THE AI TOOLCHAIN

NO. 102

Tail

THE DAILY RELEASE FIREHOSE

PUBLISHED NOVEMBER 26, 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

INFORMATION DENSITY

everything



BRIEFED | EVERYTHING

Every tool in the issue, in full.

all 886 changes

Do you prefer **scroll**? 👍 👎

▸ // CONTENTS SHOW 150 of 154 tools

▾ // THE BRIEF HIDE SAME ISSUE, SAME PROMPT, TWO WRITERS: GPT-5.5 Opus 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 today: Cotal agents can switch models and stream events without being respawned; e2e can test macOS

vim ? help

desktop flows and mock network paths; gram gives MCP servers an access-request loop so tool use can be approved instead of guessed.

BUILD

Change a live agent session without tearing it down

Cotal added streamed agent events, live model switching without despawn, thread titling, harness command discovery, and spawn-time permission gating. That removes the restart-and-recreate loop when debugging or steering an agent session, so developers can inspect, rename, repermission, and retarget a running seat instead of losing its state.

cotal

EVALUATE

Exercise desktop and web agents against controlled OS and network conditions

e2e added a macOS desktop testing engine, an HTTP `request` fixture, WebMCP tool introspection, offline network simulation, and Playwright-style URL-glob routing. Teams can now test agent behavior across local app UI, tool surfaces, and failing or mocked network paths without wiring separate harnesses for each environment.

e2e

GOVERN

Put approval and policy in front of agent tool access

gram added agent access requests, feedback for MCP servers, intent-based policy guardrails, and read-only gateway execution; OpenShell added offline policy linting, Cedar-based sandbox policies, richer audit records, workspace secrets, and SPIFFE sandbox identities; minimal added `min secret`, tamper-evident box audit trails, and scoped Box Egress Proxy controls. The common gain is replacing informal trust in agent actions with explicit permission, reviewable policy, and audit evidence before tools or networks are touched.

gram · OpenShell · minimal

EVALUATE

Measure whether a skill version actually improved

SkillEvaluator added paired-case lift statistics, skill-version comparison, grader auditing, hook waivers, and a locked CI container for Tier 2/3 evaluation. Instead of eyeballing separate benchmark runs, teams can

> compare versions on matched cases and make regressions or real lift visible in CI.
>
> SkillEvaluator

Does **GPT-5.5** read better? 👍 👎

OFFENSIVE SECURITY

◆ Exploitation & C2

xalgorix

Feedback ⓘ v4.6.209 NOTES ↗ RANK

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

> xalgorix shipped finer control over wildcard scan scope, smarter retention of security-relevant data as context ages out, and a lower-cost compact checklist mode for its 22-phase methodology.

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

01 **Subdomain exclusion for wildcard scans** Not checked yet NEW

93

A server-wide exclusion list can be set via the `XALGORIX_WILDCARD_EXCLUDE` environment variable (comma-separated) or under Settings → Runtime, to skip chosen hosts across all wildcard scans. A per-scan `exclude_subdomains` list can also be added to the scan request to skip specific subdomains for a single wildcard engagement, with both sources merged at scan time. Exclusion values support glob patterns (`dev-*` , `*-qa` , `*.internal.example.com`) or plain DNS labels (`staging`), matched case-insensitively and applied before the optional subdomain cap.

Block dev, staging, and all internal subdomains server-wide so they are never turned into full sub-scans on any wildcard engagement.

```
$ XALGORIX_WILDCARD_EXCLUDE=dev,staging,*.internal.example.com
xalgorix scan --target '*.example.com'
```

Exclude QA and ephemeral dev hosts for a single scan request without touching the global setting.

```json
{
  "target": "*.example.com",
  "exclude_subdomains": ["dev", "staging", "*-qa"]
}
```

— Names env var, config key, patterns and runnable commands v4.6.209

02 **Smart stub retention for aged-out tool results** Not checked yet

NEW

84

The new `XALGORIX_TOOL_ARCHIVE_SMART_STUB` setting (opt-in, off by default) retains security-relevant excerpts — reflected payloads, error strings, secrets and tokens, HTTP status and notable headers, endpoints — when a tool result ages out of the active context window, while the full original stays archived and retrievable.

Preserve reflected payloads, secrets, tokens and notable headers in the agent's working context even after a tool result ages out, so findings are not silently dropped mid-session.

```
$ XALGORIX_TOOL_ARCHIVE_SMART_STUB=true xalgorix
```

— Names env var and exact retained data categories, opt-in by default v4.6.208

03 **Compact checklist mode for scan methodology** Not checked yet NEW

76

Setting `XALGORIX_CHECKLIST=compact` in Settings → Scan checklist or as an environment variable runs the full 22-phase methodology at roughly half the prompt-token cost (~14K tokens vs ~27K), without requiring a restart.

— Names config key and token figures but no example command v4.6.206

WAS THIS USEFUL? 👍 👎

BUILD

◆ Model Training & Fine-Tuning

PyTorch

Feedback ⓘ announced · Nov 26, 2026 BLOG ↗

PyTorch tensors and Dynamic neural networks in Python with strong GPU acceleration

Torch-TensorRT 2.15 and ExecuTorch 1.6 add a production-ready path to compile PyTorch models to TensorRT engines packaged as ExecuTorch `.pte` programs runnable from C++ or Python without libtorch.

Compile a PyTorch model to a TensorRT-backed ExecuTorch `.pte` program using `output_format="executorch"` in torch_tensorrt.save(), runnable from C++ or Python without loading torch. Not checked yet

NEW SUPPORTED PLATFORM Install the full compile-and-run stack with `pip install "torch_tensorrt[executorch]"`; install only the lightweight robot runtime with `pip install --no-deps executorch==1.6.0 torch-tensorrt-executorch-runtime==0.2.1 nvidia-cuda-runtime==13.2.86 tensorrt-cu13-libs==11.4.0.106 numpy`. Not checked yet

Route ops TensorRT cannot handle to ExecuTorch's GPU path at compile time by passing `CudaPartitioner` to torch_tensorrt.save(), keeping the whole model in one `.pte` program on a single CUDA stream instead of failing the build. Not checked yet

Enable CUDA graph replay for every TensorRT engine call by setting `use_cuda_graphs=True` in torch_tensorrt.dynamo.compile(), reducing per-call GPU launch overhead. Not checked yet

Bundle multiple model components (e.g. a vision encoder and an action head) into one `.pte` program, one method per component, using torch_tensorrt.executorch.export(). Not checked yet

+3 more

└──▷ BREAKING ON UPGRADE

! Torch-TensorRT 2.15 requires TensorRT 11.4, but JetPack 7.2 ships with TensorRT 10 — the install pulls its own TensorRT libraries and cannot reuse the JetPack-bundled ones.

WAS THIS USEFUL? 👍 👎

Hugging Face TRL

Feedback ⓘ v1.19.0 NOTES ↗ RANK

TRL trains transformer language models with reinforcement learning, covering supervised fine-tuning, reward modeling, PPO, DPO, and GRPO.

> TRL v1.19.0 introduces multi-teacher distillation and environment-based agent distillation, extends AsyncGRPO to Mamba2 and LFM architectures with a major planning/packing speedup, broadens assistant-only-loss and tool-calling template support across many model families, and ships a batch of breaking changes including required peft/accelerate version bumps and a changed parse_response signature.

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

01 DistillationTrainer new capabilities

Not checked yet NEW

85

Adds multi-teacher distillation (MOPD) support to `DistillationTrainer`, letting a dict of teachers be routed per example via a `teacher_id` column, and adds an `environment_factory` argument matching the GRPOTrainer contract so agents can be distilled against stateful environments.

Distil a student from two specialist teachers, routing math examples to one and code examples to the other via a `teacher_id` column in the dataset.

python

```
from trl import DistillationTrainer
from datasets import load_dataset

dataset = load_dataset("my-org/math-code-mix", split="train")  #
must have a 'teacher_id' column: 'math' or 'code'

trainer = DistillationTrainer(
    model="Qwen/Qwen2.5-0.5B-Instruct",
    teacher_model={"math": "Qwen/Qwen2.5-Math-1.5B-Instruct",
"code": "Qwen/Qwen2.5-Coder-1.5B-Instruct"},
    train_dataset=dataset,
)
trainer.train()
```

— Runnable code example shows exact usage No usage example or command given v1.19.0

02 AsyncGRPO support for Mamba2 and LFM architectures

Not checked yet NEW

65

AsyncGRPO now supports Mamba2-based architectures (Nemotron-H, Granite-4.0-H, Bamba) with tree packing via new Triton tree scan, conv, and fused mixer kernels, and also supports LFM models in tree attention.

— Names architectures and kernels but no run command v1.19.0

---

03 **AsyncGRPO planning and packing speedup** Not checked yet IMPROVED

65

AsyncGRPO micro-batch planning and packing now uses array operations, cutting planning + packing time from 18s to 1s per pass and raising trained tokens/s by 28% on Qwen3-1.7B math rollouts.

— Concrete before/after numbers but no config shown v1.19.0

---

04 **Automatic end-of-turn token fix in SFTTrainer** Not checked yet

IMPROVED

65

SFTTrainer now adds missing end-of-turn tokens to `generation_config.eos_token_id` for models such as Qwen base, Gemma 1/2 instruct, and Qwen3.5 that previously omitted them from stop tokens.

— Names exact config field and affected models v1.19.0

---

05 **assistant_only_loss extended to more model families**

Not checked yet IMPROVED

65

`assistant_only_loss=True` now works with DeepSeek-V3.1, DeepSeek-V3.2-Exp, DeepSeek-R1-Distill, SmolLM2, Phi-4, Gemma 4, GLM-4.7, GLM-5, Qwen3-Coder, Llama 3.1, Llama 3.2, OLMo 3 Think, LFM2-24B-A2B, and Inkling.

— Names flag and full list of models it now covers v1.19.0

---

06 **Minimum peft and accelerate version bumps** Not checked yet

60

BREAKING

`peft>=0.18.0` is now required, with peft 0.17 compatibility code removed, and `accelerate>=1.12.0` is now required, with older accelerate fallbacks removed.

— Names exact minimum versions for upgrade v1.19.0

THINNER COVERAGE BELOW

---

07 **HF Jobs detection via HF_JOB_ID** Not checked yet NEW

55

TRL now detects HF Jobs via the `HF_JOB_ID` environment variable and adds the hardware flavor to the model card automatically.

— Names exact environment variable to set v1.19.0

---

08 **S3 multi-agent aggregation example** Not checked yet NEW

50

Added an S3 multi-agent aggregation example showing held-out team success rising from 4.7% to 60.9%.

– Concrete result numbers but no runnable snippet provided v1.19.0

09 **Tool-calling chat templates for DeepSeek models** Not checked yet

NEW

45

Added tool-calling chat-template support for DeepSeek-R1-0528, DeepSeek-V3-0324, DeepSeek-V3.1, and DeepSeek-V3.2-Exp.

– Names models but no mechanism or example v1.19.0

10 **parse_response now requires prefix argument** Not checked yet

45

BREAKING

`prefix` is now a required argument of `parse_response`; callers that omitted it will break.

– Names the exact signature change needed v1.19.0

11 **Gradient checkpointing renamed to activation checkpointing**

Not checked yet

40

BREAKING

Following the transformers 5.23 rename, gradient checkpointing is now called activation checkpointing; any config or code using the old name will need updating.

– Names the rename but not the specific keys affected v1.19.0

12 **ModelConfig.model_revision default changed** Not checked yet

30

BREAKING

`ModelConfig.model_revision` now defaults to `None` instead of its previous default.

– States the change but not the old default value v1.19.0

└──▷ **BREAKING ON UPGRADE**

! `prefix` is now a required argument of `parse_response`; callers that omitted it will break.

! `peft>=0.18.0` is now required; peft 0.17 compatibility code has been removed.

! `accelerate>=1.12.0` is now required; older accelerate fallbacks have been removed.

! `ModelConfig.model_revision` now defaults to `None` instead of its previous default.

! Gradient checkpointing is renamed to activation checkpointing following the transformers 5.23 rename; any config or code using the old name will need updating.

WAS THIS USEFUL?  

fizgig

Feedback ⓘ v9.8.7 NOTES ↗ RANK

Fizgig — LoRA & Fine-tune Studio for Klein 9B, Krea 2, MiniMax H3, Qwen Image 2.1, Anima & Z-Image Turbo and Ace-Step XL: train, fine-tune, profile, repair and extract LoRAs & LoKR for Image, Video & Music Models

> Fizgig's two releases add a GPU-specific Speed-ups panel that cuts Krea 2 LoRA training step time on an RTX 5090, alongside drag-and-drop uploads, reduced AMD memory use for Qwen-Image previews, new MiniMax H3 Image controls for reference Strength and Speed boost, native-shape reference pictures for edit models, and offline-capable caching.

