘ updated · Nov 10, 2026

NOTES ↗ RANK

LangSmith provides tracing, evaluation, and deployment tools for LLM applications.

> LangSmith's latest release deepens LLM Gateway observability and Bedrock integration, reworks thread-based automations around SmithDB queries, and adds dual-stack networking, new chat models, and sandbox diagnostics.

└──▷ WHAT SHIPPED · 8 FEATURES most completely described first ? what's the number?

01 **Gateway metadata header and trace visibility** NEW 80

The LLM Gateway now returns an `X-LangSmith-Gateway-Metadata` response header describing routing, policy limits, guard activity, and upstream errors; clients can attach it to traces as `ls_gateway_info` metadata. Trace details now surface LLM Gateway outcomes, policy limits, selected model, and fallback attempts in a dedicated Gateway section, and errored runs blocked by the gateway show a Blocked badge in the trace tree.

Capture LLM Gateway routing metadata and attach it to your own trace for end-to-end audit of model selection, policy limits, and fallback attempts.

python

```
response =
requests.post('https://api.smith.langchain.com/v1/chat/completio
ns', headers={'Authorization': 'Bearer <api_key>'}, json={...})
gateway_meta = response.headers.get('X-LangSmith-Gateway-
Metadata')
client.create_run(..., extra={'ls_gateway_info': gateway_meta})
```

— Names exact header and metadata key, with runnable code example snapshot-20261110

02 **Turns view deprecated in thread detail** DEPRECATED 70

The Turns view in thread detail is deprecated and will be removed from LangSmith Cloud on October 31, 2026, and from self-hosted in v17; use the Messages or Details view instead.

— Gives exact removal dates and migration alternative snapshot-20261110

03 **Thread-scoped automation queues and filters** BREAKING 65

Thread-grouped automations can now add one annotation queue item per matched conversation so reviewers can inspect the full thread instead of a single run. Thread webhook and annotation-queue automations now require SmithDB queries and show thread filters instead of run filters, and no longer default to `Is Trace = true`; existing automations using run filters will need to be migrated.

— Names changed default and migration need, no UI path given snapshot-20261110

04 **Improved Bedrock gateway routing and authentication** IMPROVED 60

The LLM Gateway now authenticates to Amazon Bedrock with an assumed IAM role or AWS access keys instead of requiring a static Bedrock API key (existing Bedrock API keys still take precedence when multiple credential types are configured), and routing to Bedrock now selects the API format per model family — Anthropic models use Messages, OpenAI-compatible models use Responses, and other Bedrock models use Chat Completions — preventing unsupported API errors for models such as Claude Opus.

— Explains precedence rules but no setup steps given Names model families and formats but no commands shown

snapshot-20261110

05 **CSV export for colon-containing dataset keys** IMPROVED 60

Datasets whose input, output, or metadata keys contain a colon — including `source_run_feedback:<key>` metadata keys — can now be exported to CSV.

— Names exact key format fixed, no export command shown snapshot-20261110

THINNER COVERAGE BELOW

06 **New GPT-5.6 models in LangSmith Chat** NEW 45

LangSmith Chat now offers GPT-5.6 Sol, GPT-5.6 Terra, and GPT-5.6 Luna alongside existing OpenAI models.

— Names specific models but no usage instructions snapshot-20261110

07 **Sandbox resource usage history** NEW 40

Sandboxes now retain recent CPU, memory, and network usage history to help diagnose resource consumption and performance bottlenecks.

— Describes metrics tracked but no access method given snapshot-20261110

08 **IPv4 and IPv6 dual-stack support** NEW 20

LangSmith services now accept both IPv4 and IPv6 connections.

— Bare statement with no configuration or scope detail snapshot-20261110

└──▷ BREAKING ON UPGRADE

! Thread webhook and annotation-queue automations now require SmithDB queries; existing automations using run filters will need to be migrated.

! The Turns view in thread detail is deprecated and will be removed from LangSmith Cloud on October 31, 2026, and from self-hosted in v17; use the Messages or Details view instead.

WAS THIS USEFUL? 👍 👎

Arize Phoenix

Feedback ⓘ `@arizeai/phoenix-cli@1.21.0` NOTES ↗

RANK

Arize Phoenix is an open-source platform that monitors LLM application traces and evaluates their outputs.

