THE AI TOOLCHAIN

NO. 093

Tail

THE DAILY RELEASE FIREHOSE

PUBLISHED NOVEMBER 15, 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 219 changes

Do you prefer **scroll**? 👍 👎

▸ // CONTENTS SHOW 54 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.

Three unrelated things stand out. coreai-kit's checkpoint resume drops time-to-first-token from 837.8 ms to 38.6 ms on resumed chats. Kiro

vim ? help

adds Grok 4.7 at a 500K-token context plus an Approve for Me permission mode. And nono and openclaw.net both move agent access onto real identity — session recording, live audit logging, and on-behalf-of token exchange.

DEPLOY

Resume a long chat without paying the whole prompt again

Checkpoint-based resumption restores saved state instead of reprocessing the conversation, cutting time-to-first-token from 837.8 ms to 38.6 ms; the catalog also grew from 67 to 80 models, so model swaps stay in-SDK. The warm-up pause on every reopened session goes away.

coreai-kit

GOVERN

Let an agent act as a named user, and have the log to prove it

nono adds platform authentication, session recording and live audit logging with AWS operation-level endpoint policy matching; openclaw.net runs workflows under user-delegated auth via Entra on-behalf-of or RFC 8693 token exchange. Agent actions stop landing under one shared service credential, which is what makes after-the-fact attribution possible at all.

nono · openclaw.net

BUILD

Approve only the steps you care about, in a 500K-token window

Kiro's Approve for Me mode sits between full autonomy and prompting on every tool call, and Grok 4.7 at 500K tokens means a large repo context survives a session; cezar's Manager mode gives a global agent repo connections and merge approval. Both reduce the per-step babysitting that makes long agent runs expensive in attention.

Kiro · cezar

BUILD

Tune retrieval quality per collection instead of accepting one default

Configurable HNSW search depth trades recall against latency per new collection, MMR deduplication stops near-identical chunks from crowding the context, and BM25 rebuilds move to the background so large indexes stay queryable during reindex. Index-quality diagnostics replace guessing at why a retrieval went wrong.

knowledge-rag

Does **Opus 5** read better? 👍 👎

OFFENSIVE SECURITY

◆ Exploitation & C2

xalgorix

Feedback ⓘ v4.6.190 NOTES ↗ RANK

The xalgorix platform runs AI pentesting agents for reconnaissance, vulnerability detection, and exploitation workflows.

> xalgorix now auto-generates ready-to-run curl replay commands for every HTTP/HTTPS finding's proof of concept.

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

01 **Auto-generated curl replay commands for findings** NEW

73

Every HTTP/HTTPS finding now includes a copy-pasteable `curl` command built from the finding's reported method and endpoint. It preserves literal `../` path-traversal segments, URL encoding, and shell quoting so the request replays exactly.

— Explains mechanism and gives a directly runnable command. v4.6.190

WAS THIS USEFUL? 👍 👎

DETECTION & RESPONSE

◆ Federated Security Search

gfs

Feedback ⓘ v0.5.3 NOTES ↗ RANK

Gfs is a git For database Systems

> gfs's v0.5.0 release was the biggest of the window, adding a `gfs user` CLI and MCP tool, `gfs destroy`, a file-based SQLite provider, Kubernetes/k3s runtime support, a node-local `RepoCredentialVault`, a more thorough `gfs fsck`, recoverable `branch -d`, an interactive `gfs storage reclaim`, and richer Kubernetes storage reporting; v0.5.2 and v0.5.3 followed with Kubernetes-specific environment variables for TLS secrets and node scheduling.

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

01 **gfs fsck filesystem consistency check** NEW 85

Check filesystem consistency with `gfs fsck`, which verifies object hashes, reports unreachable and dangling objects without removing anything, proves a detached working copy is safe to remove, and reports what deleting a snapshot would actually free on Kubernetes.

— Command named with detailed mechanism and scope v0.5.0

02 **Kubernetes storage reporting and configuration** IMPROVED 85

Kubernetes storage reporting and configuration improve: `gfs status` now surfaces whether snapshots share extents or copy bytes and status() reports real ZFS volume usage for Kubernetes storage; pod readiness is detected via a Kubernetes watch instead of a 2-second poll; and storage-kubernetes automatically uses the cluster-default `VolumeSnapshotClass` when none is configured.

— Multiple named surfaces and before/after but no CLI example v0.5.0

03 **gfs user CLI and MCP tool for database users** NEW 80

Adds the new `gfs user` CLI supporting `create`, `list`, `drop`, and `set-password` subcommands, plus `gfs user grant`, `gfs user revoke`, and `gfs user list-privs` to grant, revoke, and list object-level privileges. `gfs user` is also exposed as an MCP tool so AI agents can create, list, drop, and set passwords for database users without shell wrappers.

— Commands fully named but no runnable example given v0.5.0

04 **Kubernetes deployment configuration via environment variables** NEW 80

gfs adds `GFS_K8S_NODE_SELECTOR` to schedule databases on labelled Kubernetes nodes when no node is explicitly pinned, `GFS_K8S_TLS_SECRET` to mount a TLS Secret into the Postgres pod on Kubernetes, and `GFS_CONTAINER_UID_GID` to override the container run-as user.

Run GFS containers as a specific UID/GID to match the host filesystem ownership in a rootless environment.

```
$ GFS_CONTAINER_UID_GID=1001:1001 gfs init --database-provider
postgres --database-version 17
```

— Named all three env vars; only one has a runnable example v0.5.3 v0.5.2 v0.5.0

05 **Interactive `gfs storage reclaim`** IMPROVED 75

Run `gfs storage reclaim` interactively — it now prompts for confirmation when run at a terminal, and deletes OpenEBS ZFS volumes the driver deferred.

— Command named with before/after behavior change v0.5.0

06 **gfs destroy command** NEW 70

Tear down an instance and remove the local `.gfs` directory with the new `gfs destroy` command.

Tear down a GFS-managed instance and clean up the local `.gfs` directory when you are done with a project.

```
$ gfs destroy
```

— Runnable command with example; mechanism is simple teardown v0.5.0

07 **RepoCredentialVault for repository credentials** NEW 60

Store repository credentials node-locally in the new `RepoCredentialVault` at `.gfs/secrets`.

— Names exact storage path but no usage command v0.5.0

THINNER COVERAGE BELOW

08 **File-based SQLite provider** NEW 45

Use the new file-based SQLite provider with an in-process engine — no container runtime required.

— Describes behavior but no flag or command shown v0.5.0

09 **Recoverable `branch -d`** IMPROVED 45

Make `branch -d` recoverable instead of silently destructive.

— Names command but no mechanism for recovery given v0.5.0

10 **Kubernetes/k3s runtime support** NEW 30

Run GFS on k3s/Kubernetes with the new k3s/Kubernetes runtime.

— Bare statement of new runtime, no setup detail v0.5.0

NEW SUPPORTED PLATFORM PostgreSQL 18 support

NEW SUPPORTED PLATFORM Linux aarch64 binary

> v0.5.1 NOTES ↗ STABLE
>
> gfs v0.5.1 adds PostgreSQL 18 extension support and a native Linux aarch64 binary.
>
> └──▷ GET THIS VERSION ▾
>
> ```
> $ git clone --branch v0.5.1 https://github.com/Guepard-Corp/gfs.git
> # already have the repo? check out this version:
> $ git checkout v0.5.1
> ```
>
> gfs now supports PostgreSQL 18, with the gfs extension built for pg18 and `gfs-postgres` images published to GHCR for PostgreSQL 16, 17, and 18.
>
> gfs now publishes a native Linux aarch64 binary alongside x86_64, with runners pinned to Ubuntu 24.04.

WAS THIS USEFUL? 👍 👎

BUILD

◆ AI Coding Agents

Amazon Kiro

Feedback ⓘ changelog · Nov 15, 2026 CHANGELOG ↗

RANK

Kiro is an agentic development environment that uses spec-driven workflows to plan, build, and maintain software.

> Kiro launched an Android app for managing cloud coding sessions from a phone, added Grok 4.7 with a 500K-token context window across all surfaces, and rolled out unified Enterprise Management policies, a new Approve for Me permission mode, CLI compaction and chat-copy controls, full-screen Mermaid diagrams, MCP sign-in support for Client ID Metadata Document servers, and broader credit-pooling identity provider support.

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

01 **Grok 4.7 model support** NEW 90

Grok 4.7 (from SpaceXSI) is available across Kiro IDE, CLI, Crew, and Web for Pro, Pro+, Pro Max, Power, and Enterprise plan customers in `us-east-1`, with a 500K-token context window, a 2.7x credit multiplier for requests up to 200K tokens, and 5.4x for requests over 200K tokens. It offers four reasoning effort levels — `low`, `medium`, `high`, and `xhigh` — defaulting to `high`, letting users trade cost against depth of self-checking per task.



Pick the right effort level for a long, cross-file bug hunt without burning unnecessary credits.

📍 In the Kiro IDE or Crew model picker, select Grok 4.7, then set reasoning effort to `xhigh` before starting the agent session.

— Exact multipliers, context window and effort levels, with example

changelog-20261115-89133c38 launch-20261112-69821146

02 **Kiro for Android app** NEW 85

Kiro now has an Android app on Google Play: start a session by connecting repositories and optionally attaching a screenshot, while Kiro continues working in a cloud sandbox

after you put the phone down. Review ready changes with file edits rendered as mobile-friendly cards, then let Kiro open a pull request with PR status visible alongside the session. Cloud sessions share conversation history, session state, and synced Kiro configuration across Kiro CLI, Web, and IDE, currently available in `us-east-1` (N. Virginia); enterprise organizations using IAM Identity Center must have an administrator enable cloud sessions before users can access Kiro for Android.

Pick up an in-progress cloud session on your phone and unblock it by answering a clarifying question, without needing a laptop.

> 📍 1. Install Kiro from Google Play and sign in with the same identity used for Kiro CLI, Web, or IDE. 2. Open Cloud Sessions — your existing sessions appear with their conversation and session history. 3. Tap the session that needs your attention, add context or answer the question, and let Kiro continue.

Review a finished change and trigger a pull request entirely from an Android phone.

> 📍 1. Open the completed session in the Kiro Android app. 2. Tap the diff — file edits render as mobile-friendly cards for file-by-file review. 3. If the change looks right, tap Open pull request ; PR status then stays visible alongside the session.

— Full workflow with steps, region and enterprise gate specified `launch-20261112-f0f4287a`

03 **Enterprise Management policy across Kiro surfaces** NEW

80

A single Enterprise Management policy can now be applied across the Kiro IDE (version 1.4.154 or later), Kiro CLI (version 3.1.0 or later), Kiro Web, and Kiro Crew LTS, covering model rules, shell commands, file access, web search, and MCP tool calls for new Kiro profiles in `us-east-1`. Kiro Crew LTS is now available for organizations using Enterprise Management, applying both profile-wide policy rules and Crew-specific controls over messaging channels, approval modes, scheduled jobs, fetch hosts, browser access, and dashboard access.

— Names versions, region and scope but no admin console path `changelog-20261115-89133c38`

04 **CLI compaction threshold control** NEW

80

Set `compaction.autoCompactionThreshold` in `cli.json` to control when automatic compaction starts summarising older conversation history, either as a user-wide default or overridden per workspace (e.g. setting it to 85).

Set a lower compaction threshold so Kiro summarises older history earlier, keeping context lean during long coding sessions.

> 📍 `/settings` Then open Compaction, set `compaction.autoCompactionThreshold` to your preferred percentage (e.g.

60), and confirm. The value is saved to cli.json and applies to the next session you start or load.

Override the global compaction threshold for a specific workspace so it triggers later than the user default.

```json
{
  "compaction": {
    "autoCompactionThreshold": 85
  }
}
```

— Exact config key with worked example values changelog-20261115-89133c38

05 **Approve for Me permission mode in chat** NEW 70

Kiro chat adds a new Approve for Me permission mode alongside Supervised and Accept file changes, where Kiro approves routine actions automatically, blocks risky ones, and asks when unsure. Enable it by default via Kiro: Enable Approve For Me (User) from the Command Palette.

— Exact command given; approval heuristic only sketched changelog-20261115-89133c38

06 **Copy chat history with /copy command** NEW 70

Copy earlier chat responses with `/copy N` or `/copy last N`, e.g. `/copy last 3` to copy the last three messages for pasting into a bug report or sharing with a colleague.

Copy the last three messages — your prompts and Kiro replies interleaved — to paste into a bug report or share with a colleague.

```
$ /copy last 3
```

— Exact command syntax with a concrete usage example changelog-20261115-89133c38

07 **Full-screen Mermaid diagram viewing** NEW 70

Open wide Mermaid diagrams full-screen in the terminal with `Ctrl+]` or `/diagram`, and rebind the key with `chat.keybindings.viewDiagram`.

— Exact keybinding and config key, no example shown changelog-20261115-89133c38

08 **MCP sign-in for CIMD servers** NEW 70

Kiro can now sign in to MCP servers that support Client ID Metadata Documents but not dynamic client registration (such as Tableau's hosted MCP) with no `oauth` block in configuration; a configured `clientId` or `clientMetadataUrl` still takes priority.

— Named config fields but no sample configuration shown changelog-20261115-89133c38

09 **Credit pooling with Okta and Entra ID** IMPROVED 60

Credit pooling now works with Okta and Microsoft Entra ID in addition to AWS IAM Identity Center; manage users, groups, seats, and usage limits from the Access management section of the Kiro console.

— Names providers and console section, no deeper mechanism changelog-20261115-89133c38

THINNER COVERAGE BELOW

10 **Expanded model provider lineup** IMPROVED 30

Kiro now supports models from Anthropic, OpenAI, SpaceXSI, and leading open-weight providers all in one place, widening model choice across IDE, CLI, Crew, and Web.

— Names providers but no access details or config launch-20261112-69821146

└──▷ **ALSO FROM THESE RELEASES**

WAS THIS USEFUL? 👍 👎

Alibaba Qwen Code

Feedback ⓘ desktop-v0.27.4 NOTES ↗

RANK

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

> Qwen Code's latest releases add Feishu and WeCom webhook channel integrations, a Managed Web Terminal, a multi-workspace Goals page, per-turn hook recording, a workspace memory import API, and an experimental unified agent loop, alongside PR-watch operator controls and stricter Auto Mode authorization.

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

01 **Unified agent loop across TUI and ACP rounds** NEW 80

Each `AgentCore`, `OpenTUI`, `Ink TUI`, and `ACP` round, plus speculation model rounds and the `Hosted Harness` model round, now runs through a shared `runAgentLoop` call for consistency. An experimental unified agent loop, enabled via `experimental.unifiedAgentLoop` and `experimental.eventBusAgentLoop`, routes `ACP` rounds through the event-bus kernel.

— Names exact internal call and experimental flags, but status is still experimental.

v0.27.4 desktop-v0.27.4

02 **Feishu and WeCom webhook channel integrations** NEW 77

Adds `Feishu` and `WeCom` webhook support as new outbound channel integrations: a `Feishu` custom bot webhook with optional signing secret, channel control, and scheduled group delivery, and a send-only `WeCom` group webhook that delivers Notify outputs, prompt results, and scheduled-task results. New Web UI connection methods let you connect to `Feishu` and `WeCom` via webhook.

Route Feishu notifications from a running channel to a custom bot webhook with message signing enabled.

```
$ qwen channel start
```

— Names channels, webhook options, and a start command; no full setup walkthrough.

v0.27.4 desktop-v0.27.4

03 **Workspace memory import API in SDK** NEW 75

A new typed SDK API exposes workspace memory import, letting callers preview, submit, inspect, cancel, and retry Codex memory imports from a workspace-bound

context.

– Names the SDK actions available but not exact method signatures. v0.27.4

desktop-v0.27.4

04 **PR-watch dependency provisioning and operator control in serve mode**

IMPROVED 70

Serve mode now provisions PR-watch dependencies and resumes delivery on reconnect; PR-state watching is made operator-only by setting `serve.sessionPrWatch` in the serve daemon configuration.