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

01 GPU-specific Speed-ups panel in Preferences

Not checked yet NEW

95

Fizgig groups GPU acceleration settings under Preferences → Speed-ups for your graphics card: NVIDIA gets INT8 fast inference, Faster block swap, Keep swapped blocks in pinned RAM, and (Windows only) a guide to stopping the driver moving training into system RAM; AMD gets INT8 fast inference, `Force SageAttention`, Split large convolutions, and Use AMD's MIOpen. Enabling Faster block swap for Krea 2 LoRA training cut step time from 3.28 s to 2.17 s on an RTX 5090 (INT8, 16 blocks swapped).

Speed up Krea 2 LoRA training on NVIDIA hardware without increasing VRAM by enabling faster block swap.

> 📍 In Preferences → Speed-ups for your graphics card, tick Faster block swap, then start a Krea 2 LoRA run and watch s/step in the log to confirm the improvement.

— Full per-GPU option list, exact UI path and benchmark numbers given. v9.8.7

02 New MiniMax H3 Image controls

Not checked yet NEW

70

MiniMax H3 Image gains two new controls: a reference Strength control in Studio and LoRA Royale — set it to `1` to use the picture as-is, lower values reduce its influence, and `0` omits it entirely — and Speed boost rendering in the Samples tab and Studio, where the first steps render at half size (2 steps or a third of them, whichever is more) and the rest render at full size.

— Exact value semantics and locations given. Mechanism and location given, no measured speed gain.

v9.8.6

03 Tiled VAE decode cuts AMD memory use

Not checked yet IMPROVED

65

AMD cards now decode the Qwen-Image VAE in tiles and run large convolutions in strips, using roughly a third of the previous memory, so a 1024×1536 preview no longer reserves around 30 GB on a 16 GB card.

— Concrete numbers and mechanism, but automatic with no user action. v9.8.7

04 **Every epoch sheet reads typed epoch lists** Not checked yet

IMPROVED 65

LoRA Royale's Every epoch sheet now reads epochs typed directly into the Epochs box (for example `5,11,22,38`) straight from the training folder, without rendering them at the top first.

— Exact field name and example syntax given. v9.8.7

05 **Reference pictures kept at native shape for edit models**

Not checked yet IMPROVED 65

Edit models (Klein, Flux 3, Qwen Image, Fibo 2 and MiniMax H3 Image) now use reference pictures in their own shape at the Reference MP rather than cropping them to the render dimensions, and Studio's Reference MP box is no longer greyed out for edit models.

— Names all affected models and the UI field changed. v9.8.6

06 **Drag-and-drop file uploads across upload boxes** Not checked yet

NEW 60

Dragging a file from the desktop onto any box with an Upload button — training form picture rows, Samples reference pictures, Studio, LoRA Royale, the Profiler, Extract, and Metadata — is now accepted in place of using the file picker.

— Names every affected surface but gives no runnable step. v9.8.7

THINNER COVERAGE BELOW

07 **LoRA Isolation masks background for two-subject LoRAs**

Not checked yet IMPROVED 50

LoRA Isolation now masks the stranger's half of side-by-side training pictures and appends unique caption words, so two-subject LoRAs no longer split the background between characters.

— Explains mechanism but no UI path or settings named. v9.8.7

08 **Offline caching for Anima, ACE-Step and SDXL configs**

Not checked yet IMPROVED 45

Caching now works offline: Anima's and ACE-Step's tokenizers and SDXL's config load from downloaded files first, so the internet is not needed once a model is downloaded.

— Describes behavior but no file path or flag named. v9.8.6

09 **Caption search result highlighting** Not checked yet IMPROVED

20

Captions search now highlights matching words inside each caption result.

— Single bare sentence with no further detail. v9.8.7

WAS THIS USEFUL? 👍 👎

DeepSpeed

Feedback ⓘ v0.20.1 NOTES ↗ RANK

DeepSpeed is a deep learning optimization library for distributed training and inference, providing ZeRO sharding, pipeline parallelism, and mixed-precision execution.

> DeepSpeed v0.20.1 ships a wide batch of distributed-training and inference optimizations, led by ReFlow asynchronous CPU-offloaded optimizers for ZeRO-3, a fused CPU Adam step, XPU graph capture for the Hybrid Engine, and a one-shot P2P all-reduce for `TP=2` inference, alongside incremental gains to AutoTP, AutoEP, AutoSP and Ulysses sequence parallelism.

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

01 AutoEP expert gradient staging and combine overlap

Not checked yet IMPROVED 85

AutoEP adds opt-in skip of the recomputed DeepEP combine step, stages expert weight gradients across microbatches (including under `ZeRO-3`), overlaps expert weight gradients with the DeepEP backward combine, and supports `combine_impl='fused_weighted_sum'` with AutoEP+AutoTP folding.

– Names concrete config value and multiple mechanisms v0.20.1

02 Fused CPU Adam optimizer improvements Not checked yet IMPROVED

70

Adds a fused CPU Adam step for faster CPU-side optimizer updates, writes CPU optimizer results back to non-contiguous parameters, and warns when `DeepSpeedCPUAdam` is limited to one OpenMP thread.

– Names component and three concrete behavior changes v0.20.1

THINNER COVERAGE BELOW

03 Transfer schedule control in checkpointing Not checked yet NEW

55

Exposes the affine transfer schedule as `schedule_transfers` in checkpointing to control transfer ordering.

– Names exact config key but no behavior detail beyond that v0.20.1

04 XPU graph capture for Hybrid Engine decode Not checked yet NEW

55

Adds XPU graph capture (opt-in) and segment-based kernel injection for the Hybrid Engine's greedy decode.

– Describes what changed but not how to opt in v0.20.1

05 **Uneven-shard rejection and FPDT padding for Ulysses sequence parallelism**

Not checked yet NEW 50

Adds opt-in uneven-shard rejection and FPDT padding for Ulysses sequence parallelism.

– Names the mechanisms but not how to enable them v0.20.1

06 **One-shot P2P all-reduce for TP=2 inference** Not checked yet NEW

45

Adds a one-shot P2P custom all-reduce for tensor-parallel inference at `TP=2.`

– Specific scope (TP=2) but no mechanism or usage path v0.20.1

07 **AutoSP loss aggregation and label alignment** Not checked yet

IMPROVED 45

AutoSP now aggregates the causal LM loss and gradients over the SP group, and aligns cross-entropy targets after an in-model causal label shift.

– Describes two behavior changes without implementation detail v0.20.1

08 **AutoTP untied input embedding sharding** Not checked yet IMPROVED

40

AutoTP now shards an untied input embedding along the vocabulary dimension.

– Single-line behavior change with no further detail v0.20.1

09 **ReFlow async CPU-offloaded optimizers for ZeRO-3** Not checked yet

NEW 35

Adds ReFlow asynchronous CPU-offloaded optimizers for ZeRO-3 training.

– Bare feature name, no usage or mechanism detail v0.20.1

WAS THIS USEFUL?  

◆ AI Coding Agents

SST OpenCode

Feedback ⓘ updated · Nov 26, 2026 NOTES ↗ RANK

OpenCode is an open-source AI coding agent for terminal-based development tasks.

> OpenCode updated its system prompts for GPT-6.2-based model aliases and adjusted model picker ordering to surface OpenCode Go first.

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

01 **GPT-6.2 system prompt for model aliases** Not checked yet IMPROVED

45

OpenCode now uses a system prompt written for GPT-6.2 when running the Sol, Astra, and `gpt-latest` model aliases.

— Names the aliases but not prompt mechanism details snapshot-20261126

02 **OpenCode Go listed first in model picker** Not checked yet IMPROVED

30

After connecting OpenCode, the model picker now lists OpenCode Go first.

— Simple UI ordering change, thinly described snapshot-20261126

WAS THIS USEFUL? 👍 👎

Cline

Feedback ⓘ sdk/sdk/v0.2.2 NOTES ↗ CODE ↗ RANK

Cline autonomous coding agent as an SDK, IDE extension, or CLI assistant.

> Cline's SDK shipped a model catalog refresh adding new providers and pruning deprecated models, alongside breaking API changes to global settings and telemetry helpers.

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

01 **Breaking changes to SDK settings and telemetry exports**

Not checked yet BREAKING 80

`writeGlobalSettings`, `setTelemetryOptOutGlobally`, and `setAutoUpdateEnabledGlobally` no longer accept an options argument, so callers passing one will break. Several telemetry helper exports were also removed from `@cline/llms`, `@cline/shared`, and `@cline/core`, breaking any code that imports them.

— Names exact functions and packages affected for migration sdk/sdk/v0.2.2

02 **Model catalog adds providers, removes deprecated models**

Not checked yet IMPROVED 70

The bundled model catalog adds the `concentrate` and `primalabs` providers and updates default model selections for 28 providers. It also removes deprecated OpenAI model IDs — `gpt-4o`, `gpt-4o-mini`, `gpt-4.1`, `gpt-4.1-mini`, `gpt-5`, `gpt-5-mini`, `gpt-5-nano`, `gpt-5-pro`, `gpt-5.1`, `gpt-5.3-codex`, `gpt-5.3-codex-spark`, `gpt-5.4-nano`, `gpt-image-2`, `o3`, and `o3-pro` — and removes `MiniMax-M3` from the MiniMax coding plans.

— Names every added provider and removed model ID sdk/sdk/v0.2.2

└──▷ **BREAKING ON UPGRADE**

! `writeGlobalSettings`, `setTelemetryOptOutGlobally`, and `setAutoUpdateEnabledGlobally` no longer accept an options argument — callers passing one will break.

! Several telemetry helper exports were removed from `@cline/llms`, `@cline/shared`, and `@cline/core` — code importing them will break.

WAS THIS USEFUL? 👍 👎

yetone magpie

Feedback ⓘ v0.1.4837 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's heaviest work this window was under-the-hood routing and provider logic: a new PaceDrain routing strategy, per-group classifier prompts, routing via OpenAI's Decisions API, key-possession proofs for private-network failover, and a batch of Google API compatibility fixes, alongside a gateway privacy switch and smaller skill-library, UI, and desktop-detection improvements.

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

01 **GPU rendering toggle for Linux hybrid laptops** Not checked yet

NEW 85

Settings › General › Advanced › GPU rendering lets you switch WebKitGTK compositing on or off without editing environment variables, controlling `WEBKIT_DISABLE_COMPOSITING_MODE` and `WEBKIT_DISABLE_DMABUF_RENDERER` behaviour directly from the UI.

— Exact settings path and environment variables named v0.1.4835

02 **Google provider compatibility fixes in /v1/systemone**

Not checked yet IMPROVED 80

`/v1/systemone` now forwards audio sent to Google as audio blocks via `data:` URLs, alongside existing image block support. noul question criteria named `true` and `false` are mapped automatically to Google's `true_description` and `false_description`, so Google no longer rejects them with `Unknown parameter`. A Google or Bailian vendor decision model such as `google/gemini-decision-1` is now served regardless of which models a key's `AccountModels` list contains, so it can no longer be accidentally blocked.

— Endpoint, field names and model id all specified v0.1.4834

03 **Maker-grouped model ordering in provider editor** Not checked yet

IMPROVED 80

The provider editor now groups models by maker and sorts them newest-first within each group, comparing version numbers segment by segment (so 5.10 sorts before 5.9) and placing releases before previews, via `modelorder.js` and its `MAKERS` map covering Anthropic, OpenAI, Google, DeepSeek, xAI, Qwen, Moonshot AI, Zhipu, MiniMax, and others; models from an unrecognised maker fall under Other in the vendor's original order.

— Sorting mechanism and full maker list named v0.1.4832

04 **Routing classifier via OpenAI Decisions API** Not checked yet NEW

80

Routing rules can now classify requests through OpenAI's `POST /v1/decisions` endpoint by setting a routing rule's classifier to an OpenAI key's `gpt-6-luna` model.

Direct a routing rule's classifier at OpenAI's Decisions API so that `gpt-6-luna` determines which upstream each request goes to.

```
$ POST /v1/decisions with a routing rule classifier set to an OpenAI key's gpt-6-luna
```

— Exact endpoint and model id with runnable example v0.1.4830

05 **Plugin routing tables included in state reads** Not checked yet

IMPROVED

80

Plugin routing tables (`jsRoutings`) are now included in every state read — state(), groupsState(), and pluginsState() — so pickers built from any of these reads carry the full routing list without opening Settings.