> Arize Phoenix added a CLI command for listing dataset versions and expanded its Python evaluators with new decision-model backends for OpenAI and Anthropic.

└──▷ **WHAT SHIPPED · 2 FEATURES** most completely described first ? what's the number?

01 **Dataset versions list command in CLI** NEW 83

The CLI gains `px dataset-versions list <dataset-identifier>` to list all versions of a dataset, supporting cursor pagination, optional limits, and pretty, JSON, and raw output formats.

— Names exact command, flags and output formats `@arizeai/phoenix-cli@1.21.0`

THINNER COVERAGE BELOW

02 **New decision model backends for classification evaluators**

IMPROVED 56

Python classification evaluators now support the OpenAI Decisions API as a decision model backend, and the `AnthropicAdapter` now uses Anthropic native structured output instead of tool-call forcing for classification evaluations.

— Names backends and adapter but no usage steps `arize-phoenix-evals-v3.10.0`

> `CVE-2026-53998` CVE ↗ STABLE
>
> `CVE-2026-53998`: Format String Injection in applyChatTemplate (CVSS 10.0, affects `>= 13.0.0`)

WAS THIS USEFUL? 👍 👎

Langfuse

Feedback ⓘ v4.77.0 NOTES ↗ RANK

Langfuse provides tracing, evaluation, and monitoring for LLM applications.

> Langfuse removed automatic token-usage inference—callers must now submit token counts explicitly—while also adding Amazon Bedrock as an AI gateway upstream provider, IAM role authentication for S3, and audit logging for public API list routes.

└──▷ WHAT SHIPPED · 9 FEATURES most completely described first ? what's the number?

01 Automatic tokenization removed, manual submission required

60

BREAKING

Automatic tokenization is removed — the worker no longer infers observation token usage from content, so setups that relied on Langfuse computing token counts automatically will stop receiving those counts; token usage must now be submitted explicitly by the caller. A UI banner guides users to submit missing token usage after this change.

– Clear before/after and migration note, but no exact API field named v4.77.0

THINNER COVERAGE BELOW

02 Topic summaries queue isolation and cost tracking IMPROVED

45

Topics summaries now run on their own queue per trace, isolating their workload, and cost tracking is now computed for topic execution runs.

– Describes mechanism but no actionable config or flag v4.77.0

03 AI gateway connection and resource improvements IMPROVED

45

The AI gateway now supports custom base URLs and extra headers for `OpenAI` and Anthropic connections, and sizes its memory budget from the cgroup limit for container-aware resource management.

– Names the providers and capability but no exact field name Explains mechanism but is an automatic, non-actionable change

v4.77.0

04 Amazon Bedrock upstream provider in AI gateway NEW

43

Connect to Amazon Bedrock as an upstream provider in the AI gateway using API key auth.

– Names the integration and auth method but no config key shown v4.77.0

05 **IAM role auth for S3 storage and exports** IMPROVED 40

Authenticate Amazon S3 buckets and S3 exports using IAM role auth instead of static credentials.

– Names the mechanism change but no setup steps given v4.77.0

06 **Audit logging for public API list routes** NEW 25

Audit logs now record read-access events for public API list routes.

– States the change without naming specific routes or fields v4.77.0

07 **External media storage now enterprise-only** BREAKING 25

External media storage is now an enterprise feature.

– Restriction stated but no detail on affected setups v4.77.0

08 **Alert messages show metric value and unit** IMPROVED 15

Alert messages now show the metric value and unit alongside the alert notification.

– Minor UI change described in one line v4.77.0

09 **Improved Langchain input/output normalization** IMPROVED 15

Langchain input/output normalization is now improved in the normalized-io pipeline.

– No specifics on what changed in normalization v4.77.0

└──▷ **BREAKING ON UPGRADE**

! Automatic tokenization is removed — the worker no longer infers observation token usage from content. Setups that relied on Langfuse computing token counts automatically will stop receiving those counts; token usage must now be submitted explicitly by the caller.