– Names the exact config key but gives no example value. v0.27.4 desktop-v0.27.4

05 **Multi-workspace Goals page in Web Shell** IMPROVED 60

The Web Shell Goals page now aggregates goals across all workspaces, merging goals from trusted workspaces, showing workspace badges, and letting new goals target a selected workspace.

– Describes the UI behavior fully but points to no API. v0.27.4 desktop-v0.27.4

06 **Per-turn hook execution recording and inspection** NEW 60

Qwen Code now records and queries per-turn hook executions, with read-only Hook pages, SDK access, summaries, and signed snapshots available in session history.

– Lists Hook pages and SDK access but no endpoint or flag names. v0.27.4

desktop-v0.27.4

THINNER COVERAGE BELOW

07 **Multi-segment delivery for long channel replies** IMPROVED 50

Long managed-channel replies are now planned as ordered multi-segment deliveries instead of being truncated, with a visible notice if the segment limit is exceeded.

– Explains the before/after but no config to tune segment limits. v0.27.4

desktop-v0.27.4

08 **Explicit authorization required for risky Auto Mode actions** 50

BREAKING

Auto Mode now requires explicit user authorization for risky actions—destructive, discarding, shared-history, or bulk operations—rather than merely checking relevance.

– Explains the behavior change but no setting to adjust it. v0.27.4

09 **HTTP fallback for terminal tabs, previews, and commands**

IMPROVED

40

Terminal tabs, web previews, local commands, and slash commands now keep working over plain HTTP when native UUID support is unavailable.

— Names affected surfaces but no config or flag to control it. v0.27.4

10 **Managed Web Terminal for browser session access** NEW

35

Adds a Managed Web Terminal for browser-based access to managed agent sessions.

— Single sentence, no UI path or configuration given. desktop-v0.27.4

11 **Client filesystem registration in managed sessions** NEW

30

Client filesystems can now be registered into managed sessions, enabling cross-environment file access.

— One-line description with no mechanism or API detail. desktop-v0.27.4

12 **Improved session search in Web Shell** IMPROVED

20

Improves the session search experience in the Web Shell.

— Vague one-line improvement with no specifics. desktop-v0.27.4

WAS THIS USEFUL?  

yetone magpie

Feedback ⓘ v0.1.4248 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.

> Magpie added four new AI provider integrations (GitHub Copilot, Gemini Code Assist, Antigravity, Grok/SuperGrok), introduced a Daily update channel alongside Stable, and shipped a plugin API that lets agents auto-install their vendor CLI at sign-in — plus slash-command import from six other agents, quota table UI improvements, per-account Codex turn-state tracking, Gemini 3+ thinking-level translation, and a one-click warm-all accounts action.

└──▷ **WHAT SHIPPED · 8 FEATURES** most completely described first ? what's the number?

01 **Daily update channel for releases** NEW 95

A new Daily update channel picks one release per day at 20:00 UTC that passed CI and shipped all required files, alongside the existing Stable channel. Query it via `channel=daily` on `/api/latest`, and download artifacts via the new `/download/daily` and `/download/daily/<platform or file>` endpoints, mirroring the existing stable download paths.

Check which release the Daily channel currently resolves to — useful for scripting automated installs that want fresher builds than Stable.

```
$ curl 'https://api.magpie.example/api/latest?channel=daily'
```

Download the Daily-channel build for a specific platform in a CI pipeline that tracks daily releases.

```
$ curl -L 'https://api.magpie.example/download/daily/<platform or file>' -o magpie-daily
```

— Exact endpoints and runnable curl examples provided v0.1.4239

02 **Automatic vendor CLI install during plugin sign-in** NEW 90

Plugins can declare `auth.cli.name`, auth.cli.installed(), auth.cli.install({ signal }), and `auth.cli.manual` to let magpie automatically install a required vendor CLI during sign-in when it is not present on the machine. magpie also exposes CLIInstalled(ctx, id) and PluginCLIMissing(id) to query whether a plugin's declared vendor CLI is available.

— Exact plugin API functions and behavior named `v0.1.4246`

03 **Slash command import from other agents** NEW 85

magpie can import existing slash commands from agents — Pi's `prompts/*.md`, Claude Code's `commands/*.md`, OpenCode's `commands/*.md` and `command/*.md`, Gemini CLI's `commands/*.toml`, Qwen Code's `commands/*.md` and `*.toml`, and Droid's `commands/*.md` — into the magpie library via the new `commandFind` UI in the Slash commands tab. Commands with the same name and identical content across agents are merged into one library entry; those with the same name but different content are listed separately for the user to choose from.

— Every agent's file path and exact UI location named `v0.1.4238`

04 **Warm up all accounts in one action** NEW 85

A new `warm-all` API endpoint (api('warm-all', { agent })) warms every account currently in use at once, skipping accounts already running or turned off, and reporting each account's outcome individually via `.results`.

— Exact API call signature and skip/report behavior given `v0.1.4227`

05 **Per-account Codex turn-state tracking** IMPROVED 75

magpie now tracks the X-Codex-Turn-State header returned by each Codex account and replays it on every subsequent request within the same turn, so multi-turn tool calls are routed back to the originating account rather than being admitted as new turns.

— Header name and routing mechanism specified, no user action `v0.1.4242`

06 **Collapsible provider rows in quota table** IMPROVED 70

Accounts sharing a provider now collapse into a single row showing averaged column shares, expandable on click, with the open/closed state persisted in `magpie.quotaTableOpen` (keyed by provider). The Account column is now hidden from the quota table when no row carries an account name.

— Config key and exact UI behavior named `v0.1.4248`

07 **Thinking-level translation for Gemini 3+** IMPROVED 65

magpie now translates thinking budget parameters to thinking level for Gemini 3+ models when routed through an Anthropic or Chat relay (such as peta-gemini), avoiding the budget rejection error Gemini 4 returns for `thinkingBudget` requests.

— Mechanism and specific error avoided are named `v0.1.4235`

THINNER COVERAGE BELOW

08 **New AI provider integrations added** NEW 45

magpie added support for GitHub Copilot, Gemini (Code Assist), Antigravity (via the `agy` CLI), and Grok (SuperGrok) as AI providers it can switch agent models to.

— Lists four providers added, no setup mechanism given `product docs`

WAS THIS USEFUL? 👍 👎

Alibaba OpenCodeReview

Feedback ⓘ v1.13.10 NOTES ↗

RANK

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's v1.13.10 release focuses on workspace-mode CLI ergonomics, adding resumable reviews, broader diff context, a smarter default ref, and friendlier output rendering.

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

01 **`--to` defaults to HEAD** IMPROVED 80

When only `--from` is given, `--to` now defaults to `HEAD` automatically, removing the need to specify the target ref explicitly, e.g. `ocr review --from abc1234` reviews all commits up to HEAD.

Review all commits since a known good commit up to HEAD without having to specify `--to` explicitly.

```
$ ocr review --from abc1234
```

— Exact flag, default value, and runnable command given v1.13.10

02 **Resume interrupted workspace-mode reviews** NEW 60

The new `--resume` flag lets users continue an interrupted workspace-mode review instead of restarting it from scratch.

— Names the exact flag but no detail on state tracking v1.13.10

THINNER COVERAGE BELOW

03 **Comparison-side code access in workspace mode** NEW 40

Workspace mode reviews can now access the comparison-side code, enabling reviews that read both sides of a diff for richer context.

— Explains benefit but no concrete mechanism or config v1.13.10

04 **Markdown and styled CLI output** IMPROVED 35

The CLI now renders Markdown summaries and styled progress output for more readable review results.

– Thin description with no named format or path v1.13.10

WAS THIS USEFUL? 👍 👎

Letta Code Feedback ⓘ v0.34.62 NOTES ↗ RANK

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

> Letta Code improved its monitor feature to resume saved watches after interruption and correctly distinguish terminal states from in-progress ones.

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

01 Monitor resumption and terminal-state tracking IMPROVED

28

Monitors now resume saved watches after interruption, and the system distinguishes terminal states from in-progress ones so interrupted watches can pick back up correctly.

– Describes behavior change but no command or config named v0.34.62

WAS THIS USEFUL? 👍 👎

codealta

Feedback ⓘ 4.4.1 NOTES ↗ RANK

Codealta is an agentic Development Environment

> Codealta 4.4.1 lets sessions trigger Adaptive code reviews directly, adds per-space browser link routing, and extends scratchpad cleanup to macOS and Linux.

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

01 **Session-initiated Adaptive code reviews** NEW 80

Start an Adaptive code review from within the session using `alta adaptive review`, which launches critics and returns immediately; check progress with `alta adaptive status`. The verdict is delivered as a prompt rather than dropped, enabling workflows like gating a push on a passed review.

Gate a push on a completed Adaptive review: start the critics, poll until a verdict arrives, then push only if it passed.

```
$ alta adaptive review && alta adaptive status
```

— Names exact commands and behavior, runnable as shown 4.4.1

02 **Per-space link routing via config** NEW 75

The `[links]` table in `.alta/config.toml` controls which browser and profile open links in a given space, overriding the workspace-level Links open in and Browser profile settings found in General.

Route all links in a project to a specific browser and profile, overriding the workspace default.

```toml
[links]
browser = "Chrome"
profile = "Work"
```

— Exact config key, file path, and override behavior given 4.4.1

THINNER COVERAGE BELOW

03 **Cross-platform scratchpad cleanup** IMPROVED 50

Scratchpad cleanup at startup now runs on macOS and Linux, not only Windows, without following symlinks or crossing mount boundaries. `Settings > Storage` shows when

cleanup is inactive on a given system.

— Describes mechanism and UI location but no config surface 4.4.1

WAS THIS USEFUL? 👍 👎

wrongstack

Feedback ⓘ v1.7.1 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's desktop builds now check for and install signed updates automatically, with new controls to manage when and how updates happen.

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

01 **Desktop auto-update with Install & restart** NEW 70

Desktop builds check for a newer stable release at launch and every six hours, with options in Appearance settings to disable checks, run a check manually, or dismiss an offered version via `Later`. Windows and macOS builds can now `Install & restart` using verified updater packages, and the release pipeline verifies all four Windows/macOS targets before publishing `latest.json`.

— Names settings path and update file but no CLI command v1.7.1

WAS THIS USEFUL? 👍 👎

Yuval Sarel cones

Feedback ⓘ v0.4.13 NOTES ↗ RANK

Cones is a terminal dashboard for managing coding-agent sessions across projects.

> cones overhauled session organization with a new `cones place` command replacing `cones tag`, richer `cones ls` filtering, configurable sort order, and folder/tag grouping with keyboard and drag-and-drop reordering.

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

01 **New cones ls filters and launched_by field** NEW 85

`cones ls` gains `--runs` to include non-supervised runs and `--job` to show a job's full history; `--json` output now also exposes a `launched_by` field for launched sessions.

Review a job's full run history, including runs that are no longer supervised, to audit past agent activity.

```
$ cones ls --job nightly-audit --runs --json
```

— Flags, field and a runnable example command are all given. v0.4.13

02 **Folder/tag grouping and manual reordering** NEW 85

Sessions can now be grouped by folder or by tag, with each session able to appear in multiple places (its folder, `# NAME` tag groups, `folder › # NAME` headings, and pinned). A new placement checklist opened with `ctrl+l` lets you assign a session to existing or new tag groups without leaving the list; `ctrl+b` reorders sessions from the keyboard, `ctrl+t` toggles pinned state, and mouse drag now supports dropping rows onto headings.

— All keybindings and grouping labels named, no screenshots attached. v0.4.13

03 **cones place replaces cones tag** BREAKING 75

`cones tag` is replaced by the new `cones place` command for managing session placement; existing state in `state/tags.json` and `pins.json` converts automatically on upgrade.

— Names old/new command and migrated state files, no usage example. v0.4.13

04 **Configurable session sort order** NEW 65

Session sort order is now configurable via the new `sort` setting in cones config, choosing between started, last active, or manual order.

Set sessions to sort by last active time so the most recently used sessions surface at the top of the dashboard list.

```
$ cones config set sort last-active
```

— Config key and example command given, mechanism unexplained. v0.4.13

THINNER COVERAGE BELOW

05 **Unpriced fast-mode sessions** IMPROVED 40

Claude Code fast mode, US-only inference replies, and fast Codex threads are now left unpriced rather than charged at standard rates.

— States the change but gives no mechanism or UI path. v0.4.13

└──▷ **BREAKING ON UPGRADE**

! `cones tag` is replaced by `cones place`; existing tag state in `state/tags.json` and `pins.json` is converted automatically on upgrade.

WAS THIS USEFUL? 👍 👎

coffee-cli

Feedback ⓘ v4.0.3 NOTES ↗ RANK

coffee-cli is a clean AI terminal workspace with quick input, folder switching, copy & paste, and image pasting.

> coffee-cli v4.0.3 adds a responsive phone/narrow-window layout and lets users configure what closing the window does.

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

01 **Configurable window-close behavior** NEW 60

In Settings → General, users can choose what closing the window does: ask each time, keep running, or quit. Keeping running puts coffee-cli in the Windows notification area, the macOS menu bar, or minimizes it on Linux, and macOS gains a menu bar icon to open the window or jump to a running tab.

— Names exact settings path and per-OS behavior. v4.0.3

THINNER COVERAGE BELOW

02 **Phone and narrow-window responsive layout** NEW 55

coffee-cli now shows a phone layout when controlled from a mobile browser — side panels become drawers, all tabs collect into one menu, and the launch page and Settings adapt to the phone screen. On a computer, the side panels also become drawers when the window is narrower than 1000px.

— Describes behavior and a concrete 1000px threshold but no exact UI path. v4.0.3

WAS THIS USEFUL? 👍 👎

mksglu Context Mode

Feedback ⓘ v2.2.0 NOTES ↗ RANK

Context Mode is a free, local plugin that keeps tool output out of your coding agent's context window. Sandbox and SQLite FTS5 search, 99% fewer bytes in context.

> Context Mode v2.2.0 adds OMP (oh-my-pi) as its 11th supported agent, with full plugin, doctor, upgrade, and system-prompt integration.

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

01 **OMP agent support added** NEW 91

Context Mode now treats OMP as a valid `AgentId` with a Pi-shaped table row, an `omp/index.js` wrapper, and an `omp` key in `package.json`. It reads `contextMode.excludeTools` from `api.pi.settings` per call when running under OMP to gate which `ctx_*` tools are exposed per prompt, and all nine `ctx_*` tools register with `loadMode essential` while OMP registers tools at load time with one turn per prompt. The `ctx_upgrade` command on OMP prints the OMP plugin upgrade path, writes no host config, and enforces one plugin root per package; `doctor` now names OMP, reads the OMP plugin store, and prints OMP-specific verbs and host-limit info; and setup merges/uninstall drops context-mode in OMP's `disabledServers` list when OMP imports it.

Add Context Mode to an OMP environment and verify the installation is healthy.

```
$ omp plugin upgrade context-mode
```

— Names exact config keys, commands, and files; fully mechanistic. v2.2.0

WAS THIS USEFUL? 👍 👎

mjolnir

Feedback ⓘ v2.71.0 NOTES ↗ RANK

Mjolnir is a manage Codex, Claude Code, Muse Code, Kimi Code, Grok Build, and DeepSeek Harness with durable sessions, isolated environments, quotas, and remote control. A Rust terminal and web control plane for Agent Client Protocol (ACP) coding agents.

> Mjolnir laid groundwork in its Observe subsystem for exporting session observability data and added capability-gated enforcement of DSH configuration.

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

01 **Capability-gated DSH enforcement** IMPROVED 20

Mjolnir now enforces DSH configuration through capability gating within the Observe subsystem.

— Names mechanism but lacks concrete config keys or commands v2.71.0

02 **Observe export foundation** NEW 15

Mjolnir adds an export foundation to the Observe subsystem, laying groundwork for exporting session observability data. No export formats, endpoints or commands are specified yet.