— Named functions and exact behaviour change v0.1.4827

06 **Key-possession proof for private-network providers**

Not checked yet NEW

80

The `netproxy` package adds `Vouch` and `WithVouch` so a failover address on a private or link-local network must prove it holds the provider's gateway key — via `RemoteProvePath` — before any keyed request is sent to it.

— Named functions and path with clear security mechanism v0.1.4823

07 **Per-group classifier prompts** Not checked yet NEW

80

A group's own classifier instructions can be set via `ClassifyPrompt` on the `Group` struct (saved trimmed, enforced at `MaxClassifyPrompt` characters), so the classifier uses those instead of magpie's built-in prompt when routing turns for that group.

— Named field, struct and limit constant specified v0.1.4820

08 **Skill library gains Dim skills and agent filter** Not checked yet

NEW

70

The skill library now reads Dim agent skills from `~/.agents/skills` and a project's `.agents/skills` — the same folder shared with Goose — and Codex also picks up Dim's skills from that path. You can also filter the library down to a single agent's

installed skills via a new agent picker, which appears when the library has more than one skill and more than one agent.

— Named folder paths; integration with Goose and Codex detailed v0.1.4817 v0.1.4837

09 **Auto-discovered routing group for single-provider models**

Not checked yet NEW 70

A model served by only one provider can now form its own auto-discovered routing group via `AutoGroupsOfOne` / `SetAutoGroupsOfOne`, surfaced as a toggle under Advanced in the Routing groups UI when found groups are enabled.

— Named functions and exact UI location given v0.1.4836

10 **PaceDrain routing strategy** Not checked yet NEW 70

The new `PaceDrain` routing strategy picks the account with the most quota remaining per hour (by `Seat.Pace`) and stays on it until spent, then moves to the next account by pace — combining the existing Pace and Drain strategies.

— Strategy and mechanism named, no setup command given v0.1.4827

11 **Request-content retention toggle in gateway** Not checked yet NEW 65

A Show request content switch in the gateway lets you turn off request-body retention: when off, no request text is kept in memory, traces, or the Recent-calls ring, while the call and its route are still recorded.

— Named toggle and scope, but no exact settings path given v0.1.4837

12 **Detection of DSH NEXT and DSH Desktop shells** Not checked yet NEW 65

magpie now detects the DSH NEXT and DSH Desktop community Electron shells (`DSH (NEXT|Desktop)\.app/`, `dsh-desktop(-next)?` processes) as recognised DeepSeek Harness desktop app variants, alongside the official app.

— Exact process/app patterns named, limited behavioural detail v0.1.4816

13 **Higher reply token limit for thinking Gemini models**

Not checked yet IMPROVED 65

magpie now detects when a Google Gemini model thinks by default (Gemini 2.5 and later, including Gemini 4 Hydrogen and high-temperature Pro variants) and raises the reply token limit so reasoning tokens don't consume the entire budget and leave the response truncated.

— Model list named but no adjustable setting given v0.1.4815

THINNER COVERAGE BELOW

14 **Per-API routing for dsh routing groups** Not checked yet IMPROVED

55

The `dsh` agent now routes a routing group onto its Messages or Gemini API route when every member of that group is best on that API, matching the routing a single member alone would receive.

– Behaviour described but no exact config or command v0.1.4813

WAS THIS USEFUL? 👍 👎

Letta Code Feedback ⓘ v0.34.71 NOTES ↗ RANK

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

> Letta Code's single release in this window expanded subagent capabilities with a new worker backend and added persistent lifecycle tracking with a listing command.

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

01 **Persistent subagent lifecycle and ListAgents command**

Not checked yet NEW

55

Subagent lifecycle is now persisted across sessions, and a new `ListAgents` command lets you enumerate active subagents.

– Names a runnable command but lacks mechanism detail v0.34.71

02 **Antigravity CLI worker backend for subagents** Not checked yet

NEW

40

Subagents now support running Antigravity CLI workers via the new worker backend, expanding the set of execution backends available to subagents.

– Names the backend but no usage detail v0.34.71

WAS THIS USEFUL? 👍 👎

forgedock

Feedback ⓘ v1.67.0 NOTES ↗ RANK

ForgeDock — autonomous AI dev pipeline for Claude Code. GitHub as agent memory: issue in, PR out.

> ForgeDock shipped self-healing standing-order batches, a serial merge train, outbound-only agent connectivity with one-command install, and model-aware context limits, alongside engine efficiency SLOs and incremental git-bundle sandbox seeding.

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

01 Serial merge train for PR merging

Not checked yet NEW

93

A serial merge train, enabled via `orchestration.merge_train` in `forge.yaml`, exits lanes at PR-ready and merges green PRs in sequence; phase 3 moves base-sync into the train, tracks per-batch counts, and measures throughput to choose the default.

Turn on the merge train so green PRs are merged serially without manual intervention, reducing merge-queue pile-ups in a busy batch

yaml

```yaml
orchestration:
  merge_train: true
```

— Named config key with runnable YAML example and phased mechanism v1.64.0

02 Self-healing standing-order batches

Not checked yet NEW

84

Standing-order batches run via `forgedock orchestrate`, auto-joining new matching issues, auto-readmitting on label change, self-healing halted lanes, and closing delivered issues — removing the need for watcher or operator sweeps.

Run a standing-order batch that automatically picks up new matching issues and self-heals halted lanes without manual intervention.

```
$ forgedock orchestrate 41 42 43 --parallel 3
```

— Runnable command plus full mechanism of auto-join and self-heal v1.67.0

03 Outbound-only agent connectivity and installer

Not checked yet

IMPROVED

79

Agents now connect to the router outbound-only, removing the need for inbound network access to the agent host; adds one-command install/uninstall via the `/install` route, `--uninstall` on Linux, and a macOS installer, plus account-linked device-code enrollment and account pool viewing.

— Names route, flag, installer and enrollment mechanism v1.64.0

04 **Per-host lane ceiling and sandbox disk cap** Not checked yet NEW

70

A per-host lane ceiling cannot be overridden by `max_remote` in batch configuration, and a new lane disk hard cap with auto-prune, `fstrim`, and WSL sparse VHD support keeps sandbox disk use bounded.

— Names config key and disk mechanisms, no direct command v1.64.0

05 **Orchestration diagnostics and concurrency warnings**

Not checked yet IMPROVED 63

Pool status now shows borrowable versus used slots, throttle and drain reasons, and local pause state; ForgeDock also prints effective concurrency and its source, and warns when `--parallel` exceeds the host cap or `forge.yaml` is missing.

— Names flag and config file but no command example v1.64.0

THINNER COVERAGE BELOW

06 **Engine efficiency SLOs for lane health** Not checked yet NEW

53

New engine efficiency SLOs detect idle slots, looping lanes, and missed interventions automatically, auto-filing engine-gap issues and scheduling `fd-introspect` runs.

— Names the SLO mechanism but no command shown v1.67.0

07 **Model-aware context ceiling** Not checked yet IMPROVED 52

Context ceiling is now model-aware and cost-aware, dropping the hard 120K-token kill for models with up to 1M-context windows.

— Specific token limits named but no config surface v1.64.0

08 **Git bundle seeding of sandbox repos** Not checked yet NEW

48

Sandbox repos are now seeded from the dispatching host via incremental git bundle over the router, falling back to GitHub when the dispatching host is unavailable (GitHub-only fallback).

— Clear mechanism named but no usage example v1.65.0

WAS THIS USEFUL?  

Alibaba Qwen Code

Feedback ⓘ v0.28.6 NOTES ↗ RANK

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

> Qwen Code's v0.28.6 release adds reply translation and generated UI card interactivity to its Web Shell, moves core session handling onto an event-bus kernel, and expands reasoning-effort controls across Gemini and Kimi model routes.

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

01 **Expanded reasoning effort support for Gemini and Kimi models**

Not checked yet IMPROVED 73

Gemini 4 Neon and Hydrogen models now support `xhigh` reasoning effort (default `high`) across Gemini and Vertex routes, including dated, preview, and suffixed aliases; Kimi K3 models on custom OpenAI-compatible gateways now infer `low` / `high` / `max` reasoning effort with `max` as default and mandatory thinking, without per-route declarations.

— Names specific models, routes, and effort values precisely. v0.28.6

02 **Reply translation via displayLanguage setting** Not checked yet

NEW 72

A new `general.displayLanguage` setting (values: `off`, `zh-CN`, `en`) is available in Ink, OpenTUI, and Web Shell to translate replies and thoughts, while original untranslated responses remain stored server-side.

Show translated replies in Chinese in the terminal TUI while keeping original responses stored server-side.

```yaml
general.displayLanguage: zh-CN
```

— Names exact config key and values with usage example. v0.28.6

03 **Interactive generated UI cards in Web Shell** Not checked yet NEW 62

Generated UI cards now support `ui.copy` and `ui.download` actions in Web Shell for copying text or downloading a small text file, and persist view state via `ui.state`, with model guidance on restoring and saving values.

— Names three concrete action keys but lacks runnable command. v0.28.6

04 **Web Shell command handling: backgrounding and resend reliability**

Not checked yet IMPROVED 62

Background shell commands now move to the background instead of stopping when Send is used, with the model receiving the command's id, output location, and status; and prompts now submit with idempotency keys, retry transient failures twice, and show a message-level resend indicator instead of a blocking banner.

— Explains mechanism and before/after but no named API surface. v0.28.6

THINNER COVERAGE BELOW

05 **MessageRewriteMiddleware removed in favor of hook-based rewriting**

Not checked yet 50

BREAKING

`MessageRewriteMiddleware` is removed; reply and thought rewriting now goes through the hook-based route in Ink, OpenTUI, and Web Shell.

— Names removed component and replacement mechanism. v0.28.6

06 **Suspension metrics in instance metrics route** Not checked yet

NEW 48

The instance metrics route now reports suspension counts, including clean suspensions, reverted attempts, eligibility failures, and cold opens.

— Names metric categories but not the endpoint path. v0.28.6

07 **Event-bus kernel for headless and Ink TUI sessions**

Not checked yet IMPROVED 40

Headless sessions and the Ink TUI now run on the event-bus kernel by default, with cancelled runs settling within 10 seconds.

— Describes behavior change but no actionable surface. v0.28.6

└──▷ **BREAKING ON UPGRADE**

! `MessageRewriteMiddleware` is removed; reply and thought rewriting now goes through the hook-based route in Ink, OpenTUI, and Web Shell.

WAS THIS USEFUL? 👍 👎

octomind

Feedback ⓘ 0.74.0 NOTES ↗ RANK

Octomind is an open-source AI coding agent and agent runtime: one binary, any model, MCP-native. Runs in terminal, CI, or as a daemon.

> Octomind's three releases this window center on a deeper plugin and MCP tool-routing architecture — external lifecycle plugins with result-feedback judgements, a new `mcp` config block for code-mode tool dispatch, delegated context search, and freeform MCP tool parameters — alongside several breaking changes to session and runtime persistence.

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

01 **MCP tool routing and search in code mode** Not checked yet BREAKING

90

Added the `mcp` config block to control which tools code scripts can directly invoke or route externally; introduced ranked, searchable MCP tool search in code mode; routed direct and external MCP tool calls through code scripts to centralise tool dispatch logic; tracked script calls for MCP tools to give plugins and the runtime visibility into which scripts were invoked; and showed nested schemas in MCP tool call forms so compound tool arguments are fully rendered at call time. MCP tool routing is now configurable by code mode, so existing implicit routing behaviour may change.

— Names the exact config block and five concrete routing/search additions. 0.73.0

02 **External lifecycle plugins and judgement feedback** Not checked yet

75

BREAKING

Added support for external lifecycle plugins defined outside the main binary, which can receive session-model events and participate in the agent loop, plus plugin judgements with result feedback that let plugins evaluate tool results and steer runtime decisions across the full decision chain. The session model is now included in plugin events so plugins can branch on which model is active in a session. External lifecycle plugins are now a supported plugin type, so plugin registration and event dispatch contracts may have changed.

— Describes mechanism and scope but no concrete plugin API surface named. 0.73.0

THINNER COVERAGE BELOW

03 **Session Restore And Reload State Changes** Not checked yet BREAKING

55

Session restore IDs are now typed as `SessionId` instead of raw strings, so integrations passing raw ID strings may break; runtime state is now restored from session snapshots on reload, so sessions that previously started clean will now resume prior state; and plugin context is now persisted in session history across reloads, so existing sessions may replay differently.

— Names the exact type and the integration risk. States the behavioural change but no snapshot format detail. Clear persistence change but no field or format named.

0.73.0

04 **Freeform MCP tool call parameters** Not checked yet IMPROVED

45

MCP tool calls now accept freeform parameters, removing the previous restriction to structured inputs.