WAS THIS USEFUL? 👍 👎

Braintrust

Feedback ⓘ updated · Nov 10, 2026 NOTES ↗

Braintrust provides evaluation, tracing, and improvement workflows for AI applications.

Braintrust adds pattern rules for scheduled Loop runs, in-browser thread translation, Terraform provider GA with new ephemeral resources, and `bt trace run --parent` for attaching coding-agent sessions to existing spans.

Pattern rules let you narrow scheduled Patterns runs to specific traces with a SQL filter and list problems for Loop to watch for.

Translate non-English thread messages and span names in-browser using the built-in translation API in desktop Chrome or Edge, without consuming model credits.

The Braintrust Terraform provider is generally available (v1.0.0), adding an ephemeral `braintrust_service_token` resource and a `braintrust_users` data source.

Manage saved views for Patterns as code by setting `patterns` as the `view_type` on the `braintrust_view` Terraform resource.

Continue a trace from an existing SDK span by passing `--parent` to `bt trace run` with the value returned by a span's export().

+4 more

NEW INTEGRATION Firebase Genkit tracing (Go SDK)

NEW INTEGRATION MCP server tracing for fastmcp and mcp-handler (TypeScript SDK)

UPDATED INTEGRATION Expanded tracing support for AI frameworks (TypeScript SDK v3.43.0)

UPDATED INTEGRATION Anthropic Java SDK 3.x support and async client wrapping (Java SDK)

NEW INTEGRATION Temporal activity and workflow tracing (Python SDK)

NEW INTEGRATION OpenAI Responses API span recording (Python SDK)

UPDATED INTEGRATION LiveKit Agents PII redaction in traces (TypeScript SDK)

UPDATED INTEGRATION Cursor SDK subagent tool call tracing (TypeScript SDK)

└──▷ BREAKING ON UPGRADE

! The Terraform provider now requires Terraform or OpenTofu 1.11.1 or later.

! `braintrust_ai_secret` now requires `secret_version` alongside `secret`, and `braintrust_env_var` now requires `value_version` alongside `value`; changing only the value no longer updates it, and references to `secret` or `value` return `null`.

! The `braintrust_org_automation` Terraform resource is removed; delete its blocks from your configuration and run `terraform state rm` on each address before

running `terraform plan`.

WAS THIS USEFUL? 👍 👎

GOVERN

◆ AI Runtime Defense & Guardrails

PipeLab Pipelock

Feedback ⓘ v3.8.0 NOTES ↗

Pipelock is an open-source AI agent firewall that scans MCP and HTTP egress for exfiltration, SSRF, and prompt injection and emits signed action receipts.

Pipelock adds receipt-load controls (`--restart-after` , `--destinations per-request` , `--ready-timeout` , `RECEIPT_LOAD_VERIFY_TIMEOUT`) and tightens evidence completeness, URL signing, and outbound decoding enforcement.

└──▷ GET THIS VERSION ▾

Docker

```
$ docker pull ghcr.io/pipelab-org/pipelock:3.8.0
```

git

```
$ git clone --branch v3.8.0 https://github.com/pipelab-org/pipelock.git
# already have the repo? check out this version:
$ git checkout v3.8.0
```

published under the project's GitHub account · checked 10 Nov Wrong command?

Control receipt-load restart and attribution with `--restart-after` to verify continuity across a proxy restart, `--destinations per-request` to attribute destination-only receipts to individual planned requests, and `--ready-timeout` or `RECEIPT_LOAD_VERIFY_TIMEOUT` to allow more startup and verification time.

Set the maximum URL length explicitly with `fetch_proxy.monitoring.max_url_length` now that the default ceiling has changed.

NEW DETECTION Nearby-keyword gates for custom DLP patterns

UPDATED DETECTION Expanded encoded-text coverage for DLP and injection scanning

NEW DETECTION DLP inspection of long separator-split encoded values

NEW DETECTION Unicode tag character decoding for DLP and injection scanning

└──▷ BREAKING ON UPGRADE

! `/metrics` and `/stats` endpoints are now exposed only when `metrics_listen` names a dedicated listener; omitting or blanking the setting disables both routes at startup.