— Named as groundwork only; no concrete surface or format given v2.71.0

WAS THIS USEFUL? 👍 👎

zot

Feedback ⓘ v0.4.78 NOTES ↗ RANK

zot manages coding agent workflows from the command line.

> zot v0.4.78 adds a configurable timeout for shell tool execution.

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

01 **Configurable shell execution timeout** NEW 35

zot adds a configurable shell execution timeout setting that controls how long shell tool runs are allowed to execute before timing out.

– Names the setting but not its key or default value. v0.4.78

WAS THIS USEFUL? 👍 👎

◆ Model Training & Fine-Tuning

Thinking Machines Lab Tinker

Feedback ⓘ

updated · Nov 15, 2026 NOTES ↗ RANK

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

> Tinker's November 15 snapshot adds StableReinforce RL stabilization, the Dimuon optimizer for LoRA fine-tuning, free checkpoint storage, and new model availability including Inkling 2 Small, Kimi K3.1 Pro, and expanded vision support for DeepSeek v4.1 Flash and GLM 5.3 Flash.

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

01 **StableReinforce RL stabilization** NEW 80

Enable stable RL runs by sampling with `record_stability_info=True` and passing `loss_fn="stable_reinforce"` to `forward_backward_async`; StableReinforce keeps training stable when the sampling and training policies diverge. The feature is alpha and its API may change.

– Exact flags and API named but marked alpha/unstable snapshot-20261115

02 **Dimuon optimizer for LoRA fine-tuning** NEW 73

Use the Dimuon optimizer for LoRA fine-tuning by setting it at training-client creation time; it reaches lower loss in fewer steps than AdamW on supervised fine-tuning workloads, with a recommended learning rate of 1e-3 (10–100x larger than AdamW).

– Named optimizer and learning rate but no exact API call shown snapshot-20261115

03 **Free checkpoint storage plan** NEW 70

Tinker now offers a free checkpoint storage plan covering up to 100 GB or 100 checkpoints per org; additional storage is billed at $0.10 per GB per month, managed from the Billing tab of the console.

Move an org to the paid checkpoint storage plan after hitting the 100 GB or 100-checkpoint free-tier limit.

📍 1. Open the Tinker console. 2. Go to Billing tab. 3. Select the paid checkpoint storage plan.

– Exact limits and pricing with UI navigation to upgrade snapshot-20261115

04 **Inkling 2 Small fine-tuning support** NEW 65

Inkling 2 Small, a 594B-parameter (17B active) Mixture-of-Experts model that accepts image inputs and supports variable effort, is now available for fine-tuning on Tinker.

— Detailed architecture but no concrete steps to start `snapshot-20261115`

THINNER COVERAGE BELOW

05 **Vision input support for DeepSeek v4.1 Flash and GLM 5.3 Flash**

IMPROVED 40

DeepSeek v4.1 Flash and GLM 5.3 Flash now accept image inputs in prompts for both sampling and fine-tuning on Tinker, using the same input format as other vision models.

— Names both models and capability but no format specifics given `snapshot-20261115`

06 **Kimi K3.1 Pro availability** NEW 30

Kimi K3.1 Pro from Moonshot is now available on Tinker with a 256K context window.

— Bare availability note with one spec, no usage detail `snapshot-20261115`

WAS THIS USEFUL? 👍 👎

◆ Coding Agent Add-ons

cezar

Feedback ⓘ v0.20.0 NOTES ↗ RANK

cezar is an open-source orchestrator and ADE for running Claude Code, Codex, OpenCode, Pi and other AI coding agents in parallel — locally or 24/7 on your own server.

> Cezar's biggest release in this window adds a Manager mode for global agents that can connect to repositories, approve merges and chat with you, alongside tighter automation-to-manager reporting, richer task input handling, and repo-local agent command discovery.

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

01 Automation results reported to manager chat NEW 90

Link an automation's results to a manager's chat using `cez automation add` with `--report-to-manager`, `--manager <id>`, and `--report-mode post-result|failure-only|none`. Change an existing link with `cez automation report <id>`, and view it with `cez automation list --json`.

Route every automation result — including failures — to a manager's chat so it can act on them without you polling.

```
$ cez automation add --report-to-manager --manager <id> --report-mode post-result
```

— Full flag set plus runnable example command v0.20.0

02 Long-form task input via file, stdin, or message reference NEW 85

File long text into a task without truncation using `cez-manager task create` and `task say` with the new `--attach-last-user-message`, `--from-message`, `--file`, `--stdin`, and `--allow-placeholder` flags.

Paste a long transcript into a manager task without it being truncated or retyped into a command.

```
$ cez-manager task say --file transcript.md --allow-placeholder
```

— Named flags with a runnable example command v0.20.0

03 Manager task follow/unfollow notifications NEW 80

Subscribe a manager to live task-state updates with `cez-manager task follow|unfollow <run>`, receiving task-card notifications within seconds when a task finishes, fails, is cancelled, or waits for input.

Get notified the moment a dispatched task finishes or stalls so a manager can act on it immediately.

```
$ cez-manager task follow <run>
```

— Exact command and behaviour with example v0.20.0

04 **Toggleable PR-watch polling for monitoring tasks** NEW

75

Turn off PR-watch polling for monitoring tasks with `CEZ_MONITORING_PR_WATCH=0`; when enabled, a monitoring task checks the PR with one `gh pr view` per wake-up and skips the model turn if nothing changed.

— Exact env var and mechanism, no example run v0.20.0

05 **Manager mode for global AI agents** NEW

70

A new Manager mode lets you create global agents, connect repositories read-only or read-write, run saved jobs, approve merges and automation proposals, and chat with agents directly. It can be disabled with `CEZ_MANAGERS=0`.

— Disable flag given but no setup command shown v0.20.0

06 **Repo-local agent command discovery** NEW

70

Repo-local agent commands in `.claude/commands`, `.codex/prompts`, `.cursor/commands`, `.opencode/command`, and `.github/prompts` now appear in the `/` list, Skills page, and ⌘K, with `$ARGUMENTS` substitution.

— Paths named but no command, only UI surfaces v0.20.0

07 **Manager access to claude.ai connectors and MCP servers** NEW

65

Managers now access your claude.ai connectors (Gmail, Calendar, Drive, and others) and MCP servers defined in your Claude config, with per-connector Off, Read, or Read+write permissions and per-tool overrides.

— Permissions model named but no config example v0.20.0

08 **Per-child task agent account selection** NEW 60

Pick a child task's agent account with `cez task create --account <profile>`; a child on the same runner inherits its parent's account.

— Exact flag given but no example v0.20.0

THINNER COVERAGE BELOW

09 **Cockpit installable as standalone PWA** NEW 50

Install the cockpit as a standalone PWA on a phone or desktop home screen, opening directly behind a Basic auth gate with live data and no service worker.

— Describes behaviour but no setup steps v0.20.0

10 **Searchable model pickers for long option lists** IMPROVED 40

Model pickers with more than 15 options now include a search box that matches by name and id, with Enter selecting the top match.

— Minor UI change with threshold but no steps v0.20.0

WAS THIS USEFUL? 👍 👎

Worktrunk

Feedback ⓘ v0.82.0 NOTES ↗ RANK

Worktrunk manages Git worktrees from the command line for parallel AI coding-agent workflows, with hooks, status listings, and LLM-written commits.

> Worktrunk's latest release adds safety-net snapshots before forced worktree removal, a Claude Code branch-diff pane, and more resilient JSON output from `wt switch`, alongside a breaking split of Codex activity hooks into a separate package.

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

01 **Backup snapshot on forced worktree removal** NEW 85

`wt remove --force` now snapshots the worktree, including untracked files, to `refs/wt-backup/<branch>` before deleting it, and prints a recovery command so the removal can be undone.

Force-remove a worktree safely — the snapshot at `refs/wt-backup/<branch>` lets you recover uncommitted work if the removal was a mistake.

```
$ wt remove --force
```

— Names exact ref path and flag with recovery command v0.82.0

02 **Resilient JSON output from wt switch** IMPROVED 85

`wt switch --format=json` now retains confirmed results on stdout even when a later step fails, routing diagnostics to stderr and signalling failure with a non-zero exit status, so scripts can parse the confirmed outcome separately from error detail.

Capture a machine-readable result from a switch even when a post-switch step fails, so CI scripts can parse the confirmed outcome and inspect stderr for the failure detail separately.

```
$ wt switch --format=json feat 2>switch-errors.log | jq .
```

— Exact flag and stdout/stderr/exit-code behaviour given v0.82.0

03 **Codex activity hooks split into separate package** BREAKING 80

Codex activity hooks have moved from `worktrunk` to `worktrunk-markers`; existing installs must run `codex plugin add worktrunk-markers@worktrunk` to retain

activity markers after updating.

— Exact migration command and package names provided v0.82.0

THINNER COVERAGE BELOW

04 **Claude Code branch-diff pane via /wt-diff** NEW 55

A new `/wt-diff` pane in Claude Code shows diffs for every worktree the session has worked in, including work detected from subagents.

— Names the command but no mechanism detail beyond this v0.82.0

└──▷ **BREAKING ON UPGRADE**

! Codex activity hooks are split into a separate package: existing installs must run `codex plugin add worktrunk-markers@worktrunk` to retain activity markers after updating, because hooks have moved from `worktrunk` to `worktrunk-markers`.

WAS THIS USEFUL? 👍 👎

◆ Speech & Voice AI

coreai-kit

Feedback ⓘ 0.8.0 NOTES ↗ RANK

Coreai-kit is a swift SDK for running chat, vision and speech models on iPhone and Mac with Apple's Core AI. Model download and caching, FoundationModels integration, and runnable examples with documented OS, SDK and model requirements.

> coreai-kit 0.8.0 expands its model catalog from 67 to 80 models and adds checkpoint-based chat resumption that cuts time-to-first-token from 837.8 ms to 38.6 ms, alongside new quantile forecasting and typed-decision APIs, tool calling for LFM3, per-platform engine selection, and several breaking API changes.

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

01 **Checkpoint-based resumption for hybrid-model chats** NEW

90

`InferenceOptions.checkpointAfter` saves inference state at the last turn-end token so a hybrid-model chat can resume from a checkpoint instead of re-prefilling the whole conversation, cutting time-to-first-token from 837.8 ms to 38.6 ms on LFM3-1.7B on an M4 Max.

— Names exact config key and concrete before/after benchmark numbers 0.8.0

02 **Quantile forecasting with Timestral** NEW

85

The new `KitTimestralForecaster` and CoreAI.forecastQuantiles(_ options:) run on catalog id `timestral-209m` (`kind: quantileForecasting`): feed in a number series and get back `horizon` steps of 21 quantiles (0.01–0.99) as a `QuantileForecast` with `median` and quantile(_:) accessors.

— Names API, catalog id, and output shape precisely 0.8.0

03 **Typed decision scoring with KitD1Decider** NEW

80

The new `KitD1Decider` runs on catalog id `d1-3b` (`kind: tokenDecision` , `format: optionRows`): pass a text or JSON state plus images and typed questions (noul, choice, score), and each option comes back with a probability rather than generated text.

— Names API and catalog id but no runnable example given 0.8.0

04 **Init signature changes break function references** BREAKING

78

Taking the address of SpeakerSegment.init(speaker:startFrame:endFrame:), KitDiarizer.segments(from:bridgeFrames:), CatalogEntry.init(...), CatalogEntry.Variant.init(path:sizeMB:), or VLArchitecture.init(...) by their 0.7.3 full names no longer compiles because each gained new parameters; wrap in a closure instead (e.g. `let make = { SpeakerSegment(speaker: $0, startFrame: $1, endFrame: $2) }`).

— Names all affected initializers plus a concrete workaround `0.8.0`

---

05 **Thinking toggle for LFM3 and Qwen3 chats** NEW 75

`ChatSession.Configuration.enableThinking` controls whether LFM3 and Qwen3-family responses include a reasoning trace: leaving it `nil` uses the template default, while setting it to `false` requests the thinking-off prompt so the full response budget goes to the answer.

Suppress the reasoning trace on LFM3 so the full response budget goes to the answer rather than thinking.

```swift
var config = ChatSession.Configuration()
config.enableThinking = false
let chat = try await ChatSession(model: modelID, configuration:
config)
```

— Config flag with runnable code example included `0.8.0`

---

06 **Tool calling support for LFM3** NEW 75

`KitLanguageModel` declares `.toolCalling` for bundles carrying LFM3's `<function=` / `<parameter=` tokens, renders tool turns matching transformers' `apply_chat_template`, and parses XML calls with values typed by each tool's schema; the LFM3 1.7B catalog model calls tools through `LFM3ToolDialect`.

— Explains parsing mechanism in detail, no sample invocation `0.8.0`

---

07 **Model catalog expansion** NEW 75

Expands the model catalog from 67 to 80 models, adding catalog id `lfm3-1.7b` (`kind: chat`) for LFM3 1.7B — which thinks by default, calls tools through `LFM3ToolDialect`, and includes an iPhone-specific `ios-<arch>/` variant targeting the static-shape Neural Engine build — catalog id `minicpm-v-4.7-1b` (`kind: vlm`) for MiniCPM-V 4.7 1B, loaded through `VLModelID.miniCPMV47` and `VLArchitecture.miniCPMV47` with 448 px input, 64 image tokens, M-RoPE over T/H/W, and new `VLArchitecture` fields `ropeShiftFrameTail` and `assistantPrefix`, plus six new decision models across five new kinds, zero-shot text classification, `Fun-ASR-Nano`, `Audio8-TTS`, and 8-speaker diarization.

– Names catalog id and variant, lacks usage example Concrete specs named, no usage code shown Counts and names given but no per-model mechanism

0.8.0

08 **Catalog variant metadata enhancements** NEW 73

Adds `CatalogEntry.Variant.architectures` and `AIModel.deviceArchitectureName` so apps can inspect which iPhones have a compiled bundle before downloading, with ModelCatalog.entry(forID:) now throwing `unsupportedDevice` before any download begins when the device isn't covered, and adds `CatalogEntry.Variant.engine` to let a catalog variant select an engine per platform, overriding `CatalogEntry.resolvedEngine` for that platform, which ChatSession(catalog:), `TypedDecisions`, and `KitTextNormalizer` all pick up automatically.

– Names exact APIs and thrown error, lacks example Names API surface and consumers, no code example

0.8.0

09 **VibeVoicePaths.inBundleDir now throws on device mismatch** 65

BREAKING

VibeVoicePaths.inBundleDir(_:glueDir:voicesDir:embedTokens:) now throws when the folder contains only graphs compiled for other devices — callers must add `try`.

– Exact function named with required migration step 0.8.0

10 **New enum cases break exhaustive switches** BREAKING 65

Any `switch` without `default` over `CatalogEntry.Kind`, `Decision.Format`, `VisionError`, `KitBundleError`, `VibeVoiceError` or `VoxCPMError` no longer compiles — each enum gained new cases; add a `default` branch.

– Lists affected enums and required fix verbatim 0.8.0

└──▷ **BREAKING ON UPGRADE**

! VibeVoicePaths.inBundleDir(_:glueDir:voicesDir:embedTokens:) now throws when the folder contains only graphs compiled for other devices — callers must add `try`.

! Any `switch` without `default` over `CatalogEntry.Kind`, `Decision.Format`, `VisionError`, `KitBundleError`, `VibeVoiceError` or `VoxCPMError` no longer compiles — each enum gained new cases; add a `default` branch.

! Taking the address of SpeakerSegment.init(speaker:startFrame:endFrame:), KitDiarizer.segments(from:bridgeFrames:), CatalogEntry.init(...), CatalogEntry.Variant.init(path:sizeMB:), or VLArchitecture.init(...) by their 0.7.3 full names no longer compiles because each gained new parameters; wrap in a closure instead (e.g. `let make = { SpeakerSegment(speaker: $0, startFrame: $1, endFrame: $2) }`).