— Clear before/after but no example of a freeform call. 0.73.1

05 **Approval-based remote MCP skill loading** Not checked yet NEW

40

Remote MCP skills are now loaded only after explicit approval before they are added to a session.

— Clear behaviour change but no approval flow or command named. 0.74.0

06 **Deprecated sampling overrides now ignored** Not checked yet

DEPRECATED

30

Deprecated sampling overrides in config are now silently ignored rather than applied.

— Names the deprecated config area but not the specific keys. 0.73.0

07 **Model time context refreshed each turn** Not checked yet IMPROVED

30

Model time context is now refreshed before each turn so the model always has an accurate view of the current time.

— Explains the behaviour but no config or trigger surface. 0.73.0

08 **Delegated context search** Not checked yet NEW

25

The agent can now offload context lookups to a sub-agent via delegated context search, rather than performing the lookup itself.

— Names the capability but no mechanism or trigger detail. 0.74.0

09 **Formatted plugin response display** Not checked yet IMPROVED

15

Recognised plugin responses are now formatted for cleaner display of tool output.

— Bare description with no named format or surface. `0.74.0`

└──▷ BREAKING ON UPGRADE

! Plugin context is now persisted in session history (`persist plugin context in session history`) — existing sessions may replay differently on reload.

! Deprecated sampling overrides in config are now silently ignored (`ignore deprecated sampling overrides`).

! MCP tool routing is now configurable by code mode (`make MCP tool routing configurable by code mode`) — existing implicit routing behaviour may change.

! Session restore IDs are now typed as `SessionId` — integrations that pass raw ID strings may break.

! Runtime state is now restored from session snapshots (`restore runtime state from session snapshots`) — sessions that previously started clean on reload will now resume prior state.

! External lifecycle plugins are now a supported plugin type (`support external lifecycle plugins`) — plugin registration and event dispatch contracts may have changed.

WAS THIS USEFUL?  

Alibaba OpenCodeReview

Feedback ⓘ v1.14.4 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 v1.14.4 adds a skill for reviewing GitHub PRs from pinned snapshots, persists retry reports across sessions, refines C/C++ allowlist scanning, and adds Vietnamese translation.

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

01 **GitHub PR review from pinned snapshots** Not checked yet NEW

37

A new skill lets OpenCodeReview review GitHub PRs from pinned snapshots rather than the live head of the branch.

— Names the capability but no command or config to invoke it v1.14.4

02 **C/C++ allowlist scanning excludes test paths** Not checked yet

IMPROVED

35

OpenCodeReview now excludes C and C++ test and benchmark paths from allowlist scanning, reducing noise from non-production code.

— Names languages and scope but no config key given v1.14.4

03 **Session-end retry report persistence** Not checked yet IMPROVED

28

OpenCodeReview now persists retry reports at session end so they survive across sessions, instead of being lost when a session ends.

— Describes behaviour change but no interface to act on v1.14.4

04 **Vietnamese translation added** Not checked yet NEW

10

OpenCodeReview adds Vietnamese translation support, per the release summary.

— Mentioned only in summary with no further detail v1.14.4

WAS THIS USEFUL? 👍 👎

corecoder

Feedback ⓘ v0.19.0 NOTES ↗ RANK

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

> CoreCoder's latest releases add session forking and offline inspection for saved conversations, tighten up `--resume` error handling, and introduce automatic model fallback via an environment variable.

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

01 Automatic model fallback via CORECODER_FALLBACK_MODELS

Not checked yet NEW 95

Setting `CORECODER_FALLBACK_MODELS` to an ordered, comma-separated list of model names (e.g. `gpt-5.4-mini,deepseek-chat`) makes CoreCoder automatically switch providers after create-time transient retries are exhausted, without hiding 400 errors or replaying in-progress streams. Fallback selection — including calls made by sub-agents — is reported on stderr, token cost is charged against the model that actually served each response, and both OpenAI-compatible provider strings and LiteLLM cross-provider model strings are accepted.

Guard against a primary provider outage in CI by falling back to a secondary model automatically.

```
$ export CORECODER_FALLBACK_MODELS=gpt-5.4-mini,deepseek-chat
corecoder -p "add error handling to parse_config()" --yes
```

— Full mechanism, env var, example, format and billing behavior all given. v0.17.0

02 Fork saved sessions into new branches Not checked yet NEW

85

`--fork-session ID` paired with `--before-turn N` creates a new branch of a saved session at a specific retained turn; the new branch is then continued with `--resume NEW_ID`, leaving the original session untouched.

Branch a conversation at turn 3 to try a different prompt, without touching the original session.

```
$ corecoder --inspect-session abc123
corecoder --fork-session abc123 --before-turn 3
corecoder --resume NEW_ID
```

— Exact flags and a runnable branching workflow are given. v0.19.0

03 **Offline session inspection via --inspect-session** Not checked yet

NEW

70

`--inspect-session ID` lets you view a saved session's model, timestamp, message count, and turn count offline, without an API key; the conversation itself is never printed or modified.

Check what model and how many messages a saved session contains before deciding whether to resume it — works with no API key.

```
$ corecoder --inspect-session abc123
```

— Names the flag and metadata shown, read-only guarantee noted. v0.19.0 v0.18.0

04 **Safer --resume session validation** Not checked yet IMPROVED

65

`--resume ID` now validates the session file before configuring the model, reporting whether the file is missing, unreadable, or invalid, and no longer hides other saved sessions when one file is bad.

— Clear before/after behavior but no runnable example given. v0.18.0

WAS THIS USEFUL? 👍 👎

pstack-claude

Feedback ⓘ v0.9.156 NOTES ↗ RANK

pstack-claude is a port of Poteto's pstack agent framework to Claude Code, Codex, Copilot, Pi, and other AI coding assistants.

> pstack-claude's biggest window addition is a REST fallback for `watch-pr` when Claude Code cloud sessions hit a GraphQL-refusing GitHub proxy, alongside a new `/reframe` command, schema-aware Codex Arena model dispatch, expanded Codex subagent role/configuration surfaces, an `agents/openai.yaml` implicit-invocation policy, OpenCode 2 subagent registration, and assorted Windows and Pi reliability fixes.

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

01 **watch-pr REST fallback for Claude Code cloud** Not checked yet

NEW 93

In a Claude Code cloud session, `watch-pr` now reads pull requests over REST instead of stopping with exit 7 when the GitHub proxy refuses GraphQL, fetching PR facts from `pulls/<pr>`, review threads from `ccr/review_threads`, and checks from the head's check runs and statuses. Checks are de-duplicated the same way `gh pr checks` does — keeping the newest run per name, workflow, and event — while excluding Copilot review-agent runs but retaining CodeQL's default-setup runs. A blocked merge state is reported over REST when a required review is missing, rather than showing `review-required`. Partial reads are guarded: a `total_count` returning fewer rows, a full last page from a list read, a review comment that opens a thread the route did not return, and a thread route of another shape each count as query errors so a partial read cannot report ready.

— Names every endpoint, dedup rule, and partial-read guard condition. v0.9.153

02 **Codex subagent configuration: roles, context, input, enablement**

Not checked yet IMPROVED 85

The Codex mapping now names native agent roles — `general-purpose` maps to `default` and exploration maps to `explorer` — with the role omitted on full-history forks (which inherit the parent role) and when the spawn schema has no role field. Spawning a reviewer that should see only its brief is done with `fork_context: false` or `fork_turns: "none"`. `request_user_input` usage is now controlled by mode and client, with required decisions asked in plain text outside Plan mode. Codex subagents are now controlled with `agents.enabled`, and `docs/reference.md` no longer tells users to set `[features] multi_agent = true`.

— Names every flag and mapping but each addition stays terse. v0.9.154

03 **agents/openai.yaml policy suppresses implicit skill invocation**

Not checked yet NEW 85

The generator now writes an `agents/openai.yaml` file with `policy.allow_implicit_invocation: false` into every principle leaf, preventing Codex from listing `principle-*` skills in its default catalog; `generate --check` enforces this policy.

— Names config file, key, and the enforcing check command. v0.9.154

04 **Pi agent host (`hc`) behavior improvements** Not checked yet

IMPROVED 83

On Pi with `pi agent host: hc`, a resumed session now correctly ends an agent whose launching pi died between `hc run` and `hc get`, by querying `hc get` for the pid of the terminal named by the agent id and treating no answer within 2 seconds as the agent being gone; and a project's committed sheet now also controls the `pi agent host: hc` line for worktree agents, so agents run as RPC children when the committed sheet omits that line, even if the agent directory's sheet sets it.

— Full mechanism and config key, but a narrow edge case. Names the config line but not the sheet file's full format.

v0.9.156 v0.9.152

05 **/reframe command for project reframing** Not checked yet NEW

80

The new `/reframe` command analyzes your project's files and description to surface 8 to 12 expert terms ranked by how much each would change what gets built, with cross-field analogies, named gaps, and a rewritten description; it is now listed in the `/poteto-help` skill table.

— Names the command and its output shape, no example run shown. v0.9.155

06 **Codex Arena schema-aware model dispatch** Not checked yet IMPROVED

78

Skills now pass `model` and `reasoning_effort` to `spawn_agent` only when the exposed schema includes each field, avoiding dispatch errors on Codex's multi-agent v2 schema when either field is dropped. When `model` is unavailable, a diverse-model panel instead varies reasoning effort and notes reduced diversity in its verdict, and Arena runs its cross-judge on Codex only when the judge and provisional winner are known to be different families, per a rule stored in the Codex platform mapping.

— Names schema fields and fallback logic but not full mapping file. v0.9.156

07 **Windows support for open-pstack** Not checked yet IMPROVED

78

`open-pstack` now escapes each argument and the shim path with caret-escaping for `cmd.exe` when launching a `.cmd` or `.bat` shim on Windows, preventing variable names such as `%SystemRoot%` in a plugin path from being expanded. Its end-to-end tests now run on Windows in CI, using `.cmd` harness shims launched through `cmd.exe` and skill links made as junctions.

— Names escaping mechanism and CI setup, no user command. v0.9.152 v0.9.153

---

08 **OpenCode 2 subagent registration and overrides** Not checked yet

NEW

73

The OpenCode 2 plugin now registers four bundled agents as subagents under their Claude Code names (such as `pstack:poteto-agent`), so a skill's `subagent_type` can name an OpenCode agent without entries in `opencode.json`. Users can override bundled OpenCode agent fields by setting the same agent ID in `opencode.json`, which OpenCode applies after plugins.

— Names config file and override mechanism, not the full agent list. v0.9.152

---

09 **Fork-aware SessionStart routing on Codex** Not checked yet IMPROVED

60

The `SessionStart` routing hook on Codex now also runs when a chat is forked, so a forked session loads the routing instruction and honours `session hook: off` the same way a new session does.

— Explains the trigger and config flag, little further mechanism. v0.9.154

THINNER COVERAGE BELOW

---

10 **Git root lookup caching** Not checked yet IMPROVED

55

The extension now asks git for a directory's root once per process instead of on every agent start and every subagent spawn, saving 12.6 ms each on macOS.

— Gives a concrete saving but little mechanism detail. v0.9.152

WAS THIS USEFUL? 👍 👎

The Open Engine Zeroshot

Feedback ⓘ v17.2.0 NOTES ↗

RANK

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

> Zeroshot's v17.2.0 release makes CLI resumes default to checkpoint continuation, lets local scripts drive verifier signals without model calls, and adds typed filepath-validated file storage alongside a breaking bump of its OECP, target discovery and hosted run route APIs.

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

01 Checkpoint-based resume as CLI default

Not checked yet IMPROVED

80

Failed CLI runs now resume automatically from their latest checkpoint instead of restarting from scratch; use `--restart` on `zeroshot resume RUN_ID` to preserve the old full-restart behavior.

Restart a failed run fully from the beginning, bypassing the new checkpoint-resume default.

```
$ zeroshot resume RUN_ID --restart
```

— Exact command and flag given, clear before/after behavior v17.2.0

02 File storage and typed filepath validation

Not checked yet NEW

70

Adds selected-target filesystem contracts, private atomic disk storage, stable pins, ownership hooks, optional capability discovery, and local/direct Docker CLI file operations. Validates typed `filepath` outputs and publishes declared files or folders at explicit human requests and outer graph success, governed by request/run access controls.

— Names typed `filepath` surface and storage mechanisms, no exact command v17.2.0

THINNER COVERAGE BELOW

03 **Contained Windows controller startup** Not checked yet NEW

55

Adds opt-in contained Windows controller startup for the local CLI `run`, `resume`, and `answer` commands and for Python `LocalTarget` callers.