- ! Container and Helm chart images now use the `ghcr.io/pipelab-org` registry; update image repositories, registry policies, and Helm OCI URLs before adopting this release.
- ! A valid open evidence prefix now reports `VALID PREFIX, END UNPROVEN` and exits 3; use `--allow-prefix` to deliberately accept a supported prefix without proving its end.
- ! Outbound decoding now refuses when decoding depth or work bounds leave a new view uninspected; requests on URL, body, header, WebSocket, MCP, and A2A paths are affected.
- ! Signed URL targets in receipts now retain only scheme, host, and port, dropping userinfo, path, query, and fragment; update consumers that relied on receipt paths or queries.
- ! The verifier install gate now exits `3` for a retryable registry access failure and `2` for a package contradiction or unclassified failure; `receipt-load` exits `4` for a setup failure before receipt evidence exists.

WAS THIS USEFUL? 👍 👎

AI MODELS

◆ Frontier Models

Anthropic Claude Fable 5.2 Feedback ⓘ

released · Nov 10, 2026 MODEL ↗ NOTES ↗ BLOG ↗

Anthropic: "Claude Fable 5.2 and Claude Mythos 5.2, our new models for the hardest technical challenges."

Anthropic says: "Fast mode is available for $16 per million input tokens and $80 per million output tokens."

Anthropic says: "At High effort, Claude Fable 5.2 scores about the same as Opus 5.5's best score (45.3% vs. 45.5%) for roughly half the cost per task, and ahead of GPT-6 Astra's best score (42.8%) at about the same cost."

Anthropic says: "At Xhigh effort, Claude Fable 5.2 reaches 49.1%, the highest score of any released model we've tested."

Anthropic says: "At Medium effort, Claude Fable 5.2 matches Opus 5.5's top score for about 40% of the cost per task."

Anthropic says: "At Medium effort, Claude Fable 5.2 scores 66.4%, ahead of Opus 5.5 at Max effort (64.4%) for less than one-tenth of the cost per task."

+12 more

└──▷ BREAKING ON UPGRADE

- ! Claude Fable 5.2 and Claude Mythos 5.2 enforce the prefix check for every account regardless of age; if your code edits the system prompt, tools, or an earlier message and replays a thinking block from a later turn, the request returns a 400 error.
- ! A request using the `computer_20251124` tool on Claude Fable 5.2 or Claude Mythos 5.2 returns a 400 error on all platforms except Amazon Bedrock, which still accepts the earlier tool — migrate to `computer_toolset_20260801`.
- ! When a conversation moves from Claude Fable 5.2 or Claude Mythos 5.2 to any other model (including Claude Opus 5.5 and Claude Fable 5.1), the API silently drops the thinking blocks those models produced.
- ! A Claude Fable 5.2 or Claude Mythos 5.2 executor using the `advisor` tool returns a 400 error if the advisor model is anything other than Claude Fable 5.2 or Claude Mythos 5.2.

WAS THIS USEFUL? 👍 👎

Meta Muse Ember

Feedback ⓘ released · Nov 10, 2026 MODEL ↗

Muse Ember from Meta was released on 9 Nov 2026. 1049K-token context; $0.10 in / $0.50 out per million tokens (OpenRouter).

WAS THIS USEFUL? 👍 👎

Meta Muse Ember (Meta)

Feedback ⓘ released · Nov 10, 2026 MODEL ↗

Muse Ember (Meta) from Meta was released on 9 Nov 2026. 1049K-token context; $0.10 in / $0.50 out per million tokens (models.dev).

WAS THIS USEFUL? 👍 👎

◆ Open-Weight Models

Nex AGI Nex-N2.6

Feedback ⓘ released · Nov 10, 2026 MODEL ↗

Nex-N2.6 from Nex AGI was released on 10 Nov 2026. 262K-token context; $0.12 in / $0.40 out per million tokens (OpenRouter).

WAS THIS USEFUL? 👍 👎

// NAME CHANGES 1

Tools and vendors that now go by another name. Search and old links still find them under the name they had.

Superset is now Paradev

Name change announced Nov 5, 2026

// END OF TRANSMISSION

The AI Toolchain · curated daily · November 10, 2026

Releases & features from the open-source and commercial AI tools that practitioners actually run.

▷ Subscribe Feedback on this issue