WAS THIS USEFUL? 👍 👎

◆ AI App Builders

Replit Agent

Feedback ⓘ announced · Nov 12, 2026 BLOG ↗ RANK

Replit Agent is an AI coding agent that builds, tests, and deploys applications in Replit.

> Replit Agent expanded beyond coding into business operations, adding the ability to run Google Ads campaigns and make agentic purchases directly from the chat interface.

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

01 **Agentic commerce purchasing for Core and Pro users** NEW

80

Given a plain-language request, Replit Agent locates products, prepares a merchant checkout via a Browserbase-isolated browser session, and waits for owner approval before any charge is made, powered by Stripe's Link network and Visa payment infrastructure. Available to Core and Pro users in the US.

Have Replit Agent source physical supplies for your business, with a spend cap and owner approval before any charge is made.

> 📍 In Replit chat, type: 'Find 50 shipping boxes that fit this product, under $75 including delivery.' Agent will locate options, prepare the merchant checkout, and pause for your approval before completing the purchase.

— Names infrastructure and approval mechanism; example shows prompt but no API

launch-20261112-8c680708

02 **Google Ads campaign management in chat** NEW

75

Pro users can connect or create a Google Ads account and have Replit Agent run campaigns across Search, YouTube, Display, and Discover — drafting ad copy, generating images, recommending daily budgets, and launching campaigns without leaving Replit chat. Agent also sets up conversion tracking so campaigns can be measured by sign-ups and sales rather than clicks alone.

Launch a targeted Google Ads campaign from Replit without a separate ads platform login, specifying goal, budget, and duration in plain language.

> 📍 In Replit chat, type: 'Get us in front of contractors who bid on public projects. I'm willing to spend $40 a day for 30 days.' Agent drafts ad copy, generates images, recommends targeting, and guides you through review and launch.

— Names platforms and steps; no UI path or command syntax given launch-20261112-8c680708

WAS THIS USEFUL? 👍 👎

AGENT

◆ AI Agent Frameworks

Arcade

Feedback ⓘ changelog · Nov 15, 2026 CHANGELOG ↗

Arcade provides an actions runtime that lets AI agents securely call external services with authorization and governance controls.

Arcade adds a `purge-execution-logs` CLI command, API-key/gateway columns on the execution-history table, multi-account OAuth profiles, and new tools across GitHub, Linear, Slack, Google Docs, and Microsoft Teams.

Filter and report tool execution history by which API key and gateway ran each call, once the automatic schema migration adds `api_key_id` and `gateway_id` columns to the tool-execution table.

Run `purge-execution-logs` on the engine image to delete old tool-execution rows; without flags it reports what it would delete rather than acting.

GitHub toolkit now adds tools to search issues and pull requests with GitHub's query syntax, and a new tool lists conversation comments across a repository.

Linear toolkit adds a tool to list a project's or initiative's change history, a tool to list comments across a team or project since a given time, and `ListInitiatives` rows now report health, owner, and key dates.

Slack toolkit adds a tool to delete owned files in one batch call.

+8 more

NEW INTEGRATION Snipe-IT toolkit

└──▷ BREAKING ON UPGRADE

- ! The Figma toolkit tools are removed entirely and the previous 1.x line is no longer served; review pinned toolkit versions before upgrading.
- ! Word, Excel, and PowerPoint file-creation tools now fail by default when the target file already exists instead of silently overwriting it.
- ! Word and Excel now request narrower Graph file scopes; review stored Graph app consent before upgrading.
- ! Google Forms: `CopyForm` with no destination now places the copy in the caller's My Drive instead of beside the source, dropping any inherited shared-drive or folder sharing.
- ! Google Slides: `CreatePresentationFromTemplate` with no destination now places the copy in the caller's My Drive instead of beside the template, dropping any inherited sharing.
- ! The tool-execution history API only reports which API key and gateway ran a call after the automatic schema migration (`api_key_id` and `gateway_id` columns) has

been applied.

WAS THIS USEFUL? 👍 👎

AutoGPT

Feedback ⓘ autogpt-platform-beta-v0.11.3 NOTES ↗ RANK

AutoGPT automates multi-step workflows with continuous AI agents that users can build and deploy.

> AutoGPT's beta platform opened account deletion to all users with a scriptable CLI flag, added Sign in with Apple, visual answer cards for Otto, and an improved expert chat sidebar, alongside brief mentions of mobile server foundations, remote Local PC execution, and new Gemini model support.

└──▷ **WHAT SHIPPED · 8 FEATURES** most completely described first ? what's the number?

01 Account deletion opened up with cleanup tooling NEW

72

Account deletion no longer requires admin access and is now available to every user. Deletion can be performed via the `erase-account --no-email` CLI flag to skip the confirmation email in automated or scripted contexts, and a one-off cleanup job deletes storage bucket folders belonging to already-deleted accounts.

Erase a user account immediately in automated or scripted contexts where sending a confirmation email would block the flow.

```
$ erase-account --no-email
```

— Names exact CLI flag and cleanup job, with runnable example

autogpt-platform-beta-v0.11.3

THINNER COVERAGE BELOW

02 Expert chats sidebar improvements IMPROVED

50

The sidebar now shows all chats for every expert, with real message counts and per-expert paging.

— Names concrete UI behaviour but no exact navigation path autogpt-platform-beta-v0.11.3

03 Sign in with Apple NEW

35

AutoGPT adds Sign in with Apple support for both native and browser authentication flows.

— States scope but no config or UI path given autogpt-platform-beta-v0.11.3

04 **Expert reset to original state** NEW 26

Users can reset an expert back to its original state.

— One-line capability with no further detail `autogpt-platform-beta-v0.11.3`

05 **Otto visual answer cards** NEW 23

Otto can now answer with visual cards instead of plain text responses.

— Bare description with no mechanism or example `autogpt-platform-beta-v0.11.3`

06 **Gemini 4 Hydrogen and Carbon model support** NEW 20

AutoGPT adds support for the Gemini 4 Hydrogen and Carbon models, as noted in the release summary.

— Names specific models but no usage detail `autogpt-platform-beta-v0.11.3`

07 **Remote Local PC execution targets** NEW 13

AutoGPT adds support for remote Local PC execution targets, per the release summary.

— Named but undescribed in source material `autogpt-platform-beta-v0.11.3`

08 **Native mobile server foundations** NEW 10

This release lays foundations for a native mobile server, as noted in the release summary.

— Only mentioned in passing, no detail given `autogpt-platform-beta-v0.11.3`

WAS THIS USEFUL? 👍 👎

cotal

Feedback ⓘ v0.164.14 NOTES ↗ RANK

Cotal is the open standard for agent coordination

> Cotal shipped native connection introspection across delivery lanes, hosted delivery services, endpoints and feeds, new space-level signing-key rotation and validation APIs, verified update delivery for Windows CLI and macOS desktop, and a breaking change that makes signing-key rotation preserve account policy instead of resetting it.

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

01 Space account signing-key rotation and validation APIs NEW

95

Rotate or roll back a space's data signing key without rewriting the shared broker record using replaceSpaceAccountAuth(store, record) in `@cotal-ai/workspace`, which writes only that space's account record through a `SecretStore` and reads it back after writing. validateSpaceAccountAuth(broker, record, space?) in `@cotal-ai/core` exposes the account-level checks of `validateSpaceAuth` for any space on a broker, with `validateSpaceAuth` itself restricted to the space the operator is named for.

— Exact function signatures and modules given, runnable v0.164.10

02 Native connection introspection across delivery APIs NEW

85

Query a delivery lane's data-account connections with nativeConnections() on the `DeliveryLane` returned by `openDeliveryLane`, surfacing the server ID and client ID of each connection. `startDeliveryService` now returns a `DeliveryServiceHandle` whose nativeConnections() lists endpoint, membership-rw, and timer-writer connection IDs; CotalEndpoint.nativeConnections() and MembershipFeedHandle.nativeDataConnections() expose the same bindings for a single endpoint or feed.

— Names four concrete APIs but no usage example v0.164.12 v0.164.9

03 Signing-key rotation now preserves full account policy

85

BREAKING

`rotateDataAccountSigningKey` now preserves an account's full policy — connection limit, leaf-node limit, per-user revocations, and expiry — across signing-key rotations instead of resetting to stock defaults. Callers that relied on rotation to reset limits or clear revocations must now change the claim themselves, and rotation now refuses to re-issue claims it cannot preserve whole: claims signed by another operator, claims for a different account than the auth's account seed, expired or not-yet-valid claims, claims

with an invalid revocation map, claims with top-level JWT fields outside the registered set, and claims with a scoped signing key.

– Clear migration note but no code example v0.164.9

04 **Windows CLI automatic update via Task Scheduler** NEW

70

Windows CLI installations now stay current automatically through a Task Scheduler update helper that guards the upgrade with a native job, confirms the process list is empty before proceeding, restarts the verified daemon service under one activation deadline, and records activation separately from the installed release.

– Mechanism detailed but update is automatic, nothing to invoke v0.164.14

05 **macOS desktop update installation and verification** IMPROVED

65

Cotal's macOS desktop app now installs verified app updates while hidden or after Restart to update is pressed, holding replacement back during active app operations and relaunching without adding a daemon restart path. It also verifies signed update downloads when the build includes an updater key, and the Check now button re-checks the update manifest on demand.

– Describes UI paths but mechanism partly implicit v0.164.14 v0.164.9

06 **Native connection custody for auth authority plane** NEW

60

Hosts can now compare tracked server and connection IDs with the broker's account census and await terminal connection closure for a standalone auth authority plane via native connection custody, without receiving credentials or raw connection objects.

– Mechanism named but no API signature given v0.164.9

THINNER COVERAGE BELOW

07 **CLI selects verified retained execution generations at startup**

NEW

50

Cotal's CLI now selects verified retained execution generations at startup, so the built-ins-only bootstrap accepts an artifact-bound descriptor, validates the retained package inventory, and runs the selected generation.

– Behaviour described but no flag or command named v0.164.10

└──▷ **BREAKING ON UPGRADE**

! Callers that relied on `rotateDataAccountSigningKey` to reset an account's limits or clear its revocations must now change the claim themselves; rotation no longer resets those fields

to stock defaults.

! `rotateDataAccountSigningKey` now refuses to re-issue claims it cannot preserve whole: claims signed by another operator, claims for a different account than the auth's account seed, expired or not-yet-valid claims, claims with an invalid revocation map, claims with top-level JWT fields outside the registered set, and claims with a scoped signing key.

WAS THIS USEFUL? 👍 👎

opencrabs

Feedback ⓘ v0.6.0 NOTES ↗ RANK

Opencrabs is the all-in-one AI agent. Build mobile apps, frontend, backend, manages files, deep research, schedule tasks and events.

> OpenCrabs v0.6.0 ships a full desktop GUI alongside the terminal app, bundling packaging across all five targets, plus a new settings card, theme/font/transparency controls, Seeker universal search, transcript forwarding with provenance, and an in-window Projects screen.

└──▷ WHAT SHIPPED · 9 FEATURES most completely described first ? what's the number?

01 **Theme, font, and transparency settings** NEW 95

Open a theme modal with `ctrl+shift+t` or `/theme` offering four modes — `auto` (follows system), `orgn` (dark palette), `light` (soft light palette), and `custom` (accent from a hue wheel) — set via `[gui] theme` with `[gui] custom_accent` for the accent color. Set `[gui] font` and `[gui] mono_font` from the Font row (`mono_font` walking JetBrains Mono, SF Mono, Menlo, Consolas, Cascadia Code, DejaVu Sans Mono, Noto Sans Mono, Liberation Mono, Courier New, then `monospace`), and set `[gui] transparency` to a percentage, clamped to 40, via the Window transparency slider to composite the whole window over the desktop.

Pin a custom accent color and follow the system light/dark appearance for the window theme.

toml

```
[gui]
theme = "auto"
custom_accent = "#e06c75"
```

Set window transparency so the transcript is composited over the desktop wallpaper at 25 percent.

toml

```
[gui]
transparency = 25
```

— Full config keys, defaults, and runnable examples v0.6.0

02 **Completion chimes for tasks and plans** NEW 85

Enable completion chimes by setting `[agent] completion_notification_enabled` to true; the `notification-sound` feature plays an embedded WAV when a task settles or a plan reaches 100 percent.

Opt into audio chimes so the agent plays a sound when an autonomous plan finishes, without watching the screen.

```toml
[agent]
completion_notification_enabled = true
```

— Exact config key with a runnable example v0.6.0

03 **Settings card UI in GUI** NEW 83

Open a settings card with `ctrl+comma` to configure approval policy, notifications, redaction, and credentials without leaving the window; credential keys write through `write_keys_key` and absent-key removal uses the new `remove_key` path on `Config`. Settings now organizes into a category rail with sections General, Notifications, Privacy, Advanced, Memory, and Credentials instead of a single column.

— Exact shortcut, config function names, and section list v0.6.0

04 **Desktop GUI packaging and distribution** NEW 78

v0.6.0 bundles a desktop GUI (`opencrabs-gui`) in every archive across all five targets; the running app owns the instance lock and keeps background work alive, and macOS gets a double-click bundle. The Linux archive now includes `opencrabs-gui.desktop`, installing the window app as `OpenCrabs Desktop` in the applications menu; the macOS archive now includes `OpenCrabs Desktop.app` with `CFBundleExecutable` set to `opencrabs-gui`, alongside the existing `OpenCrabs.app` terminal bundle. Windows archives now bundle the four MSVC runtime DLLs beside both executables so the app starts on a clean install with no Visual Studio or vc_redist required.

— Names exact bundle files and IDs, no sizes or timings v0.6.0

05 **Seeker universal search and session pinning** NEW 78

Open Seeker with `Ctrl+Shift+F` on both surfaces or `cmd-k` in the window to search conversations, slash commands, screens, saved skills, and documentation pages in one centered field. Pin a session with `Ctrl+Shift+P` in both the window and the terminal, freeing `Ctrl+Shift+F` for Seeker.

— Exact shortcuts, but no detail on index scope v0.6.0

06 **Theme config key migration** BREAKING 70

The `[gui] dark_theme` and `[gui] light_theme` config keys are now ignored; a GUI selection of a TUI preset resets to `auto`. Migrate to `[gui] theme` set to `orgn`, `light`, `custom`, or `auto`.

— Exact migration path, no deprecation timeline given v0.6.0

07 **Projects screen for session management** NEW 70

Open the Projects screen with `alt+p` or `/projects` in the window to create, delete, assign, and open project and session cards, including bulk assign, move, and clear across multiple sessions.

— Exact shortcut and actions, no data model detail v0.6.0

08 **Message interactions and transcript forwarding** NEW 65

Messages in the window now support Edit on hover, React from the crab pack, Reply with quoted rows, Forward staged into the destination composer, and Copy as Markdown. Forward transcript messages to another session or channel using press-and-hold or shift-click selection; each forwarded row carries a `source_session_id` provenance column.

— Names actions and provenance field, no exact UI path v0.6.0

THINNER COVERAGE BELOW

09 **Persisted plan completion state** NEW 55

Completed plans now write a `plan_completed` row keyed on session id and plan id, displaying a read-only checked card in both apps after completion and reload.

— Names schema but no command or UI path v0.6.0

└──▷ **BREAKING ON UPGRADE**

! The `[gui] dark_theme` and `[gui] light_theme` config keys are ignored; a GUI selection of a TUI preset resets to `auto`. Migrate to `[gui] theme` set to `orgn`, `light`, `custom`, or `auto`.

WAS THIS USEFUL?  

openclaw.net

Feedback ⓘ v0.5.0 NOTES ↗ RANK

Openclaw.net is a self-hosted AI + agent runtime in .NET (NativeAOT-friendly) for Enterprise and Personal use.

> `OpenClaw.NET` v0.5.0 adds user-delegated workflow authentication via Entra on-behalf-of or RFC 8693 token exchange, hardens channel delivery semantics and gateway access controls, tightens payment idempotency handling, and ships ready-to-run desktop bundles for Windows, macOS and Linux.