— Names exact CLI commands and Python interface, opt-in mechanism unclear v17.2.0

04 **Runtime configuration replacement on resume** Not checked yet

IMPROVED

55

CLI and Python resumes can now replace exact or uniform runtime plans while preserving the saved graph, input, source, delivery lineage, checkpoint progress, and conversation restoration.

— Explains scope and preserved state, but no concrete command shown v17.2.0

05 **OECP, discovery, and run route API version bump** Not checked yet

55

BREAKING

OECP now uses v3, target discovery v4, and hosted run routes v3; clients and targets must be upgraded together to remain compatible.

— Names exact versioned APIs and upgrade requirement v17.2.0

06 **Script-emitted verifier signals** Not checked yet NEW

50

Instruction-free local scripts can now emit declared verifier signals, route graph choices, and record diagnostics without making any model calls.

— Describes mechanism but no named API or flag to act on v17.2.0

07 **Static Linux builds without libc** Not checked yet IMPROVED

45

Zeroshot is now built for static Linux targets using the kernel rename operation directly, keeping file publication atomic on Linux musl without libc.

— Names build target and mechanism but no reader action v17.2.0

08 **Stable public run ID across resume and suspend** Not checked yet

IMPROVED

40

Resumed, suspended, answered, or re-resumed runs now keep their original public run ID, so the run list shows a single entry and Steps and Graph views replay all parts at the same position.

— Clear behavior change but no named flag or endpoint v17.2.0

└──▷ **BREAKING ON UPGRADE**

! OECP now uses v3, target discovery v4, and hosted run routes v3 — upgrade clients and targets together.

WAS THIS USEFUL? 👍 👎

Amazon Kiro Feedback ⓘ

Models: GPT 6.2 Astra Preview Begins Rolling Out to Kiro Enterprise (2026-11-25)

BLOG ↗ RANK

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

> Kiro expanded Enterprise capabilities this window with GPT 6.2 Astra model access and granular credit-pool usage limits, alongside a cluster of CLI 3.6.x improvements covering headless agent/tool-trust controls, permission approval flow, output display, and guided compaction.

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

01 Credit pooling usage limits and overages Not checked yet NEW

95

Enterprise administrators can cap pooled credits each user may consume per billing cycle with the `credit_pool.usage_limit` control, settable as a profile-level default and overridden at the group or user level, with precedence: user's own limit, then highest group limit, then profile default. Overages can be turned on or off for the whole enterprise account via the `overages` control on the profile.

Set a per-user monthly credit cap at the profile level so every user in the enterprise defaults to that limit — then tighten or raise it for specific groups or individuals.

> 📍 In the Kiro Console, go to Enterprise Management › Controls › Profile, add a control with key `credit_pool.usage_limit` and your chosen credit value. To override for a team, go to Controls › Groups, select the group, and set a different `credit_pool.usage_limit`. To override for one user, go to Controls › Users, select the user, and set their own `credit_pool.usage_limit`.

Prevent users from consuming credits beyond the pool balance by disabling overages at the account level.

> 📍 In the Kiro Console, go to Enterprise Management › Controls › Profile and set the `overages` control to off.

— Named config keys, precedence rules, and exact console steps

General: Credit Pooling Usage Limits an…

02 GPT 6.2 Astra model for Enterprise Not checked yet NEW

90

Kiro adds GPT 6.2 Astra model support for Enterprise organizations, enabled via model governance in the admin console (Settings > Model Governance). The model supports a 1M context window, carries a 7.2x credit multiplier (rising to 14.4x for requests over 272K tokens), and runs inference in the US geography only.

Grant your Kiro Enterprise organization access to GPT 6.2 Astra during the Preview period via the model governance settings in the Kiro admin console.

> 📍 In the Kiro admin console, go to Settings › Model Governance, find GPT 6.2 Astra in the available models list, and enable it for your organization.

— Exact numbers and console path given, but no API/CLI command

Models: GPT 6.2 Astra Preview Begins Ro…

03 **Headless run agent and tool-trust controls** Not checked yet NEW

90

Non-interactive CLI runs can trust specific tools with `--trust-tools=web_search` (a `use_aws` entry in the same flag now warns it has no effect), and can start V3's default or planner agents in `--no-interactive` runs with `--agent kiro_default` or `--agent kiro_planner`. A `chat.defaultAgent` setting that references a missing agent file now warns and continues instead of stopping the run.

Run a V3 headless pipeline using the built-in planner agent without any interactive prompts.

```
$ kiro-cli --no-interactive --agent kiro_planner
```

Restrict a headless run to trust only web search tools, so other tool calls still require approval.

```
$ kiro-cli --no-interactive --trust-tools=web_search
```

Run a headless V3 pipeline and trust web-search tool calls without an approval prompt.

```
$ kiro-cli --no-interactive --agent kiro_default --trust-
tools=web_search
```

— Exact flags and runnable commands for every named control

CLI 3.6.2: Subagent Stop Hooks and Guid… CLI: Subagent Stop Hooks and Guided Com…

04 **V2-to-V3 migration via managed-settings.json** Not checked yet

NEW

90

Administrators can turn off CLI V2 on managed devices by setting `cli_v2.enabled` to `false` in the `settings` object of `managed-settings.json`; interactive launches that would have defaulted to V2 now start V3 with a notice instead.

Disable V2 for all users on a managed device so interactive launches default to V3.

```json
{
  "settings": {
    "cli_v2.enabled": false
  }
}
```

Prevent managed-device users from launching V2 sessions so everyone runs V3.

```json
Set `cli_v2.enabled` to `false` in `managed-settings.json`:
{
  "settings": {
    "cli_v2.enabled": false
  }
}
```

— Exact config file, key and resulting behavior change given

CLI 3.6.2: Subagent Stop Hooks and Guid…

CLI: Subagent Stop Hooks and Guided Com…

05 **CLI output display and guided compaction** Not checked yet IMPROVED

75

The Lean density preset now shows why each tool call runs via tool-call reasons enabled in `/settings` output-details; tool output can be expanded and collapsed with Ctrl+O (Esc no longer collapses it); and one-off context compaction can be directed with the `/compact` command plus custom instructions, e.g. `/compact Keep the migration decisions and remaining validation steps`, to preserve specific context instead of the default summary.

Direct a one-off compaction to preserve specific context rather than using the default summary.

```
$ /compact Keep the migration decisions and remaining validation steps
```

— Exact keybindings and runnable /compact command given

CLI 3.6.2: Subagent Stop Hooks and Guid… CLI: Subagent Stop Hooks and Guided Com…

THINNER COVERAGE BELOW

06 **'Approve for me' one-step permission grant** Not checked yet

IMPROVED

55

In local V3 sessions, permission prompts that previously offered only Allow once now also offer 'Allow once and enable Approve for me', confirming workspace trust in one step instead of two.

— Behavior described but only as UI prompt, no command

CLI 3.6.2: Subagent Stop Hooks and Guid…

WAS THIS USEFUL? 👍 👎

opencode-telegram-bot

Feedback ⓘ v0.29.0 NOTES ↗

RANK

opencode-telegram-bot is a Telegram client for OpenCode that lets you run and monitor AI coding tasks from your phone while they run locally.

> opencode-telegram-bot's v0.29.0 release adds commands to monitor and cancel in-progress searches and context compaction, plus LaTeX math rendering and Polish language support.

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

01 **Abort and status controls for in-progress operations**

Not checked yet NEW 73

Use the `/abort` command to cancel any in-progress search — project, session, model, command, or skill — at any point, whether still entering a query or already viewing results. Separately, `/status` lets you check the status of a running context compaction, cancel it with `/abort`, or use `/detach` to continue working in another session.

— Names exact commands but limited on internal mechanism v0.29.0

THINNER COVERAGE BELOW

02 **Native LaTeX math rendering in replies** Not checked yet NEW

45

opencode-telegram-bot now renders LaTeX equations in assistant replies as native Telegram math blocks, while leaving LaTeX inside code blocks unchanged.

— Describes behaviour but no user action involved v0.29.0

03 **Polish interface language** Not checked yet NEW 35

Users can choose Polish as the interface language by selecting Polski in the setup wizard.

— Thin description, but gives exact UI path v0.29.0

WAS THIS USEFUL? 👍 👎

Tester Army e2e

Feedback ⓘ @e2e-dev/revyl@0.0.2 NOTES ↗

RANK

e2e is an end-to-end testing framework that automates UI tests across web and mobile applications.

> This window adds a new macOS desktop testing engine and a built-in HTTP `request` fixture, replaces the `requires` test option with `platforms` scoping, and introduces WebMCP tool introspection, offline network simulation, and Playwright-style URL-glob routing in the web engine. It also ships automatic Codex sign-in and secret redaction, new mobile device and lease APIs, a `ctrf` reporter, and a `collected` reporter event replacing the deprecated `plan`.

└──▷ **WHAT SHIPPED · 35 FEATURES** most completely described first ? what's the number?

01 **macOS desktop testing engine** Not checked yet NEW 94

The new `@e2e-dev/desktop` package adds engine: desktop() to drive native macOS apps via accessibility commands, including `screen` locators, the `desktop` fixture, masked screenshots, and secret fills over stdin. Run such tests in the foreground with `e2e run --headed`; desktop() returns `DesktopEngine` for type inference, and desktop({ maxDepth }) (1 to 50) controls accessibility tree depth, with `POLICY_DENIED` returned for screenshots that cannot mask a password field or for windows deeper than `maxDepth`.

— Named package, config option with range, and failure codes @e2e-dev/desktop@0.0.1

02 **Playwright-style URL glob routing** Not checked yet BREAKING 91

`browser.route`, `browser.unroute`, and `browser.waitForResponse` now parse string patterns as Playwright URL globs: `?` is literal, `{a,b}` alternates instead of matching literal text, `''` matches every URL, a relative pattern resolves against the app URL, and a pattern ending in `**` right after the host matches only URLs ending in `/`; escape literal `{`, `}`, or `*` with `\`, or pass a `RegExp`. `browser.waitForURL` now also accepts Playwright URL globs such as `'**/checkout'`, and waitForURL('') now matches any URL instead of the app URL. Patterns that previously matched specific requests may now match different or more requests.

— Full before/after glob semantics with escape guidance @e2e-dev/web@0.17.0

03 **Secret redaction in Codex integration** Not checked yet NEW 91

Redact API keys and secrets across terminal output, `report.json`, `summary.md`, trace pages, and model inputs using `registerProviderKey`,

`registerProviderKeys`, `redactSecrets`, and `redactStepText` from `e2e/agent` in an executor or config — covering `OPENAI_API_KEY`, `ANTHROPIC_API_KEY`, `AI_GATEWAY_API_KEY`, `OPENROUTER_API_KEY`, other model-provider variables, subscription login tokens, and the Codex login in `CODEX_HOME/auth.json`, for values 20 to 8,192 characters, rendered as `<secret:NAME>`. Redaction also covers whole UTF-8 text downloads up to 16 MiB (labelled `incomplete` when it can't complete), and errors with code `MODEL_PROVIDER_FAILED`, `CONTEXT_OVERFLOW`, or `MODEL_OUTPUT_INVALID` now have their text redacted and omit the raw `cause` when caught from a step.

— Every env var, function, and limit named verbatim `@e2e-dev/acp@0.1.4`

---

04 **Grant-all permissions before launch** Not checked yet NEW

88

Grant every permission an engine can grant before each launch with `app.permissions: 'all'` in config, or for a single launch with device.openApp(app, { relaunch: true, permissions: 'all' }). Engines must declare `grantsAllPermissions: true` to support this — engines released before this change, including older `@e2e-dev/web` or `@e2e-dev/mobile` releases, reject the option at config load with `INVALID_CONFIG`.

Grant every permission automatically before each fresh app launch, so tests do not stall on permission dialogs.

typescript

```
import { test } from 'e2e';

test('onboarding skips all permission prompts', async ({ app,
device }) => {
  await device.openApp(app, { relaunch: true, permissions: 'all'
});
  // test continues with all permissions already granted
});
```

Grant all runtime permissions automatically before every app launch so permission prompts never interrupt a mobile test suite.

```typescript
export default defineConfig({
  projects: [{
    name: 'ios',
    app: {
      bundleId: 'com.example.MyApp',
      permissions: 'all'
    }
  }]
});
```

Grant all permissions automatically before each launch on iOS, Android, and web so tests do not need to handle permission dialogs.

```yaml
app:
  permissions: 'all'
```

— Config key, method signature, and compatibility rule named `@e2e-dev/web@0.17.0` `@e2e-dev/mobile@0.12.9` `e2e@0.25.0` `@e2e-dev/core@0.25.0`

---

05 **File-URL app targets across web, mobile, and desktop**

Not checked yet NEW 86

Point `app.url` at a `file:` URL to test a local HTML file with no server, across web, mobile, and desktop engines, e.g. `app: { url: pathToFileURL('dist/index.html').href }`. On web, `browser.goto` and agent `navigate` are restricted to files in the same directory or below it; on mobile, navigation outside that directory returns `POLICY_DENIED`.

Run tests against a local build artifact with no dev server, restricting file access to the build directory.

```typescript
import { pathToFileURL } from 'url';

export default defineConfig({
  app: {
    url: pathToFileURL('dist/index.html').href,
  },
  engine: web(),
});
```

Open a local built HTML file with no dev server — useful for testing a production build of a static site in CI.

```typescript
import { pathToFileURL } from 'url';

// e2e.config.ts
export default {
  app: { url: pathToFileURL('dist/index.html').href },
};
```

— Exact config key and restriction behaviour for three engines `@e2e-dev/web@0.17.0` `@e2e-dev/mobile@0.12.9` `@e2e-dev/desktop@0.0.1`

06 **Typed launch arguments** Not checked yet NEW 86

Pass launch arguments as a typed object to `app.launchArguments` or device.openApp({ launchArguments }), mapping names to string, number, boolean, or list of strings — e.g. `{ AppleLanguages: ['fr'], AppleLocale: 'fr_FR' }` for iOS locale or typed Android `am start` extras — without leading dashes; a dashed key fails with `INVALID_CONFIG` or `INVALID_ARGUMENT`. `@e2e-dev/core/engine` now exports `checkLaunchArguments` and `appleUserDefaults` for engine authors, and engines declaring `launchArgumentValues: true` receive the object in `EngineAppInfo.launchArgumentValues`.

— Full object shape and engine-author exports named `@e2e-dev/mobile@0.12.9` `@e2e-dev/desktop@0.0.1`

07 **platforms option replaces requires** Not checked yet BREAKING 86

Tests can be scoped to platforms with the `platforms` option — e.g. `{ platforms: ['web'] }` or `{ platforms: ['ios', 'android'] }` — replacing the removed `requires` option; any test, group, or setup still using `requires` now fails at collection, and the Capability type and `capability-unavailable` skip cause are gone. Fixtures are now typed with the config's type as a third type argument via test.extend<{}, {}, typeof config>() with `as const satisfies E2EConfig`, and `platforms` on a test scopes fixture types to the targets that run it.

Scope a test to run only on iOS and Android, skipping web runs entirely.

```typescript
import { test, expect } from 'e2e';

test('camera permission granted', { platforms: ['ios', 'android']
}, async ({ agent }) => {
  await agent.assert('the camera permission prompt is not
visible');
});
```

— Exact replacement mapping and typing pattern given `@e2e-dev/desktop@0.0.1` `e2e@0.25.0` `@e2e-dev/core@0.25.0`

08 **WebMCP tool introspection** Not checked yet NEW 84

Call WebMCP tools from test code with browser.webmcp.tools() and browser.webmcp.callTool(name, input), enabled by setting web({ webmcp: true }) on Chromium or Firefox — useful for verifying AI-agent-facing surfaces in isolation with no model involved.

Inspect and call tools a page exposes via WebMCP from test code, with no model involved — useful for verifying AI-agent-facing surfaces in isolation.

```typescript
// e2e.config.ts
export default defineConfig({
  engine: web({ webmcp: true }),
});

// test
const tools = await browser.webmcp.tools();
const result = await browser.webmcp.callTool('summarise', { text:
'hello' });
```

— Exact API and config with runnable example `@e2e-dev/web@0.17.0`

09 **New test modifiers: todo, fixme, describe.only, describe.fail**

Not checked yet NEW 84

Register an unfinished test placeholder that never runs and reports as skipped with test.todo(title) (report shows `skip.cause` as `todo`); skip tests at registration or from inside their body with `test.fixme` (`skip.cause` as `fixme` or `fixme: <reason>`); focus every test in a group with test.describe.only(title, fn) or the `only` option on `describe` (CI rejects a focused group with `ONLY_IN_CI`); and mark every test in a group as expected to fail with test.describe.fail(title, fn) or the `fail` option on `describe`. `report-1` reports can now contain `skip.cause` values `fixme` and

`todo`, so pinned copies of `schema/report-v1.schema.json` must be updated and exhaustive skip-cause checks must handle both.

— Four named APIs with exact skip-cause values and schema impact `e2e@0.25.0` `@e2e-dev/core@0.25.0`

10 **`ctrf test report format`** `Not checked yet` NEW 82

Add the `ctrf` reporter with `--reporter ctrf` or `reporters: ['ctrf']` to write `<output>/ctrf.json` in Common Test Report Format 0.3.0, including flaky flags, earlier attempts, steps, and attachments.

Export test results to any CTRF-compatible dashboard alongside the default reporter.

```
$ npx e2e run --reporter ctrf,list
```

— Exact flag, output path, and format version with example `e2e@0.25.0`

11 **Reporter factories and custom reporter API** `Not checked yet` NEW 82

Call reporter factories from `e2e/reporters` — list(), dot(), json(), junit({ outputFile }), markdown({ outputFile }) — to pass options to built-in reporters. Implement custom reporters against the public `Reporter` contract using `onEvent` for live events, `reportFiles` for atomic file writes, `onRunFinished` for upload or notify, and onSummary(rows, run) for stdout output; mark the terminal reporter with `interactive: true`.

— Every factory and contract hook named verbatim `e2e@0.25.0` `@e2e-dev/core@0.25.0`

12 **`toBeOK()` HTTP response matcher** `Not checked yet` NEW 82

Assert HTTP responses with expect(response).toBeOK(), which passes for status 200–299 and works with `.not`, `expect.soft`, and `expect.poll`; failures name the status and origin path, never credentials, query, or fragment.

Assert that an API call returns a successful status in a test that uses the built-in `request` fixture.

```typescript
import { test, expect } from 'e2e';

test('creating an order succeeds', async ({ request }) => {
  const response = await request.post('/api/orders', { json: {
sku: 'A1' } });
  await expect(response).toBeOK();
});
```

— Exact matcher, status range, and security note with example `@e2e-dev/core@0.25.0`

13 **GitHub reporter additions: sharding, reruns, and step keys**

Not checked yet NEW 80

Sharded runs now record shard identity as `run.shard { index, total }` and mark each test handed to another shard with `skip.otherShard: true` in the report. A `--last-failed` rerun writes `report-with-reruns.json` alongside the report, merging the first run and every subsequent rerun into one document. Earlier runs are accessible via the new `foldLastRun` export from `e2e`, which reporters receive as `lastRun`. Interactive step keys `s` (screen) and `l` (app log) now appear below the live summary during `--step` and screen.pause().

— Four named report fields/exports grouped from one package `@e2e-dev/github@0.4.7`

14 **Built-in request fixture for HTTP calls** Not checked yet NEW

80

Tests now get a built-in `request` fixture for HTTP calls — e.g. await request.post('/api/orders', { json: { sku: 'A1' } }) — which sends to a path on the app URL or an absolute URL, appears in the report as a step, and returns a `WebResponse`.

Make an authenticated HTTP call inside a test without touching the browser's cookies — useful for seeding data or asserting API state alongside UI assertions.

typescript

```
import { test, expect } from 'e2e';

test('order appears after API seed', async ({ app, request,
screen }) => {
  await request.post('/api/orders', { json: { sku: 'A1' } });
  await app.open('/orders');
  await expect(screen.getByText('A1')).toBeVisible();
});
```

Make an authenticated API call mid-test and verify its effect in the UI, without touching the browser's cookie jar.

typescript

```
test('order appears after API call', async ({ request, screen })
=> {
  const response = await request.post('/api/orders', { json: {
sku: 'A1' } });
  await expect(screen.getByRole('row', { name: 'A1'
})).toBeVisible();
});
```

— Exact fixture usage with runnable example `@e2e-dev/desktop@0.0.1`

15 **Automatic local build installation for Revyl and EAS**

Not checked yet NEW 78

Testing a local build no longer requires manual installation: omitting `app` from revyl() uploads and installs the target's `app.appPath` on each device, and expoSimulators() does the same automatically before the first test when no `buildId` or `applicationArchiveUrl` is provided (requires `@e2e-dev/mobile` 0.12.9 or later), removing the need to call device.installApp() manually.

Let expoSimulators() handle local build installation automatically, skipping the manual device.installApp() call that was previously required.

typescript

```
// e2e.config.ts
import { expoSimulators } from '@e2e-dev/eas';

export default {
  target: expoSimulators({
    app: { appPath: './build/MyApp.app' }
    // no buildId or applicationArchiveUrl — the build is
uploaded and
    // installed on each simulator automatically before the first
test
  })
};
```

— Names config pattern and version requirement, lacks limits `@e2e-dev/eas@0.3.6` `@e2e-dev/revyl@0.0.2`

16 **e2e doctor browser dependency check** Not checked yet IMPROVED 78

`e2e doctor` now fails the `browser` check on Linux when installed browsers have unresolved system libraries, naming the missing libraries and suggesting `npx @e2e-dev/web install <name> --with-deps` or `pnpm exec e2e-web install <name> --with-deps`.

— Exact remediation commands given `@e2e-dev/web@0.17.0`

17 **Mobile simulator session isolation** Not checked yet BREAKING 78

`@e2e-dev/mobile` now requires `@e2e-dev/core` 0.24.1 or newer; two mobile() hosts from launch(), or a host and the test's own device, no longer share one simulator

session — each gets its own session named `e2e-<target>-<token>-<slot>`, and a host that finds no free device fails instead of sharing one.

— Exact session naming scheme and version requirement given `@e2e-dev/web@0.17.0` `@e2e-dev/mobile@0.12.9`

18 **Foldable device pose control for agents** Not checked yet NEW

78

Give an agent fold/unfold control over foldable devices by spreading foldTools(...engines) from `@e2e-dev/mobile/tools` into the agent's `tools`; the tool puts the device in `closed`, `half-open`, or `open` pose and works with `agent.act`, `e2e explore`, and the MCP server's `call`.

— Exact import path, poses, and integration points named `@e2e-dev/mobile@0.12.9`

19 **Automatic app installation for leased devices** Not checked yet

NEW

76

Install a leased app before tests run with installLeasedApp(request, lease) exported from `@e2e-dev/mobile/lease`, which installs `request.appPath` through the lease's daemon and returns the lease with `installedApp`. agentDeviceLeases() gains an `installAppPath` option to install the request's `appPath` automatically through each lease's daemon once a lease is granted, releasing the lease automatically on a failed install.

— Named function and option with failure behaviour `@e2e-dev/mobile@0.12.9`

20 **Reporter event stream: collected replaces plan** Not checked yet

DEPRECATED

74

The `plan` run event is deprecated in favor of `collected`, which follows it with the same `total` and `files` fields; `plan` will be removed no earlier than April 1, 2027. Reporters can also subscribe to `paused` and `resumed` (when a test waits at a `--step` prompt or screen.pause()), `output` (what the runner process prints, with no `target`), and a step `start` event that now includes `source` (the file line that called it).

— Names every new event and the deprecation timeline `@e2e-dev/html-reporter@0.0.3` `e2e@0.25.0` `@e2e-dev/core@0.25.0`

21 **Automatic Codex sign-in via API key** Not checked yet NEW

74

codex() now signs Codex in automatically using `CODEX_API_KEY` or `OPENAI_API_KEY` (first nonblank wins) when Codex has no saved login, instead of failing with 'the agent is not signed in', and saves the key via `codex login --with-api-key` unless the credential store is `ephemeral`.

– Exact env vars and fallback command named `@e2e-dev/acp@0.1.4`

22 **Offline network simulation** Not checked yet NEW

71

Simulate offline and online network conditions in a test with browser.setNetwork('offline') and browser.setNetwork('online'), letting tests assert on an app's offline state including service-worker-cached pages that a simple request-abort route cannot exercise.

Test how your app behaves when the network drops mid-session, including service-worker-cached pages that a simple request-abort route cannot exercise.

typescript

```
await browser.setNetwork('offline');
// assert the app shows an offline state
await browser.setNetwork('online');
```

– Exact method shown with example, little extra mechanism `@e2e-dev/web@0.17.0`

23 **Device inspection and interaction methods** Not checked yet NEW

70

device.isKeyboardShown() returns true or false from agent-device reports and the accessibility tree without changing device state. device.shake() shakes a local iOS simulator or Android emulator once, recorded as a `device.shake` step; physical devices and provider-backed devices return `UNSUPPORTED_CAPABILITY`.