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

01 **Channel delivery outcome classification and Discord fixes** 75

BREAKING

Channel adapters now distinguish transport acceptance, suppression, rejection, partial delivery, cancellation, and unknown outcomes, with retries limited to known transient rejections; partial or uncertain sends are not replayed. Unsupported content sent via `IChannelAdapter.SendAsync` is now explicitly rejected, and plugin send handlers must confirm transport acceptance, with missing acknowledgments treated as failures. Discord now preserves message chunks and handles malformed rate-limit metadata safely.

— Interface method named; outcome states described but no API endpoint. v0.5.0

02 **Desktop bundles for Windows, macOS and Linux** NEW 75

Desktop bundles are now available for Windows x64 (`openclaw-desktop-win-x64.zip`), Apple Silicon macOS (`openclaw-desktop-osx-arm64.zip`), and Linux x64 (`openclaw-desktop-linux-x64.zip`), each including AgentQi Companion, the NativeAOT gateway, and the NativeAOT CLI, with matching `.sha256` checksum files.

— Exact downloadable filenames given, ready to use directly. v0.5.0

03 **User-delegated workflow authentication via Entra OBO/RFC 8693**

NEW 72

Authenticated HTTP integration and MCP workflow calls now support user-delegated identity via Entra on-behalf-of or RFC 8693 token exchange, with bounded token acquisition and caching and no silent fallback to shared credentials. Delegated workflows require explicit HTTPS/OIDC and backend configuration; cookies, bootstrap/account tokens, chat, and background jobs cannot supply a delegated user assertion.

— Mechanism and config requirements named, but no exact config key given. v0.5.0

04 WhatsApp send failures now surfaced, not suppressed

BREAKING

65

`WhatsApp.BridgeSuppressSendExceptions` no longer hides failed sends; failed sends are now surfaced through `IChannelAdapter.SendAsync`, so plugin send handlers that relied on suppressed failures will now see them as errors.

— Config key and interface named, enabling direct verification. v0.5.0

05 Gateway access restrictions and log redaction

BREAKING

65

Gateway now rejects filesystem-link escapes and redacts corrupt-memory payloads from logs, and restricts workspace and archive operations, media uploads and downloads, webhook delivery reservations, and file publication. Workspace administration and digital-employee installation now require an administrator role, media uploads require an operator or administrator role, and cookie-authenticated mutations require CSRF protection.

— Role requirements described without specific flags or endpoints. v0.5.0

THINNER COVERAGE BELOW

06 Payment idempotency record reconciliation requirements

NEW

50

Pending or unreadable payment idempotency records must not be deleted to force a retry — they represent potentially completed provider operations and must be reconciled against provider records; separate storage directories do not share duplicate protection.

— No named config key or command for idempotency handling. v0.5.0

└──▷ BREAKING ON UPGRADE

! `WhatsApp.BridgeSuppressSendExceptions` remains readable but no longer hides failed sends — plugin send handlers that relied on suppressed failures will now see them surfaced as errors through `IChannelAdapter.SendAsync`.

! Unsupported content sent via `IChannelAdapter.SendAsync` is now explicitly rejected; plugin send handlers must confirm transport acceptance, and missing acknowledgments are treated as failures.

! Delegated workflows require explicit HTTPS/OIDC and backend configuration; cookies, bootstrap/account tokens, chat, and background jobs cannot supply a delegated user assertion.

! Workspace administration and digital-employee installation now require an administrator role; media uploads require an operator or administrator role; cookie-authenticated mutations require CSRF protection.

! Pending or unreadable payment idempotency records must not be deleted to force a retry — they represent potentially completed provider operations and must be reconciled against provider records. Separate storage directories do not share duplicate protection.

WAS THIS USEFUL? 👍 👎

platypus

Feedback ⓘ v4.5.0 NOTES ↗ RANK

Platypus is a self-hosted AI Agents for your whole team — on your infrastructure, your models, around the clock.

> Platypus ships a round of UI refinements this release: read-only Chat during Trigger runs, structured Provider header editing, revealable secret inputs, and tidier admin tables.

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

01 **Revealable secret input fields** IMPROVED 45

Secret input fields now offer a reveal toggle via `FormTextField`, letting admins verify credentials without leaving the form.

— Names component but lacks exact UI path v4.5.0

02 **Structured Provider header editing** IMPROVED 35

Provider headers can now be edited as individual name-and-value rows in the UI instead of raw text.

— Names UI change but no exact path given v4.5.0

03 **Read-only Chat during Trigger runs** NEW 28

Platypus now marks a Trigger run's Chat as read-only, preventing accidental edits during automated runs.

— Describes behavior but no config or flag named v4.5.0

04 **Overflow menus for admin table actions** IMPROVED 28

Admin table row actions are now grouped into overflow menus (⋮) to reduce clutter on crowded tables.

— UI tweak described without further mechanism v4.5.0

WAS THIS USEFUL? 👍 👎

HeyGen HyperFrames

Feedback ⓘ v0.9.48 NOTES ↗ RANK

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

> HyperFrames shipped a Studio keyboard shortcuts panel and expanded step targets to accept any folder containing an index.html.

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

01 **Keyboard shortcuts panel in Studio** NEW 55

Pressing `?` while in the HyperFrames Studio UI opens a keyboard shortcuts panel listing every available shortcut.

Discover all available Studio keyboard shortcuts without leaving the app.

> 📍 Press ? while in the HyperFrames Studio UI to open the keyboard shortcuts panel.

— Exact key and UI path given, but no shortcut list v0.9.48

02 **Folder with index.html as step target** IMPROVED 45

HyperFrames now accepts a folder containing an `index.html` as a valid target for a step, independently of project discovery, rather than requiring a project root.

— Explains behavior change but no command example v0.9.47

WAS THIS USEFUL? 👍 👎

persona

Feedback ⓘ @runtypelabs/persona@4.39.0 NOTES ↗ RANK

Persona is a create agentic front-end experiences for the web in VanillaJS. Lightweight, extensible, and WebMCP-native.

> Persona's 4.39.0 release focuses entirely on script-tag install performance: a new critical-CSS inlining API cuts initial payload nearly in half, CSS failures now self-heal, and asset prefetching becomes smarter about timing and network conditions.

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

01 Critical CSS inlining via installCriticalStyles() NEW

90

Calling window.AgentWidgetLauncher.installCriticalStyles() inlines the launcher's critical CSS so it paints without waiting for the external stylesheet. Script-tag installs now carry the launcher's required CSS in the bundle itself, cutting the bytes needed before the launcher appears from 37.7 KiB to 20.0 KiB Brotli.

Trigger critical-style injection manually for installs that need the launcher to paint immediately without any external stylesheet fetch.

```javascript
window.AgentWidgetLauncher.installCriticalStyles();
```

— Names exact API and before/after byte numbers @runtypelabs/persona@4.39.0

02 Automatic retry on CSS load failure NEW

72

A stylesheet that fails to load on widget open is now reported via onError({ phase: "css" }) and retried automatically on the next click, instead of leaving the widget silently broken.

Handle a CSS load failure gracefully — catch the `onError` phase and show a fallback message rather than leaving the widget silently broken.

```javascript
AgentWidget.install({
  onError(err) {
    if (err.phase === "css") {
      console.warn("Widget stylesheet failed to load; will retry
on next open.");
    }
  }
});
```

— Names the exact error phase and retry trigger `@runtypelabs/persona@4.39.0`

03 **Smarter asset prewarming after launcher render** IMPROVED

66

Script-tag installs now pre-warm the stylesheet, main bundle, and lazy chunks after the launcher renders so the first open and first message run from cache, including in Safari. Inline, docked, and auto-expanded installs fetch the stylesheet, bundle, and chunks in parallel instead of sequentially, while visitors with `Save-Data` or a 2G-class connection are only warmed once they reach for the launcher to avoid unnecessary prefetch on slow or metered links.

— Describes mechanism and conditions but no direct API to invoke

`@runtypelabs/persona@4.39.0`

WAS THIS USEFUL? 👍 👎

pi-fabric

Feedback ⓘ v0.137.0 NOTES ↗ RANK

Pi-fabric is a programmable tool and agent runtime for Pi

> pi-fabric's v0.137.0 release introduces the `memory.outline` API for bounded session previews, ships a standalone `packages/pi-recall` package for TypeScript and Python, and improves CLI continuation and session expansion filtering.

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

01 **Bounded session previews via memory.outline** NEW 67

The new `memory.outline` API gives ordered, bounded session entry previews with source and lineage binding, supporting enforce-mode and read-only speculation the same way existing memory reads do.

– Names the API and its modes but no usage example v0.137.0

THINNER COVERAGE BELOW

02 **Standalone pi-recall memory skills package** NEW 55

A new engine-free `packages/pi-recall` source package adds standalone Fabric memory skills for TypeScript and Python. It is opt-in and tested on Linux and Windows.

– Names package and platforms but no setup steps v0.137.0

03 **CLI continuation preserves bindings and page budget** IMPROVED 45

CLI continuation commands now preserve canonical selection, source and lineage bindings, branch mode, role, and page budget across continuations.

– Describes state preserved but no command names v0.137.0

04 **Role filtering for session expansion** IMPROVED 40

Session expansion can now be filtered by exact role, and that filter persists across paginated reads.

– Describes behavior but no flag or endpoint named v0.137.0

WAS THIS USEFUL? 👍 👎

◆ MCP Servers & Tooling

claudexor

Feedback ⓘ v6.1.2 NOTES ↗ RANK

Claudexor is a multi-harness control plane for Claude Code, Codex, Cursor, and OpenCode with quota-aware rotation across multiple subscriptions.

> Claudexor v6.1.2 adds a stale-read mode for capability queries so callers aren't blocked on recompute, expands daemon status visibility, and ships a breaking data-root version floor that blocks downgrades.

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

01 **Stale-read mode for capability endpoints** IMPROVED 90

The `GET /v2/agent-capabilities` API endpoint and the MCP `claudexor_capabilities` tool now return their last cached composition, dated by `generatedAt`, for up to five minutes past the 15-second recompute window, so callers are never blocked waiting for a full capability sweep.

— Names endpoint, MCP tool, field and exact timing windows v6.1.2

02 **Data root version floor blocks downgrades** BREAKING 70

Older engines — including v6.1.1 — now refuse a data root that v6.1.2 has served, surfaced as `root_authority_floor_regression`; reverting requires the complete stopped backup procedure described under v6.1.0. The SSH-runtime manifest of v6.1.2 also requires Claudexor.app 6.1.2.

— Names error code and version requirements; procedure only referenced v6.1.2

THINNER COVERAGE BELOW

03 **Expanded worstSync daemon status field** IMPROVED 40

The daemon's `worstSync` status field now names six additional sections, surfacing which work still runs on the main thread.

— Field named but the six sections aren't enumerated v6.1.2

04 **Harness updates invalidate cached status** IMPROVED 40

A harness install or update through the daemon now drops cached status projections, so the next read — including Accounts harness statuses — reflects the new program immediately.

— Describes behavior change but no config/flag named v6.1.2

└──▷ BREAKING ON UPGRADE

! Older engines — including v6.1.1 — refuse a data root that v6.1.2 has served (`root_authority_floor_regression`); reverting requires the complete stopped backup procedure described under v6.1.0.

! The SSH-runtime manifest of v6.1.2 requires Claudexor.app 6.1.2.

WAS THIS USEFUL? 👍 👎

gflow-cli

Feedback ⓘ v0.99.0 NOTES ↗ RANK

Gflow-cli is a command-line tool for Google Flow: generate Veo videos and Imagen images via script or batch. Ships an MCP server for coding agent integration.

> gflow-cli's v0.99.0 release adds a dedicated output-file flag that replaces older output flags, and introduces an opt-in stable exit-code system for scripting.

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

01 **Output file flag replaces old output flags** BREAKING

70

Command output can now be routed to a file with `--out-file` , which replaces the deprecated `-o` / `--output PATH` flags.

— Names exact flags but no further mechanism detail v0.99.0

02 **Opt-in stable exit codes** NEW

65

Setting the environment variable `GFLOW_CLI_EXIT_CODES=2` opts into a stable, structured exit-code system.

— Names exact env var but lacks exit-code schema detail v0.99.0

WAS THIS USEFUL? 👍 👎

◆ Browser & Computer Use

Skyvern

Feedback ⓘ v1.0.58 NOTES ↗ RANK

Skyvern automates browser-based workflows with vision models and agents, completing form-driven tasks on sites it has not seen before.

> Skyvern's v1.0.58 release expands Gmail integration and Copilot controls, introduces Ember-based chat API paths, adds run-group limits and Nango connection cataloging, and retires legacy AES-CBC encryption and query-string credential authentication.

└──▷ **WHAT SHIPPED · 9 FEATURES** most completely described first ? what's the number?

01 **Query-string credential fallback removed** BREAKING 65

The query-string credential fallback on streaming sockets is removed; clients that passed credentials via query string must switch to header-based authentication.

— Clear before/after with explicit migration instruction v1.0.58

02 **Gmail integration for inbox and Copilot runs** IMPROVED 60

The Email Inbox editor now supports downloading attachments from connected Gmail accounts, and Copilot's Gmail card lets users start Connected Gmail directly when the org has no native Gmail account. The Gmail run capability has also been extended to scheduled, event, and child-workflow runs.

— Names Email Inbox editor, Gmail card, and run types but no config keys v1.0.58

03 **Copilot trigger, run-approval, and diagnostics additions** NEW 60

Create Copilot triggers are now switched off by default, with a card that opens them in the Triggers panel. Copilot can start one real run of a saved workflow behind an approval card, and it now shows the failed run's at-failure screenshot when reading or diagnosing a run.

— Names Triggers panel, approval card, and screenshot feature with UI paths v1.0.58

THINNER COVERAGE BELOW

04 **Ember chat API paths** NEW 55

Skyvern now serves the chat API under Ember paths, headers, and chat-id keys alongside the old ones, enabling clients to migrate to the Ember-based scheme.

— Names Ember paths/headers/keys but gives no exact route strings v1.0.58

05 **AES-CBC encryption retired** DEPRECATED 50

The AES-CBC encrypt path is retired; only decryption of legacy AES-CBC data is still supported, so new data can no longer be encrypted with AES-CBC.

— States scope of deprecation but no migration path given v1.0.58

06 **Per-org run group limits** NEW 45

Run groups now enforce a per-org item limit and bounded concurrent dispatch, capping how many items an organization can execute simultaneously within a run group.

— Describes mechanism but no concrete limit numbers given v1.0.58

07 **Nango connections in account catalog** NEW 40

Each user's Nango connections are now listed as catalog accounts without hiding other credential sources.

— Names Nango integration but no further mechanism detail v1.0.58

08 **File upload allowlist before denylist** NEW 30

Uploads are now gated on an allowlist of expected file types checked before the existing denylist, tightening control over which files can be uploaded.

— Behavior named but no flag or config surface given v1.0.58

09 **Browser session infrastructure selection** NEW 30

A browser session can now choose its infrastructure based on its target site or bot-protection profile.

— Brief description with no mechanism or configuration detail v1.0.58

└──▷ **BREAKING ON UPGRADE**

! The query-string credential fallback on streaming sockets is removed; clients that passed credentials via query string must switch to header-based authentication.

! The AES-CBC encrypt path is retired; only decryption of legacy AES-CBC data is still supported.

WAS THIS USEFUL?  

◆ Agent Sandboxes

NoLabs nono

Feedback ⓘ v0.82.0 NOTES ↗ CODE ↗ RANK

The nono runtime runs AI agents in isolated environments with policy-controlled filesystem, network, and credential access.

> nono's v0.82.0 release adds platform authentication and session recording commands, AWS operation-level endpoint policy matching, Unix-socket webhook approvals, and live audit logging, while replacing the `connect` command with `nono remote`.

└──▷ WHAT SHIPPED · 8 FEATURES most completely described first ? what's the number?

01 AWS operation-level policy and aws_auth routing IMPROVED

95

Adds an `operations` field to `EndpointRuleConfig` in endpoint policy to match specific AWS operation-name globs on `aws_auth` routes, letting a credential grant restrict access to specific API calls (e.g. S3 `GetObject` / `PutObject`). The `aws_auth` proxy routes also now derive the signing region automatically from the request host for wildcard upstreams, select the signing profile by placeholder access key ID, and accept AWS GovCloud regions and S3 bucket-subdomain wildcards.

Restrict an AWS credential grant so only S3 `GetObject` and `PutObject` operations are permitted, rejecting all other AWS API calls through the proxy.

```json
{
  "endpoint_policy": [
    {
      "url": "https://s3.amazonaws.com/*",
      "operations": ["GetObject", "PutObject"]
    }
  ]
}
```

— Names config field, route type, and a runnable policy example v0.82.0

02 Platform authentication and enrollment commands NEW

70

Adds `nono login` and `nono logout` for OAuth 2.0 device-flow authentication and platform enrollment, and a unified `nono platform` command that now covers sign-in, tenant management, and device enrollment.

Authenticate to the nono platform from a new machine using device-flow OAuth before starting secured agent sessions.

```
$ nono login
```

— Names exact commands and auth flow but not full tenant-management scope v0.82.0

03 **Session terminal recording and replay** NEW 70

Adds `nono replay` to play back a session's terminal recording and `nono transcript` to print it as text, implementing NEP-0010 session terminal capture.

Replay a past agent session's terminal output to review exactly what commands ran and what was printed.

```
$ nono replay
```

— Commands named with runnable example, mechanism spec referenced v0.82.0

04 **Unix-socket webhook approval transport** NEW 65

Adds a Unix-socket webhook transport for approval backends in `nono proxy`, enabling approval requests over `AF_UNIX` pathname sockets, and supports approval elevation for `AF_UNIX` pathname socket connect and bind operations.

— Mechanism and socket type named but no config/CLI example given v0.82.0

05 **`nono remote` replaces `connect` for session management** 60

BREAKING

Adds `nono remote` for launching, listing, and connecting to remote sessions. The former `connect` subcommand is removed, with its functionality folded into `nono remote`; scripts or aliases calling `nono connect` must be updated.

— Clear migration note but no example command shown v0.82.0

THINNER COVERAGE BELOW

06 **Expanded plain-HTTP network approvals** IMPROVED 40

Extends network approvals in `nono proxy` to cover plain-HTTP forward requests and `allow_domain` endpoints.

— Names the config key but no example or scope detail v0.82.0

07 **Live proxy audit log writes** IMPROVED 35

Proxy audit events are now written to the audit log live as they occur, rather than only at session finalization.

— Explains before/after behavior but no config surface named `v0.82.0`

---

08 **Phantom token format for custom credentials** NEW 35

Adds a `phantom_format` field to custom credentials configuration for phantom token format control.

— Names the config field but gives no further mechanism `v0.82.0`

└──▷ **BREAKING ON UPGRADE**

! The `connect` subcommand is removed; its functionality is now part of `nono remote`. Update any scripts or aliases that call `nono connect`.

WAS THIS USEFUL? 👍 👎

◆ Agent Memory

Volcengine OpenViking

Feedback ⓘ react@0.1.12 NOTES ↗

RANK

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

> OpenViking's react package adds analytics and frontend monitoring through OpenVikingProvider, exports a new ContextPicker component, supports resource uploads in chat, and expands FileBrowser navigation and search, while removing SSR/hydration support and shipping several breaking changes to HITL and FileTree APIs.

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

01 Analytics and monitoring via OpenVikingProvider NEW

90

Adds optional, strongly typed analytics and frontend monitoring through `OpenVikingProvider`, disabled by default, covering interactions, outcomes, viewport exposure, component and request failures, resources, streams and tasks, with instance identification and bounded traces. It reuses TanStack Query v5 clients through `OpenVikingProvider` with per-endpoint concurrency and identity-scoped directory and search caches supplied by the default SDK runtime.

Enable analytics and frontend monitoring in your app so that interaction outcomes and request failures are captured without any setup overhead.

jsx

```jsx
<OpenVikingProvider analytics={analyticsInstance} monitoring=
{monitoringInstance}>
  <YourApp />
</OpenVikingProvider>
```

— Full provider usage shown with concrete monitoring scope react@0.1.12

02 FileBrowser navigation and search improvements IMPROVED

85

`FileBrowser` gains preview-only navigation via the `reveal: false` option, root navigation with automatic expand and refresh of the target directory, advanced search supporting directory scopes and date ranges, and built-in directory interactions supporting ancestor expansion, pagination, child-directory creation, host-provided permissions and validation, with custom answers able to delegate selection to the host.

— Names exact `reveal: false` option and interaction scope react@0.1.12

03 **Standalone VikingBotInteraction stylesheet** NEW 70

Import the `VikingBotInteraction` standalone stylesheet from `@openviking/react/components/viking-bot/interaction/style.css` without requiring Ark Design.

Bring in the VikingBotInteraction stylesheet independently when you are not using Ark Design.

```javascript
import '@openviking/react/components/viking-bot/interaction/style.css';
```

— Exact importable path given, runnable immediately `react@0.1.12`

---

04 **FileTree label and trailing content changes** BREAKING 70

The `FileTreeLabels.readOnly`, `noAccess` and `traverseOnly` overrides have been removed, and `FileTree` trailing content now defaults to `rowTrailingPosition: 'viewport-end'`; set `'row-end'` to retain the previous horizontal-scrolling behavior.

— Names removed overrides and exact migration flag `react@0.1.12`

---

05 **HITL text requests require non-blank input** BREAKING 65

HITL text questions now require non-blank input by default; set `required: false` to allow optional text.

— Exact flag given with clear migration action `react@0.1.12`

THINNER COVERAGE BELOW

06 **Exported ContextPicker component** NEW 55

Export `ContextPicker` for controlled file and directory selection, search, single or independent multiple selection, asynchronous selection restoration and custom footer actions.

— Component named but no usage example given `react@0.1.12`

---

07 **Resource uploads in chat** NEW 55

Support resource uploads in chat when the Console service provides attachment uploads, waiting for completion before sending and cancelling uploads when chat identity, explicit session or request lifetime changes.

— Behavior described but no API surface named `react@0.1.12`

08 **Multiline HITL text requests** IMPROVED 55

HITL text requests now honor single-line and multiline modes via the `multiline` configuration.

— Names exact config key, limited mechanism detail `react@0.1.12`

09 **Multi-step HITL answer shape change** BREAKING 50

Multi-step HITL answers now use `steps` as the answer shape; single-question answer shapes are unchanged.

— Names new shape key but little mechanism detail `react@0.1.12`

10 **SSR and hydration support removed** BREAKING 35

SSR and hydration paths are no longer supported; components must be mounted in the browser.

— States impact but offers no migration path `react@0.1.12`

└──▷ **BREAKING ON UPGRADE**

! SSR and hydration paths are no longer supported; components must be mounted in the browser.

! HITL text questions now require non-blank input by default; set `required: false` to allow optional text.

! Multi-step HITL answers now use `steps` as the answer shape; single-question answer shapes are unchanged.

! The `FileTreeLabels.readOnly`, `noAccess` and `traverseOnly` overrides have been removed.

! `FileTree` trailing content now defaults to `rowTrailingPosition: 'viewport-end'`; set `'row-end'` to retain the previous horizontal-scrolling behavior.

WAS THIS USEFUL? 👍 👎

DEPLOY

◆ Inference Providers & GPU Clouds

Crusoe Cloud

Feedback ⓘ changelog · Nov 15, 2026 CHANGELOG ↗

Crusoe Cloud provides GPU infrastructure and managed inference for AI training and deployment.

Crusoe Cloud adds a `crusoe_load_balancer` Terraform resource, Kubernetes node pool autoscaling, CMK v2 clusters with audit logs and auto-repair, and in-place InfiniBand partition moves.

Manage external load balancers via the new `crusoe_load_balancer` Terraform resource, configuring `vpc_id`, `protocol` (`tcp` or `udp`), `listen_ports_and_backends` (each with `listen_port` and `backends` of `ip` and `port`), optional `health_check_options`, and reading the assigned `vip`.

Autoscale Kubernetes node pools by adding an `autoscaling_config` block (`enabled`, `min_node_size`, `max_node_size`) to `crusoe_kubernetes_node_pool`, letting the cluster autoscaler manage node count within those bounds on clusters with a supported autoscaler version.

Rotate pull-through cache credentials in place via `crusoe_registry_repository` using the new `credentials_version` field and write-only `upstream_registry_credentials.username` / `.password` (requires Terraform 1.11 or later; credentials are never stored in plan or state).

Move a VM to another InfiniBand partition in place by changing `transport_partition_id` or `ib_partition_id` in `host_channel_adapters` on `crusoe_compute_instance`; the VM is stopped and restarted, and plans now reject mismatched `ib_partition_id` / `transport_partition_id` values.

Manage ICMP firewall rules natively via `crusoe_vpc_firewall_rule` with optional port fields — dropping the `source_ports = ""` workaround — and switch an existing rule to `icmp` in place; plans now reject ports on ICMP rules and port values outside 1–65535.

+2 more

NEW SUPPORTED PLATFORM Google Workspace SSO

NEW INTEGRATION SCIM provisioning from Okta

└──▷ BREAKING ON UPGRADE

! Existing `crusoe_registry_repository` state is upgraded and the stored `username` and `password` are dropped from it.

! A `crusoe_kubernetes_node_pool` with autoscaling enabled outside Terraform and no `autoscaling_config` block in configuration will have autoscaling disabled on

the next `apply`.

! Changing `upstream_registry.provider`, `upstream_registry.url`, or `project_id` on `crusoe_registry_repository` now replaces the repository instead of failing.

WAS THIS USEFUL? 👍 👎

Anyscale

Feedback ⓘ changelog · Nov 15, 2026 CHANGELOG ↗ RANK

Anyscale provides a managed Ray platform for distributed AI development, training, batch processing, and model serving.

> Anyscale shipped several breaking changes around multi-cloud resource scoping — system cluster termination now defaults to every cloud resource and traffic weights are capped at 100 — alongside improvements to multi-cloud artifact uploads and automatic GPU metrics setup.

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

01 Cloud-wide system cluster termination

BREAKING 90

Terminating a system cluster without naming a cloud resource now terminates the system cluster of every cloud resource in the cloud, starting with the primary, instead of only the primary cloud resource's system cluster as before. Scope the action to one cloud resource with `--resource` / `--resource-id` (CLI) or `cloud_resource_id` (SDK), and wait for completion with `--wait`.

Terminate system clusters across all cloud resources in a cloud, waiting for completion, while scoping the action to a single resource to avoid broad impact.

```
$ anyscale cloud terminate-system-cluster --resource my-cloud-resource --wait
```

— Full before/after behaviour with exact flags and a runnable command.

changelog-20261115-91f78643

02 StaticTrafficPolicy weight range enforcement

BREAKING 80

The CLI and SDK now enforce a weight range of 1–100 for each weight in a `StaticTrafficPolicy`; the CLI fails with "The weight of cloud resource `<name>` must be from 1 to 100" and the SDK raises a `ValueError` with the same message for weights above 100, whereas previously any weight of 1 or more was accepted.

Cap traffic weights at 100 when splitting traffic across cloud resources in a multi-cloud service to avoid SDK validation errors.

```python
from anyscale.service import StaticTrafficPolicy

policy = StaticTrafficPolicy(
    weights={
        "cloud-resource-us": 70,
        "cloud-resource-eu": 30,
    }
)
```

— Exact error message and SDK class named, with sample code. changelog-20261115-91f78643

THINNER COVERAGE BELOW

03 Multi-cloud working_dir and py_modules upload IMPROVED

50

Upload a local `working_dir` or `py_modules` directory to the object storage of every cloud resource in a service's compute config, so each cluster downloads from its own cloud resource rather than only the primary.

— Describes mechanism and scope but no config key or command given.

changelog-20261115-91f78643

└──▷ **BREAKING ON UPGRADE**

! The CLI and SDK now reject `StaticTrafficPolicy` weights above 100; previously any weight of 1 or more was accepted. Existing configs or SDK calls with weights above 100 will fail with 'The weight of cloud resource `<name>` must be from 1 to 100'.

! Terminating a system cluster without naming a cloud resource now terminates the system cluster of every cloud resource in the cloud, starting with the primary. Previously only the primary cloud resource's system cluster was terminated. Use `--resource` or `--resource-id` (CLI) or `cloud_resource_id` (SDK) to scope termination to one cloud resource.

! On Kubernetes-stack workspaces, `anyscale.workspace.run_command` now starts in the workspace project directory (e.g. `/home/ray/default`) instead of `/home/ray`. Commands that relied on starting in `/home/ray` must be updated to use an absolute path or an explicit `cd`.

WAS THIS USEFUL? 👍 👎

Eigen Labs Darkbloom

Feedback ⓘ v0.10.13 NOTES ↗

CODE ↗ RANK

Darkbloom is a private Inference Network on Idle Macs

> Darkbloom's v0.10.13 release adds fine-grained control over loop-exit thinking and tool rendering across more model families, hardens two-Mac pair security with Keychain-backed link keys and faster dead-link detection, and tightens JSON/tool-call behaviour for GPT-OSS, GLM, and Qwen models.

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

01 Per-model loop-exit thinking control NEW

90

The `DARKBLOOM_LOOP_EXIT_THINKING` environment variable controls loop-exit thinking per model — set to `0` to disable for Gemma 4, or to a comma-separated model list such as `gemma4,gemma4_text,qwen3_5_moe` to enable for Qwen 3.5/3.6 35B (adding `qwen3_5` covers the 9B and Qwen 3.8 27B); on by default for Gemma 4.

Enable loop-exit thinking for Qwen 3.5/3.6 35B models so runaway reasoning loops end in an answer rather than a `length` stop.

```
$ DARKBLOOM_LOOP_EXIT_THINKING=gemma4,gemma4_text,qwen3_5_moe
darkbloom start
```

— Exact env var, accepted values, and default documented with a runnable command.

v0.10.13

02 GLM-5.3-Flash schema-miss passthrough NEW

86

Setting `DARKBLOOM_GLM_JSON_SCHEMA_MISS=release` exposes GLM-5.3-Flash JSON answers that miss the requested schema to the caller's validator instead of refusing with a 422.

Let GLM-5.3-Flash pass schema-mismatching JSON answers through to your own validator instead of getting a 422 from the provider.

```
$ DARKBLOOM_GLM_JSON_SCHEMA_MISS=release darkbloom start
```

— Exact env var, value, and before/after behaviour with a runnable example. v0.10.13

03 Transformers-style tool rendering request flag NEW

83

Sending `"_darkbloom_template_json": "transformers"` in a chat request requests Hugging Face Transformers-style tool, call, and result rendering for Qwen 3.5/3.6/3.8, Bonsai 2, MiMo, GLM-5.3, and GPT-OSS.

Use Transformers-style tool rendering in a chat request so Qwen or GLM-5.3 sees tools exactly as Hugging Face would format them.

```
$ curl https://api.darkbloom.dev/v1/chat/completions \
  -H 'Content-Type: application/json' \
  -d '{"model": "qwen3-5-35b", "_darkbloom_template_json":
"transformers", "messages": [{"role": "user", "content": "Call
the search tool"}], "tools": [{"type": "function", "function":
{"name": "search", "parameters": {"type": "object", "properties":
{"q": {"type": "string"}}}}}]}'
```

— Named flag, exact model list, and runnable curl example. v0.10.13

04 **GPT-OSS tool_choice required and named functions** NEW

81

GPT-OSS now supports `tool_choice` `required` and named functions, routing the request to a declared function and returning `tool_calls`; `DARKBLOOM_GPTOSS_FORCED_TOOLS=off` disables GPT-OSS forced tool-choice support.