– Two exact methods and limits named, no example `@e2e-dev/mobile@0.12.9`

24 **Full JavaScript execution in run_code MCP tool** Not checked yet

70

BREAKING

Run full JavaScript in `e2e mcp`'s `run_code` tool, with variables, loops, `try` / `catch`, return values, and `console.log`; `open_session` lists the API. The result format changed: it now shows what the code returned, then each line that ran; statements no longer have restrictions on secret fields or locators, and lines need `await` or they start together and no longer run in order.

– Capability and result-format change described, no example `e2e@0.25.0`

25 **Empty-value assertion for locators** Not checked yet NEW

62

Assert that an input or element is empty with expect(locator).toBeEmpty(), following Playwright semantics for `input` and `textarea` values.

Assert that a form field is cleared after the user submits, without writing a custom value check.

```typescript
import { test, expect } from 'e2e';

test('search field clears after submit', async ({ app, screen })
=> {
  await app.open('/search');
  const field = screen.getByRole('searchbox');
  await field.fill('typescript');
  await screen.getByRole('button', { name: 'Search' }).click();
  await expect(field).toBeEmpty();
});
```

— Exact matcher and example, no edge cases detailed `@e2e-dev/web@0.17.0`

26 **Built-in request fixture naming conflict** Not checked yet BREAKING

62

A project fixture already named `request` now extends the built-in one; a fixture declared with its own type through test.extend<{ request: … }>({ request: … }) no longer compiles and must be renamed. An engine can no longer contribute a fixture named `request`.

— Exact migration step (rename) given `@e2e-dev/desktop@0.0.1` `e2e@0.25.0` `@e2e-dev/core@0.25.0`

THINNER COVERAGE BELOW

27 **agent-device close command** Not checked yet NEW

59

Run `agent-device close --session <name>` to free a device left held after a host exits without calling close().

— Runnable command but minimal context `@e2e-dev/mobile@0.12.9`

28 **Chromium OS sandbox default change** Not checked yet IMPROVED

56

Chromium now launches with its OS sandbox on by default on macOS and Windows; enable it on Linux with web({ chromiumSandbox }).

– Exact config key and platform scope named `@e2e-dev/web@0.17.0`

29 **Pause-aware run duration field** Not checked yet NEW

56

Reports now include a `run.durationMs` field, and time spent while every worker is paused at a `--step` prompt or screen.pause() is excluded from both the terminal's Duration row and that field.

– Exact field name and exclusion rule, no example `@e2e-dev/html-reporter@0.0.3` `@e2e-dev/github@0.4.7`

30 **Reporter output moves to stderr** Not checked yet BREAKING

55

Without `list`, `dot`, or another interactive reporter, output your code or config prints now goes to stderr instead of stdout; config-load output also moves to stderr, so stdout now holds only what a reporter explicitly prints there.

– Clear before/after but no migration command `@e2e-dev/html-reporter@0.0.3` `e2e@0.25.0` `@e2e-dev/core@0.25.0`

31 **In-memory file upload for inputs** Not checked yet NEW

54

Upload in-memory file payloads via locator.setInputFiles({ name, mimeType, buffer }), including payload arrays.

– Exact signature named but no example `@e2e-dev/web@0.17.0`

32 **Cookie filtering by request URL** Not checked yet NEW

53

Filter browser.cookies(urls) to only the cookies a request to one or more URLs would carry, matched by domain, path, and Secure.

– States mechanism but no runnable example `@e2e-dev/web@0.17.0`

33 **Fixture scope/auto inheritance change** Not checked yet BREAKING

53

A fixture that extends another (such as an override of `device`) no longer inherits `scope` or `auto` from the parent fixture — both now come from its own options.

– States exact behaviour change, no migration example `e2e@0.25.0`

34 **Engine fixtures undefined for unsupported targets** Not checked yet

IMPROVED

46

Engine fixtures are now `undefined` on targets whose engine does not provide them, so a test using `device` can run on web without a `COLLECTION_ERROR`.

— Behaviour change explained, nothing to configure `@e2e-dev/web@0.17.0`

---

35 **Original casing preserved in HTML report** Not checked yet IMPROVED

42

The HTML report now preserves the original casing of hook step tags, test tags, session names, skip reasons, screen URLs, and token counts instead of lowercasing them.

— Lists affected fields, no actionable surface `@e2e-dev/html-reporter@0.0.3`

└──▷ BREAKING ON UPGRADE

! Remove any fixture that calls device.installApp() for a build managed by expoSimulators() — without `buildId` or `applicationArchiveUrl`, it will now upload and install the build a second time.

! `browser.route`, `browser.unroute`, and `browser.waitForResponse` now parse string patterns as Playwright URL globs: `?` is literal (not a single-character wildcard), `{a,b}` alternates instead of matching literal text, `''` matches every URL, a relative pattern resolves against the app URL, and a pattern ending in `**` right after the host matches only URLs ending in `/`. Escape literal `{`, `}`, or `*` with `\`, or pass a `RegExp`. Patterns that previously matched specific requests may now match different or more requests.

! `browser.waitForURL` now accepts Playwright URL globs such as `'**/checkout'`; waitForURL('') now matches any URL instead of the app URL, and `{`, `}`, and `\` are glob syntax rather than literals.

! Engines released before this version do not declare `grantsAllPermissions: true`, so using `app.permissions: 'all'` with `@e2e-dev/web` or `@e2e-dev/mobile` releases predating this change fails at config load with `INVALID_CONFIG`.

! `@e2e-dev/mobile` now requires `@e2e-dev/core` 0.24.1 or newer; two mobile() hosts from launch() or a host and the test's own device no longer share one simulator session — each gets its own session named `e2e-<target>-<token>-<slot>`, and a host that finds no free device fails instead of sharing one.

! An engine released before this change that does not declare `grantsAllPermissions: true` rejects `app.permissions: 'all'` at config load with `INVALID_CONFIG`; this includes `@e2e-dev/web` and `@e2e-dev/mobile` released before this version.

! `@e2e-dev/mobile` now requires `@e2e-dev/core` 0.24.1 or newer; a host that finds no free device now fails instead of sharing one.

! Without `list`, `dot`, or another interactive reporter, output your code prints now goes to stderr instead of stdout; config-load output also moves to stderr. Stdout now holds only what a reporter explicitly prints there.

! The `plan` run event is deprecated; read `collected` instead, which follows it with the same `total` and `files` fields. `plan` will be removed no earlier than April 1, 2027.

! `@e2e-dev/github` now requires `e2e` 0.25.0 or later.

! A project fixture already named `request` now extends the built-in one; a fixture declared with its own type through test.extend<{ request: ... }>({ request: ... }) no longer compiles — rename it. An engine can no longer contribute a fixture named `request`.

! `@e2e-dev/acp` now requires `e2e` 0.25.0 or newer.

! A fixture that extends another (such as an override of `device`) no longer inherits `scope` or `auto` from the parent fixture — both now come from its own options.

! The `run_code` result format changed: it now shows what the code returned, then each line that ran. Statements no longer have restrictions on secret fields or locators; lines need `await` or they start together and no longer run in order.

! The `requires` test option is removed; any test, group, or setup that still sets `requires` fails at collection. Replace `{ requires: ['browser'] }` with `{ platforms: ['web'] }` and `{ requires: ['device'] }` with `{ platforms: ['ios', 'android'] }`. The `Capability` type and `capability-unavailable` skip cause are gone.

! Without `list`, `dot`, or another interactive reporter, code your tests print now goes to stderr instead of stdout. Stdout holds only what a reporter prints there.

! The `plan` run event is deprecated; read `collected` instead, which follows it with the same `total` and `files`. `plan` will be removed no earlier than April 1, 2027.

! `report-1` reports can now contain `skip.cause` value `fixme`; update pinned copies of `schema/report-v1.schema.json` and handle `fixme` in exhaustive skip-cause checks.

! `report-1` reports can now contain `skip.cause` value `todo`; update pinned copies of `schema/report-v1.schema.json` and handle `todo` in exhaustive skip-cause checks.

! The `requires` test option is removed; any test, group, or setup still using `requires` fails at collection. Replace `{ requires: ['browser'] }` with `{ platforms: ['web'] }` and `{ requires: ['device'] }` with `{ platforms: ['ios', 'android'] }`. The `Capability` type and the `capability-unavailable` skip cause in reports are gone.

! Without `list`, `dot`, or another interactive reporter, output your code or config prints now goes to stderr instead of stdout. Stdout now holds only what a reporter explicitly prints there.

! A project fixture already named `request` now extends the built-in `request` fixture. Any fixture declared with its own type via test.extend<{ request: ... }>({ request: ... }) no longer compiles — rename it. An engine can no longer contribute a fixture named `request`.

! `report-1` reports can now contain `skip.cause` values `fixme` and `todo`; pinned copies of `schema/report-v1.schema.json` must be updated from this release, and exhaustive skip-cause checks must handle both new values.

WAS THIS USEFUL? 👍 👎

Zed

Feedback ⓘ v1.30.2 NOTES ↗ RANK

Zed is an open-source code editor that runs collaborative development workflows.

> Zed's latest release adds a dedicated AI model setting for commit messages, a title bar visibility setting, a new CLI debug flag, and keyboard navigation for the agent thread switcher, alongside smaller editor and Vim keybinding additions.

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

01 **Title bar visibility setting** Not checked yet NEW 75

The new `title_bar.show` setting hides the title bar always or only in fullscreen, accepting `"never"`, `"hide_in_full_screen"`, or `"always"` (default).

Hide the title bar when working in fullscreen to maximise editor space.

json

```
{
  "title_bar": {
    "show": "hide_in_full_screen"
  }
}
```

— Named setting with all accepted values and a runnable config example. v1.30.2

02 **Model picker for AI commit messages** Not checked yet NEW 70

A new `agent.commit_message_model` setting lets you save your preferred AI model for commit message generation, selectable from a model picker now shown in the Git Panel's commit editor.

Pick a specific model for AI commit message generation and persist it across sessions.

📍 In the Git Panel, open the commit editor, click the model picker, select your preferred model — Zed saves the choice to `agent.commit_message_model` in your settings.

— Named config key and a usage walkthrough, but no mechanism detail. v1.30.2

03 **`--debug-state` CLI flag** Not checked yet NEW 65

The new `zed --debug-state` CLI flag dumps the running app's window, display, and platform state as JSON for debugging.

Print the running Zed instance's window and display state as JSON to diagnose multi-monitor layout issues.

```
$ zed --debug-state
```

— Named flag with a runnable command, limited explanation of output format. v1.30.2

---

04 **Keyboard navigation in agent thread switcher** Not checked yet

NEW

65

The agent thread switcher — opened with `ctrl-tab` from the Agent Panel or Threads Sidebar — now supports keyboard navigation via `up` / `down`, `ctrl-n` / `ctrl-p`, `home` / `end`, and `ctrl-up` / `ctrl-down`.

— Names all keybindings and entry point, no runnable example provided. v1.30.2

THINNER COVERAGE BELOW

05 **New Vim keybindings matching Neovim defaults** Not checked yet

NEW

55

Adds the `g r t` Vim keybinding for `editor: go to type definition`, and `] n` / `[ n` Vim keybindings in visual mode for `vim: select next syntax node` and `vim: select previous syntax node`, matching Neovim's defaults.

— Names all three keybindings and their mapped commands. v1.30.2

---

06 **Fold at current indentation level** Not checked yet NEW

40

Adds the `editor: fold at current indentation level` command to fold the whole document at the current line's indentation level.

— Named command but no invocation path or example given. v1.30.2

---

07 **Extra High thinking effort for Gemini 4 Hydrogen** Not checked yet

NEW

40

Adds support for the Extra High thinking effort level for Gemini 4 Hydrogen (`gemini-4-hydrogen`), which now defaults to High.

— Names model and effort level but no usage steps. v1.30.2

08 **GitHub Copilot context window and pricing display** Not checked yet

IMPROVED

40

GitHub Copilot models now use their full context window when there is no long-context surcharge, and display per-token pricing where available.

— Describes behavior change but no concrete numbers or UI path. v1.30.2

09 **Which-key popup in open dialogs** Not checked yet IMPROVED

30

Now shows the which-key popup inside open dialogs such as Go to Line and the commit editor.

— Thin one-line description with no further mechanism. v1.30.2

WAS THIS USEFUL? 👍 👎

OpenQodex

Feedback ⓘ openqodex@0.25.0 NOTES ↗ RANK

OpenQodex is an open source AI code review for coding agents. A separate reviewer checks every change your agent writes before you push, in Claude Code, Codex, Cursor and Cline.

> OpenQodex now gates agent-driven pre-push reviews behind an explicit per-repository opt-in, and adds project-level reviewer fallback chains, anonymous telemetry controls, and a GitHub Action update across two releases.