— Named env var and tool_choice values, but no runnable example given. v0.10.13

05 **Two-Mac pair security and reconnect hardening** BREAKING

78

Two-Mac pairs now store their link key in the Mac Keychain instead of a file, verify each Mac's identity on the link, and check the processes they start; the coordinator's dead-link detection window is now 6–8 seconds (down from up to 40 s), with an immediate reconnect after the link is declared dead. Both Macs of a pair must be on 0.10.13 to pair — a Mac on 0.10.12 cannot pair with one on 0.10.13.

— Mechanism and before/after numbers given, but no config key or command. v0.10.13

THINNER COVERAGE BELOW

06 **Reasoning/thinking field precedence fix** IMPROVED

52

Earlier thoughts sent back as `reasoning` or `thinking` fields now reach the model; the first non-null of `reasoning_content`, `thinking`, and `reasoning` wins.

— Explains field precedence order but offers no example. v0.10.13

07 **`finish_reason: "length"` on max_tokens truncation** IMPROVED

50

Answers that reach `max_tokens` now end with `finish_reason: "length"` for JSON answers and tool calls, including mid-stream cuts.

— Behaviour change described but no example or affected models. v0.10.13

---

08 **Qwen JSON schema support for Pydantic, OpenAI, zod** NEW

45

Qwen JSON answers now follow schemas written with Pydantic, OpenAI, or zod.

— Names three schema libraries but gives no mechanism or example. v0.10.13

---

09 **`parallel_tool_calls: false` support across models** NEW

40

Darkbloom now supports `parallel_tool_calls: false` on all models, returning the first tool call instead of failing the response.

— Thin description with no example, limit, or model list. v0.10.13

---

10 **Loop guard extended to image and video answers** IMPROVED

35

The loop guard now covers image and video answers, stopping verbatim-repeating media responses early.

— Describes scope extension without thresholds or examples. v0.10.13

└──▷ **BREAKING ON UPGRADE**

! Both Macs of a two-Mac pair must be on 0.10.13 to pair — a Mac on 0.10.12 cannot pair with one on 0.10.13.

WAS THIS USEFUL? 👍 👎

CoreWeave

Feedback ⓘ announced · Nov 14, 2026 BLOG ↗ RANK

CoreWeave is an AI cloud platform offering GPU compute, managed Kubernetes and storage for training and inference.

> CoreWeave shipped Physical AI Launchables, a one-command template catalog for robotics and autonomous-vehicle simulation, and CoreWeave Database, a managed PostgreSQL service running inside CKS clusters with pgvector and built-in HA/backups.

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

01 CoreWeave Database managed PostgreSQL service NEW

83

CoreWeave Database is a new managed PostgreSQL service that runs on CPU nodes inside a CoreWeave Kubernetes Service (CKS) cluster, connecting to applications via an in-cluster Kubernetes Service with no cross-cloud egress. It includes the `pgvector` extension for embedding queries from RAG services and AI agents, automatic failover across a primary and replicas on different hosts, and continuous point-in-time recovery by default with scheduled backups, on-demand backups before schema changes or upgrades, and full-database restore to a new cluster for validation. Live data is stored on CoreWeave Distributed File Storage (DFS), encrypted at rest and in transit with CoreWeave-managed TLS certificates, and backups stream to CoreWeave-managed storage without user-held credentials; access control uses native PostgreSQL roles combined with Kubernetes network policies, with audit logging of database activity. It is built on the open-source CloudNativePG (CNPG) Kubernetes operator and can be managed via the CoreWeave Cloud Console or `kubectl` and GitOps workflows.

— Mechanism and management surfaces detailed but no concrete command shown

launch-20261111-0bac9733

02 Physical AI Launchables template catalog NEW

80

A catalog of one-command Templates lets teams run robotics and AV simulation workloads by passing `name=value` pairs on the command line to adjust scenarios, duration, parallelism, or which model is under test. Templates include NVIDIA Isaac Sim's `Franka-pick-place`, `Scene-sdg`, `Vehicle-dynamics`, and `Warehouse-actors`; CARLA's `Scenario-matrix`; MuJoCo's `Playground-rl-sweep` and `Kitchen-pick-place`; NVIDIA AlpaSim's `Closed-loop-driving`; and NVIDIA Cosmos 3's `Scenario-generation`, plus NVIDIA Cosmos Data Factory agentic skills for generative-AI data augmentation (auto-labeling, data curation, retrieval) across video, image, and text. Every tool in the catalog also ships a `Blank` Template — a tested environment with the workload left open. Coding agents (Claude Code, Codex, and others) can launch, monitor, and retrieve results from any Template through a documented, machine-readable interface, and

integrations with Cadence, Convergent Science, Encord, nTop, PhysicsX, and Voxel51 let those tools run on CoreWeave compute while keeping their own interfaces.

— Names exact template commands and partners but lacks a full runnable example

launch-20261113-68650a19

NEW INTEGRATION SCIM identity provider integrations for AUP (Okta, Microsoft Entra)

announced · Nov 14, 2026 BLOG ↗

CoreWeave adds AUP and SUP to automate end-to-end identity sync from enterprise IdPs into Slurm-on-Kubernetes clusters via SCIM and NSSCache.

Automated User Provisioning (AUP) uses SCIM to continuously sync users and groups from enterprise identity providers into CoreWeave IAM, covering CKS, AI Object Storage, and the CoreWeave Console.

SUNK User Provisioning (SUP) automatically creates POSIX users and groups, SSH keys, and Slurm users and accounts on each cluster via a per-cluster SCIM endpoint and NSSCache, with access correct on first login.

Researchers can self-serve SSH key registration and view their POSIX UID and GID in the CoreWeave user settings page.

Teams without an identity provider can create IAM groups such as `slurm-users` and `slurm-sudo` and point NSSCache to the SCIM endpoint to provision Slurm access manually.

WAS THIS USEFUL? 👍 👎

◆ ML Platforms & SDKs

AWS Amazon Bedrock AgentCore

Feedback ⓘ

changelog · Nov 15, 2026 CHANGELOG ↗ RANK

Amazon Bedrock AgentCore provides managed agent runtimes, memory, identity, gateways, observability, and evaluation services.

> Amazon Bedrock AgentCore added a gateway-free web search API, new cold-start and resource controls for V2 agent runtimes, and Harness upgrades covering IAM-authenticated MCP connections and Strands-based managed defaults.

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

01 **Lifecycle hooks for V2 agent runtime warmup** NEW 95

Set lifecycle hooks in `hooksConfiguration` on `CreateAgentRuntime` or `UpdateAgentRuntime` for V2 agents; the build warmup hook issues a `warm` call to preload session-wide state and a `disk` call to pre-read files, reducing cold-start times for large binaries, heavy Python packages, or headless browsers.

— Names config key, APIs, hook calls, and the cold-start mechanism

changelog-20261115-9a7ef845

02 **Resource configuration for V2 agent runtimes** IMPROVED 90

Set a committed resource configuration with `minimumMemoryInMiB` on V2 agents to unlock container images up to 10 GB (compressed), up from 2 GB, for packaging models, large dependencies, and data. Resource configurations on V2 agent runtimes can also reserve a baseline from 0.25 vCPU / 1 GB memory up to 2 vCPU / 8 GB memory, with burst capacity up to 8 vCPU and 32 GB per session.

— Names config key and exact memory/vCPU limits and defaults changelog-20261115-9a7ef845

03 **Strands Harness managed defaults** IMPROVED 85

Harness now uses Strands Harness for managed defaults, providing the current recommended model (Claude Opus 5 in commercial regions), task-oriented instructions, conversation summarization, automatic prompt caching, and built-in tools for planning, web retrieval, filesystem access, context offloading, and subagent delegation; explicit model, system prompt, truncation, and `allowedTools` settings take precedence.

— Lists concrete defaults and override fields, no exact config path

changelog-20261115-9a7ef845

04 **InvokeWebSearch API for agent web search** NEW 60

The new `InvokeWebSearch` API, callable via AWS SDKs or the AWS CLI, lets agent loops or data pipelines search the web directly without needing a gateway.

— Names the API and call paths but no request signature changelog-20261115-9a7ef845

WAS THIS USEFUL? 👍 👎

◆ AI / LLM Gateways

BerriAI LiteLLM

Feedback ⓘ v1.108.9 NOTES ↗ RANK

LiteLLM is an AI gateway that calls over 100 model APIs in OpenAI format with cost tracking, guardrails, load balancing, and logging.

> LiteLLM's v1.108.9 LTS backport adds caller identity attribution for Vertex AI requests and support for Vercel AI Gateway decision models.

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

01 **Caller identity labels on Vertex AI requests** NEW 35

Vertex AI requests now carry caller identity labels, letting you attribute upstream request origin in Vertex AI audit logs and billing.

— Describes benefit but no config key or flag given v1.108.9

02 **Vercel AI Gateway decision models support** NEW 25

LiteLLM now supports Vercel AI Gateway decision models, expanding the set of routing backends available through the gateway.

— No usage example or config details provided v1.108.9

WAS THIS USEFUL? 👍 👎

◆ Inference Serving & Optimization

vLLM

Feedback ⓘ announced · Nov 12, 2026 BLOG ↗

The vLLM inference engine runs large language models with high-throughput, memory-efficient GPU execution.

vLLM adds day-0 support for Inkling 2 Small with DSpark speculative decoding, sequence parallelism for prefill, mixed-precision MoE, a new Gluon decode kernel, and LoRA.

NEW SUPPORTED PLATFORM Serve Inkling 2 Small with native DSpark speculative decoding by passing `--speculative-config '{"method":"dspark","num_speculative_tokens":7,"draft_sample_method":"probabilistic"}'` to `vllm serve`, drafting 7 tokens per step via an RNN head.

Enable Inkling 2 Small's reasoning and tool-call parsing with `--reasoning-parser inkling`, `--tool-call-parser inkling`, and `--enable-auto-tool-choice` flags on `vllm serve`.

vLLM adds sequence parallelism for prefill on Inkling 2 Small, cutting an 8K-token prefill step by 22% at TP4 and 26% at TP8, and freeing 1.3–1.9 GiB per GPU for KV cache.

vLLM adds a warp-specialized Gluon relative attention kernel for decode on Inkling 2 Small, dropping a batch-1 decode step at 100K context from 15.1 ms to 4.8 ms.

vLLM now supports LoRA fine-tuning for Inkling 2 Small.

NEW SUPPORTED PLATFORM MiniMax H3-Image on NVIDIA Vera Rubin (GB300/GB200)

NEW SUPPORTED PLATFORM MiniMax H3-Image on AMD Instinct MI325X and MI350X

NEW SUPPORTED PLATFORM MiniMax H3-Image on Intel Arc Pro B70 XPU

NEW SUPPORTED PLATFORM MiniMax H3-Image on Huawei Ascend NPUs

NEW SUPPORTED PLATFORM MiniMax H3-Image on Moore Threads MTT S5000 via MUSA

NEW INTEGRATION MiniMax H3-Image model support in vLLM-Omni

▾ 1 more release in this issue · 2026-11-12

> announced · Nov 12, 2026 BLOG ↗
>
> vLLM-Omni adds MiniMax H3-Image serving with MXFP8 DiT quantization, Ulysses SP4, parallel VAE tile decoding, AdaLN caching, and layerwise CPU offload for consumer GPUs.
>
> vLLM-Omni now supports MiniMax H3-Image, a 33.1B text-to-image and image-editing model that handles both tasks in one checkpoint with up to 16 reference

images at native 4K resolution.

vLLM now connects H3-Image DiT linears to online `MXFP8` loading and kernel selection on NVIDIA Rubin, using FlashInfer CuTe-DSL MXFP8 GEMM for eligible DiT linears and `TRTLLM_ATTN` for attention.

vLLM now distributes VAE image-tile decoding across GPUs and combines CUDA Graph replay with multiple CUDA streams per GPU, reducing VAE decode time from 2.635 s to 0.213 s on four GB300 GPUs.

vLLM now supports layerwise CPU offload for H3-Image DiT and text-encoder weights, enabling single-GPU deployment on RTX 5090, RTX 5080, RTX 6000D and RTX PRO 6000.

vLLM now supports `AdaLN` caching for H3-Image, reusing exact conditioning projections when inputs and weights match across the six denoising steps.

+1 more

WAS THIS USEFUL? 👍 👎

DATA

◆ Document Parsing

kordoc

Feedback ⓘ v4.24.80 NOTES ↗ RANK

Kordoc is a 한국 문서를 읽고·고치고·새로 쓰는 AI 문서 도구 — HWP·HWPX·PDF·오피스·스캔(OCR)을 Markdown으로, 신구대조·서식 보존 편집·Markdown→HWPX 작성·HWP→HWPX 변환·양식 채우기·공문서 검수까지. 라이브러리·CLI·MCP | Read, diff, edit, write and convert Korean documents (HWP, HWPX, PDF, Office, scans) — library, CLI and MCP server

> Kordoc now generates legacy HWP 5.0 binary files directly from Markdown, not just the modern HWPX format.

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

01 HWP 5.0 binary output from generated documents NEW

90

Generated documents can be saved as HWP 5.0 binary files via the CLI with `kordoc generate 보고서.md --preset 보고서 -o 보고서.hwp`, by setting `output_path` to a `.hwp` extension in the MCP `generate_document` tool, or in the library by calling hwpxToHwp(await markdownToHwpx(md)). This requires the optional dependency `@rhwp/core` (version 0.9.0 or later, MIT licensed) installed separately via `npm install @rhwp/core`.

Generate a Korean government report from Markdown and save it directly as an HWP 5.0 binary for recipients still on legacy HWP software.

```
$ npm install @rhwp/core
npx kordoc generate 보고서.md --preset 보고서 -o 보고서.hwp
```

— Names exact CLI command, MCP field, library call and dependency v4.24.80

WAS THIS USEFUL? 👍 👎

◆ Vector DBs & RAG

liliang-cn CortexDB

Feedback ⓘ v2.152.0 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 v2.152.0 release expands its query and analysis surface: richer aggregate queries, new Cypher functions, read-only SQL over imported tables, and a new analytics page in the live view, alongside friendlier Python client output.

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

01 **Cypher date and percentile functions** NEW 67

CortexDB's Cypher query support now includes `stDev` and percentile functions, date and datetime truncation to buckets, and the `created_at` field.

— Names exact functions and fields usable directly in Cypher queries. v2.152.0

02 **Read-only SQL over imported tables** NEW 60

Imported tables now stay queryable after import, with read-only SQL available over them through `importflow`.

— Names the `importflow` surface though no exact query syntax shown. v2.152.0

THINNER COVERAGE BELOW

03 **Aggregate API multi-measure queries** IMPROVED 57

CortexDB's aggregate API now accepts several measures, time buckets, operator filters, and percentiles in a single request.

— Describes request shape but no endpoint path or example call. v2.152.0

04 **Python client tool-result formats** NEW 46

CortexDB's Python client now returns tool results as records, CSV, and DataFrames for direct analysis of brain output.

— Names output formats but no method signature or example given. v2.152.0

05 **Analysis page in live view** NEW 40

The live view now includes an Analysis page that charts the brain's totals and trends.

— UI path is clear but mechanism and metrics detail are thin. v2.152.0

WAS THIS USEFUL? 👍 👎

Crow Desktop

Feedback ⓘ v1.0.1 NOTES ↗ RANK

Crow Desktop is a crow: build your second brain from the documents on your computer. Downloads for macOS, Windows and Linux.

> Crow Desktop's v1.0.1 release adds Microsoft Store packaging, Italian localization, streaming answers, a Sources panel that links responses back to source notes, and faster, non-blocking local-model responses while documents are being filed.

└──▷ **WHAT SHIPPED · 8 FEATURES** most completely described first ? what's the number?

01 **Microsoft Store packaging via build_msix.py** NEW 95

Crow now packages for the Microsoft Store via `scripts/build_msix.py`, bundling the local engine (CPU and Vulkan, no CUDA), icons at every display scale, with test packages signed per feature branch.

Build a signed MSIX package for Microsoft Store submission from a release branch.

```
$ python scripts/build_msix.py
```

— Names exact build script, platforms, signing and engine details v1.0.1

02 **Italian localization added** NEW 70

Crow now speaks English and Italian; switch languages in Settings → Language, and the first run follows the system language, with numbers, sizes, and dates following the locale too.

Switch Crow's interface to Italian so locale-aware number and date formatting also updates.

📍 In Crow, go to Settings → Language and select Italiano.

— Exact settings path and locale behaviour given v1.0.1

03 **Responsive local models during filing** IMPROVED 70

With local models, questions asked while documents are being filed are now answered immediately, interrupting and resuming the filing call rather than queuing behind it, and models now load as soon as Crow opens, with a status pill above the chat while loading, so the first document or question is not delayed.

— Explains concurrency mechanism and load-timing behaviour clearly v1.0.1

THINNER COVERAGE BELOW

04 **Sources panel in answers** NEW 55

An answers panel now lists Sources — the notes each answer was drawn from, ordered by citation — and highlights them in the brain while the list is open.

— Describes behaviour but no concrete UI path or command v1.0.1

05 **New default OpenRouter model** IMPROVED 55

With OpenRouter, the default model is now DeepSeek V4.1 Flash; Gemini 3.8 Flash and Claude Sonnet 5.5 remain available.

— Names specific models but no action to take v1.0.1

06 **Source file path and Explorer link in notes** NEW 50

Open notes now show the source file path under their folder, with an Open action and a Show the note in Explorer footer link.

— Names the two UI actions but no further mechanism v1.0.1

07 **Streaming answers** IMPROVED 35

Answers now stream to the screen as they are written instead of appearing all at once.

— States the change but no mechanism detail v1.0.1

08 **GitHub star link in About settings** NEW 30

Settings → About now includes a GitHub star link that acknowledges when you have already starred.

— Thin, single-line cosmetic addition v1.0.1

WAS THIS USEFUL?  

knowledge-rag

Feedback ⓘ v5.1.0 NOTES ↗ RANK

Knowledge-rag is a local RAG server for Claude Code, Cursor and other MCP clients: hybrid BM25 + vector search with cross-encoder reranking on by default, 13 MCP tools, 33 file types including code, re-indexing on file changes. One Python process with embedded ChromaDB and FastEmbed ONNX: no Docker, database server or API key; offline after the first model download.

> knowledge-rag v5.1.0 adds configurable HNSW search depth for new collections, richer reindex/index-quality diagnostics, a more efficient and exclusion-aware file watcher, background BM25 rebuilds for large indexes, and MMR deduplication alongside reranking — plus a breaking change to retrieval evaluation baselines.

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

01 **Configurable HNSW search depth** NEW 78

New collections can set HNSW search depth via the `search.ef_search` setting, letting large-index deployments tune recall depth before a collection is created.

Tune recall depth on a large collection by setting HNSW search depth in config before the collection is created.

yaml

```yaml
search:
  ef_search: 200
```

— Names exact config key with a working snippet v5.1.0

02 **Reindex and index-quality reporting** IMPROVED 76

`get_reindex_status` now shows which files failed the last reindex and why, `get_index_stats` reports a new `self_recall` metric alongside existing stats, and reindex passes remove orphan chunks (chunks whose document left the catalogue) and report the count as `orphan_chunks_removed`.

— Names three concrete diagnostics, no usage example given v5.1.0

03 **More efficient, exclusion-aware file watcher** IMPROVED 76

The file watcher now reindexes only changed paths via the new internal index_paths() path instead of walking every indexed document on any event, applies `exclude_patterns` before scheduling reindex work so files written inside excluded folders no longer trigger reindex passes, and exclude patterns of the form `folder/*` now prune the entire folder during the scan walk instead of visiting every file inside it.

– Names internal mechanism and pattern syntax, no direct command v5.1.0

04 **Background BM25 rebuilds for large indexes** IMPROVED

68

On indexes above 20,000 chunks, BM25 rebuilds now run on a background thread so searches are no longer blocked after a document write or server restart; a keyword-only search (`hybrid_alpha=0`) still waits for the first build.

– Names threshold and flag but no direct config action v5.1.0

05 **Retrieval evaluation default alpha changed** BREAKING

67

`evaluate_retrieval` now searches with `hybrid_alpha=0.3` (matching the `search_knowledge` default) instead of the previous 0.5, making figures from earlier versions not comparable; a new baseline should be recorded.

– Names exact values and tells reader to rebaseline v5.1.0

THINNER COVERAGE BELOW

06 **MMR deduplication with reranker** NEW

36

MMR deduplication now runs when the reranker is enabled, suppressing near-duplicate chunks in favour of distinct results.

– Describes behavior only, no config or flag named v5.1.0

└──▷ **BREAKING ON UPGRADE**

! Figures recorded by `evaluate_retrieval` in earlier versions are not comparable to v5.1.0 results: the tool now searches with `hybrid_alpha=0.3` (the `search_knowledge` default) instead of 0.5. Record a new baseline.

WAS THIS USEFUL?  

knowhere

Feedback ⓘ v1.2.57 NOTES ↗ RANK

Knowhere extracts, parses, and outputs structured chunks ready for AI Agents and RAG.

> Knowhere's v1.2.57 release adds upload integrity checks, source size reporting, smarter retrieval cancellation, and groundwork for a safe OpenSearch migration alongside a new indexing timeout control.

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

01 **Configurable indexing timeout** NEW 75

The API and worker's indexing timeout is now configurable via the `KNOWHERE_INDEXING_TIMEOUT_SECONDS` environment variable, shown with a default of 30 seconds in this release.

Cap how long the indexing worker waits before timing out, to prevent stalled jobs from blocking the queue.

```env
KNOWHERE_INDEXING_TIMEOUT_SECONDS=30
```

— Exact env var and default given, directly settable v1.2.57

02 **OpenSearch cutover readiness** NEW 60

Ahead of a lexical search backend migration, releases now render the lexical backend recorded in the database control row and halt if a backend transition is in progress, so PostgreSQL and OpenSearch are never read simultaneously during cutover.

— Explains mechanism and safety guarantee, no direct user action v1.2.57

THINNER COVERAGE BELOW

03 **Upload integrity and size reporting** NEW 45

Guarded file uploads are now verified against their exact size and SHA-256 checksum to detect corruption or tampering at upload time, and completed file and URL jobs now report source byte sizes so callers can confirm full payload delivery.

— Mechanism named but no API/field to act on v1.2.57

04 **Retrieval cancellation on client disconnect** IMPROVED 35

v2 retrieval is now cancelled when the client disconnects, avoiding wasted work on abandoned requests.

— Clear behaviour but no surface to invoke v1.2.57

WAS THIS USEFUL? 👍 👎

Pinecone

Feedback ⓘ announced · Nov 12, 2026 BLOG ↗

Pinecone is a managed vector database that stores and queries embeddings for AI applications.

Pinecone adds Gemini Embedding 2 to the Documents API's integrated embedding, billing tokens on the Pinecone invoice via Python SDK 10.3.0 and the console.

NEW INTEGRATION Set `'model': 'gemini-embedding-2'` in the `embed` setting of a text field when creating a Documents API index to have Pinecone embed every write and query using Gemini Embedding 2 — no Google Cloud account or API key required.

Control embedding task semantics per field by setting `task_type` to `classification`, `clustering`, or `sentence_similarity`; override per query with `task_type` values `question_answering`, `fact_verification`, or `code_retrieval_query` in the `score_by` parameter.

Pinecone now supports Gemini Embedding 2 at 1,536 dimensions (or the model's full size when `dimension` is omitted) on serverless Documents API indexes via the console and Python SDK 10.3.0.

WAS THIS USEFUL? 👍 👎

EVALUATE

◆ AI Observability & Evals

UK AI Security Institute Inspect AI

Feedback ⓘ

0.3.290 CODE ↗

Inspect is the UK AI Security Institute's framework for large language model evaluations, covering datasets, solvers, scorers, and agentic eval tasks.

Inspect AI adds active_model_event_id(), stable message IDs across the agent bridge, excluded/zeroed score reason fields, and improved sentinel sidecar accounting.

└──▷ GET THIS VERSION ▾

```
$ git clone --branch 0.3.290
https://github.com/UKGovernmentBEIS/inspect_ai.git
# already have the repo? check out this version:
$ git checkout 0.3.290
```

Get the event ID of the in-flight model call attempt with the new active_model_event_id() function, enabling correlation of live calls to log events.

Score reasons now gain `excluded` and `zeroed` fields, and excluded epochs are left out of epoch reduction.

Each message now keeps one stable ID through the agent bridge and the log, and tool events record the ID of the assistant message that proposed them.

Logs of errored or cancelled evals now keep their total sample count.

Sentinel sidecar calls now sign requests instead of sending the eval token, and judging time is counted as waiting with judge model usage included in the eval.

NEW SUPPORTED PLATFORM Gemini 4 Carbon support

NEW SUPPORTED PLATFORM Gemini 4 Hydrogen support

NEW SUPPORTED PLATFORM Mistral Large 4 support

NEW SUPPORTED PLATFORM Muse Ember 1.0 and Muse Wonder 1.0 support

WAS THIS USEFUL? 👍 👎

GOVERN

◆ AI Runtime Defense & Guardrails

Prismor

Feedback ⓘ v1.75.0 NOTES ↗ RANK

Prismor is a self-hosted runtime control plane for AI agents. Observe or HITL approve or Block rogue tool calls before it executes: secret leaks, prompt injection, supply chain etc in a local dashboard. Agent agnostic (Claude, codex, langchain etc.)

> Prismor's latest release strengthens MCP gateway security with bearer-header authentication and repeated-call limiting, plus richer audit trails that capture resource, policy version, and approver details.

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

01 **Richer audit records with policy version and approver**

IMPROVED 70

Audit records now carry `resource` and policy version fields on entries, and approver identity on approvals, with a re-check automatically triggered after approval.

– Names specific fields and behavior but no API or UI path given v1.75.0

THINNER COVERAGE BELOW

02 **Bearer-header auth for hosted MCP endpoints** NEW 55

Hosted MCP endpoints can now be authenticated by passing the endpoint key as a bearer header, allowing secured access without additional credential setup.

– Names mechanism but no exact header syntax or endpoint path given v1.75.0

03 **Repeated-call limiting in MCP gateway** NEW 50

The MCP gateway can now limit repeated identical tool calls, reducing abuse or runaway automation from agents.

– Describes purpose but lacks configuration details or limits v1.75.0

WAS THIS USEFUL? 👍 👎

◆ Model & Agent Protection

Minimal Software Research minimal

Feedback ⓘ

release-a42a9d6d NOTES ↗ RANK

Minimal is a secure, declarative runtimes for Claude Code, Gemini, Codex, Pi and more, so you can run multiple agents safely on your stack from a verified software supply chain.

> Minimal shipped a new Box Egress Proxy for certificate-backed, audited egress control, made GitHub sign-in the primary login path via `min login`, and added a daily release-check notification.

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

01 Box Egress Proxy (`bep`) for audited, certificate-based egress

NEW 91

The new Box Egress Proxy component (`bep`) proxies declared egress for boxes: it issues leaf certificates under a name-constrained root and signing CA, appends hash-chained audit records per proxy decision, and rotates the signing CA through a control socket. Proxy keys are held in the host keychain — in the Secure Enclave on macOS (inside `bep.app`), in the TPM on Linux, or in Secret Service where no usable TPM is present. Store rules live under `min --config-dir` in step with `bep serve` , client-key binding is controlled with `--max-client-keys` on `bep serve` , `minzoned` and a signed, notarized `bep` ship alongside `minvmd` in a signed, notarized macOS app bundle, and a git credential helper inside grant boxes routes `git` and `gh` traffic through the Box Egress Proxy.

Inspect the full audit trail for a box, merging its parent chain, to verify no proxy decisions or egress events have been tampered with.

```
$ min box audit --parent mybox
```

Run a task with explicit egress controls so only declared upstreams are reachable, without changing the box definition.

```
$ min task run --network host --egress declared mybuildtask
```

— Names every flag, binary, socket and storage mechanism involved release-a42a9d6d

THINNER COVERAGE BELOW

02 **Daily release-check notification** NEW 25

Minimal now checks once a day whether the channel has a newer release available and reports it.

— Bare behaviour description with no config surface named `release-a42a9d6d`

└──▷ **ALSO FROM THESE RELEASES**

Store a secret token for a box so the agent can access it without it appearing in the environment at session start.

```
$ min secret set MY_API_KEY
```

WAS THIS USEFUL? 👍 👎

AI MODELS

◆ **Open-Weight Models**

DeepSeek V3 Feedback ⓘ price cut · Nov 14, 2026 MODEL ↗

DeepSeek V3 is a chat model from DeepSeek.

Price cut on OpenRouter (deepseek/deepseek-chat): $0.32 → $0.2574 in, $0.89 → $1.0287 out per million tokens (-20%), first seen 14 Nov 2026.

WAS THIS USEFUL? 👍 👎

Meta Llama 3.2 3B Instruct Feedback ⓘ

retirement scheduled · Nov 14, 2026 MODEL ↗

Llama 3.2 3B Instruct is a chat model from Meta.

OpenRouter lists a retirement date of 20 Nov 2026 for meta-llama/llama-3.2-3b-instruct, first seen 14 Nov 2026.

WAS THIS USEFUL? 👍 👎

MiniMax M2.5 Feedback ⓘ context change · Nov 14, 2026 MODEL ↗

MiniMax M2.5 is a reasoning model from MiniMax.

Context window shrank from 200K to 198K tokens on OpenRouter (minimax/minimax-m2.5), first seen 14 Nov 2026.

WAS THIS USEFUL? 👍 👎

Alibaba Qwen3.8 27B Feedback ⓘ

context change · Nov 14, 2026 MODEL ↗

Qwen3.8 27B is a multimodal model from Alibaba.

Context window shrank from 1049K to 1024K tokens on OpenRouter (qwen/qwen3.8-27b), first seen 14 Nov 2026.

WAS THIS USEFUL? 👍 👎

OTHER / UNCATEGORIZED

◆ UNCATEGORIZED

Tokenrip

Feedback ⓘ changes since 2026-10-16 CODE ↗

Tokenrip's command-line client.

Tokenrip: Adds `rip auth add` and `rip auth logout` for named per-agent sign-ins, plus `rip job`, `rip output`, `rip feedback`, and `rip message`.

Filter task lists by `--type`, `--step`, and `--label` tags, by `--unassigned` (tasks assigned to no one, equivalent to `assignee=none`), and by `--job` name.

Track task history with `rip task` timeline output via the new `formatTaskTimeline` formatter.

Identities file now enforces tight file-system permissions via `tightenMode` on every load and save.

Table row writes now report limit hints when a row-write limit is reached, and conflict errors on row writes include `currentTableRevision`, `artifact`, and `current` (id and version number) in `DomainErrorDetails` to support optimistic-concurrency retries.

Remote MCP server files for directory listings and a Cursor plugin manifest (`mcp.json`) are now included.

+1 more

└──▷ **BREAKING ON UPGRADE**

- ! `rip trash restore` no longer takes a `<type>` argument before `<id>`; callers passing `rip trash restore <type> <id>` must drop the type.
- ! `rip comment` is replaced by `rip message` for workspace threads; scripts using `rip comment` will break.
- ! The `--audience`, `--membership`, and `folder share-contents` options are removed (flat workspace access model); any automation using these flags will fail.
- ! The `parent` option on task list is renamed to `parentId`; callers passing `--parent` must switch to `--parent-id` or the equivalent.

WAS THIS USEFUL? 👍 👎

// END OF TRANSMISSION

▷ Subscribe Feedback on this issue