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

01 **Project-level reviewer fallback chains** Not checked yet BREAKING

95

Define an ordered fallback chain of reviewers in `.openqodex/config.yaml` under `review.reviewers`, listing each entry as a tool (`claude` or `codex`) with an optional model and effort; the next entry is tried automatically when one hits a usage or rate limit, missing model, or cannot start. Reviewer settings move from personal to project scope: `review.reviewer_web` in `.openqodex/config.yaml` replaces the personal `reviewer_web` key, and the personal `reviewer` , `reviewer_model` , and `reviewer_effort` keys in `~/.openqodex/config.yaml` are no longer read. Pick a single reviewer for one run with `--reviewer <tool>:<model>` and `--reviewer-effort` ; the `--reviewer-model` flag is removed. `init` writes a `review.reviewers` list with `min_version: 0.24.0` when the project has none.

Define a fallback reviewer chain so the team automatically retries with Codex when Claude hits a rate limit.

yaml

```yaml
review:
  reviewer_web: on
  reviewers:
    - tool: claude
      model: claude-opus-4-5
      effort: high
    - tool: codex
      effort: medium
min_version: 0.24.0
```

Override the project reviewer list for a single run, targeting a specific model, without touching shared config.

📍 openqodex review `--reviewer` claude:claude-opus-4-5 `--reviewer-effort` high

– Multiple named config keys and flags with runnable examples and migration path

openqodex@0.24.0

02 **Telemetry opt-out and inspection** Not checked yet NEW

90

Disable anonymous usage telemetry by setting `openqodex_telemetry: 0` in `.openqodex/config.yaml`, or by setting the environment variable `OPENQODEX_TELEMETRY=0` or `DO_NOT_TRACK=1`; telemetry is always off in CI and with `--offline`. Inspect every field sent by telemetry with `openqodex guide telemetry`.

Disable telemetry for a repository so no usage data is sent from that project.

```yaml
openqodex_telemetry: 0
```

– Exact config key, two env vars, and inspection command given openqodex@0.25.0

03 **Per-repository opt-in for agent push reviews** Not checked yet

85

BREAKING

Repositories must now opt in to agent-driven pre-push reviews by adding `.openqodex/config.yaml` or `.openqodex.yaml` at the repo root and running `openqodex init`; push hooks for Claude Code and Codex only act in opted-in repositories. Previously hooks ran in any repository.

– Names exact config files and init command, clear opt-in requirement openqodex@0.25.0

04 **Interactive reviewer list in init** Not checked yet NEW

60

`openqodex init` now shows the planned reviewer list and prompts you to keep it, reorder it, drop reviewers, or write no list; pass `--yes` to skip the prompt and write the list automatically.

Opt a repository in to pre-push agent reviews and pin the reviewer list for the whole team.

📍 openqodex init

– Flag and prompt behavior given, but no output example shown openqodex@0.25.0

THINNER COVERAGE BELOW

05 **GitHub Action reads config from base commit** Not checked yet

IMPROVED

45

The GitHub Action now reads reviewer config from the pull request's base commit rather than the base branch tip, and can review with Codex when the runner is logged in.

— Describes behavior change but gives no command or config to act on `openqodex@0.24.0`

└──▷ BREAKING ON UPGRADE

! Agent push hooks (Claude Code and Codex) no longer run in repositories without a `.openqodex/config.yaml` or `.openqodex.yaml` settings file — run `openqodex init` in each repository you want reviewed to opt it in.

! The personal `reviewer`, `reviewer_model`, and `reviewer_effort` keys in `~/.openqodex/config.yaml` are no longer read; move them into the project's `.openqodex/config.yaml` under `review.reviewers` and `review.reviewer_web`.

! The `--reviewer-model` CLI flag is removed; use `--reviewer <tool>:<model>` instead.

! A review made under a different reviewer list no longer satisfies push hooks.

WAS THIS USEFUL? 👍 👎

GitHub Copilot CLI

Feedback ⓘ v1.0.116 NOTES ↗ RANK

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

> GitHub Copilot CLI shipped a process-wide `env` setting, new Windows and macOS sandbox policy controls, and expanded custom/BYOK provider model management in `/model`, alongside fixes to `--available-tools`, `--max-ai-credits` on resumed sessions, and MCP server config loading and authentication.

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

01 Sandbox policy controls for Windows and macOS

Not checked yet

NEW

90

Windows sandbox package queries are enabled by default; set `sandbox.userPolicy.capabilities` to `[]` to opt out. On macOS, `sandbox.userPolicy.seatbelt.extraMachLookups` allows specific Mach service lookups in sandboxed commands without enabling full Keychain access.

Opt out of Windows sandbox package queries when your policy requires a stricter baseline

```yaml
# In your sandbox user policy config
sandbox:
  userPolicy:
    capabilities: []
```

Allow a specific macOS Mach service in sandboxed commands without granting full Keychain access

```yaml
# In your sandbox user policy config
sandbox:
  userPolicy:
    seatbelt:
      extraMachLookups:
        - com.apple.security.cryptexd
```

— Exact config keys with runnable YAML examples for both platforms v1.0.115

02 `--max-ai-credits` applies on resumed sessions

Not checked yet

IMPROVED

80

Setting `--max-ai-credits` together with `--resume` or `--continue` now applies the credit limit to the resumed session, fixing a case where the value was previously ignored.

Resume a session with a hard AI-credit cap so long-running agents cannot exceed your budget

```
$ copilot --resume --max-ai-credits 500
```

— Runnable command example with exact flags v1.0.115

---

03 **MCP server config loading and authentication** Not checked yet

IMPROVED

78

Workspace MCP servers now load from `.mcp.json` and `.github/mcp.json` even when those files contain comments or trailing commas, and `/mcp auth <name>` signs in to organization-managed MCP servers; startup now summarizes sign-in needs instead of showing individual warnings.

Sign in to an organization-managed MCP server that needs authentication before its tools are available in the session.

```
$ /mcp auth <name>
```

— Names both config files and the auth command with an example v1.0.116

---

04 **Single-run plugin language servers via `--plugin-dir`**

Not checked yet NEW

75

`--plugin-dir` loads a plugin's language servers for a single run; the loaded servers are visible in `/lsp list` but not in `copilot lsp list`.

— Exact flag and command distinction named v1.0.115

---

05 **Custom provider model management in `/model`** Not checked yet

NEW

73

Add and optionally switch to custom provider models in local sessions with Ctrl+X then A in `/model`, and manage configured provider models directly from `/model`. Per-model vision defaults for bring-your-own-key models can be overridden with `COPILOT_PROVIDER_VISION=true` or `false`.

— Names the keystroke, command, and env var but no full walkthrough v1.0.115

---

06 **Extension log routing via `--log-dir`** Not checked yet NEW

70

The `--log-dir` flag routes extension process logs to the same directory as the CLI process log, consolidating logs for post-incident review.

Collect all extension and CLI process logs in one directory to simplify post-incident review.

```
$ copilot --log-dir /tmp/copilot-logs
```

— Runnable flag example included v1.0.116

07 **`--available-tools` empty-value fix** Not checked yet IMPROVED

70

Passing an empty `--available-tools=` or `--available-tools ''` now gives the model no tools, matching the flag's documented help text, instead of granting every tool as before.

— Exact flag and before/after behavior named v1.0.115

08 **Process-wide environment variables via `env` setting**

Not checked yet NEW

65

A new `env` setting lets you set environment variables for every process Copilot starts — shell commands, local LSP and MCP servers, hooks, and extensions — with or without the sandbox enabled.

— Names the setting and its scope but gives no file path or example v1.0.116

09 **`--effort none` honored on first GPT-5.4 request** Not checked yet

IMPROVED

65

Copilot CLI now honors `--effort none` on the first request with GPT-5.4 instead of falling back to medium reasoning effort.

— Names exact flag, model, and prior fallback behavior v1.0.116

10 **Fleet mode entry moved to `/fleet` command** Not checked yet

60

BREAKING

Fleet mode now starts with `/fleet` after plan approval; plan approval no longer offers the fleet option directly, so workflows relying on that prompt must switch to the explicit command.

— Names the exact replacement command and the removed behavior v1.0.116

11 **Persistent init-suggestion suppression** Not checked yet NEW

60

`/init suppress` keeps init suggestions hidden across CLI sessions that share the same profile.

– Exact command given but mechanism left thin v1.0.116

THINNER COVERAGE BELOW

12 **Sub-agent event timestamps in session log** Not checked yet

IMPROVED

50

Sub-agent events in a session's event log now record when they occurred in the `occurredAt` field, not only when they reached the main session.

– Names the field but shows no usage path v1.0.116

13 **ACP editor integration: context size and clarification forms**

Not checked yet NEW

50

In ACP-compatible editors, you can control context size for supported models without changing saved defaults, and add structured clarification questions in ACP editors that support forms.

– Two thin ACP additions with no exact setting names v1.0.116 v1.0.115

14 **Timeline timestamps via `showTimelineTimestamps`** Not checked yet

NEW

50

The `showTimelineTimestamps` setting shows start times, end times, and elapsed durations on timeline items.

– Names the setting but gives no file location v1.0.115

15 **Enabled/Disabled controls for experimental extensions**

Not checked yet IMPROVED

45

Experimental extensions now use Enabled/Disabled controls; existing Load Only settings migrate to disabled with a one-time notice.

– Describes migration but no UI path or config key v1.0.116

└──▷ **BREAKING ON UPGRADE**

! Plan approval no longer offers the fleet option; use `/fleet` explicitly to enter fleet mode.

WAS THIS USEFUL? 👍 👎

pairlet

Feedback ⓘ v2.6.6 NOTES ↗ RANK

Pairlet (formerly CC Pocket / cc-pocket) — Continue your local AI coding tasks from phone, tablet, or desktop.

> Pairlet added support for sending richer media from mobile and desktop clients, including video uploads and multi-image sending.

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

01 **Video upload and multi-image sending** Not checked yet NEW

25

Pairlet now supports video upload and supports sending multiple images at once.

– States what was added but no mechanism or limits given v2.6.6

WAS THIS USEFUL? 👍 👎

palot

Feedback ⓘ v0.24.0 NOTES ↗ RANK

Palot is a just another desktop client for OpenCode 2. Manage projects, sessions, parallel agents, requests, and changes on Linux and macOS.

> Palot shipped native desktop packages for Linux and macOS plus a way to verify their integrity via checksums and GitHub attestations, while cutting auto-update support for legacy Palot 0.11.

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

01 **Package integrity verification via SHA256SUMS and GitHub attestation**

Not checked yet NEW 80

Downloaded packages can be verified against `SHA256SUMS` or checked for GitHub build attestations using `gh attestation verify <downloaded-package-file> --repo ItsWendell/palot`.

Confirm a downloaded palot package has not been tampered with using GitHub build attestations.

```
$ gh attestation verify <downloaded-package-file> --repo
ItsWendell/palot
```

– Gives a runnable verification command and file name v0.24.0

THINNER COVERAGE BELOW

02 **Desktop packages for Linux and macOS** Not checked yet NEW 57

palot now ships desktop packages for Linux x64 in `AppImage`, `DEB`, `RPM`, and `tar.gz` formats, and for macOS Apple Silicon as `DMG` and `ZIP`.

– Lists formats and platforms but no install steps v0.24.0

03 **No auto-update feed for legacy Palot 0.11** Not checked yet 35

BREAKING

No auto-update feeds for legacy Palot 0.11 are published, so updates from that version must be performed manually.

– States the break but no migration steps v0.24.0

└──▷ **BREAKING ON UPGRADE**

! No auto-update feeds for legacy Palot 0.11 are published; updates are manual.

WAS THIS USEFUL? 👍 👎

All Hands AI OpenHands

Feedback ⓘ v1.46.0 NOTES ↗

RANK

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

> OpenHands shipped a major platform expansion with Agent Canvas (supporting the Agent Client Protocol), a scheduling-and-webhook Automation Server, a Python Software Agent SDK with REST/WebSocket APIs, and new OpenHands Enterprise features, while also adding team conversation sharing, a feedback flow, and a local Git Sync toggle for automations.

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

01 **Agent Canvas with Agent Client Protocol support** Not checked yet

NEW

65

OpenHands adds Agent Canvas, installable via npm, npx, or Docker, with a desktop app currently in preview; it can connect to a VM, Docker, or OpenHands Cloud as a backend. External agents such as Claude Code or Codex can connect to Agent Canvas through the Agent Client Protocol.

— Names install methods and protocol but no exact command. product docs

02 **Software Agent SDK and Agent Server APIs** Not checked yet NEW

60

OpenHands adds a Software Agent SDK — a Python library for building agents — alongside an Agent Server that exposes agent execution, conversations, tools, and workspaces through REST and WebSocket APIs.

— Names SDK language and API surfaces, no concrete endpoints. product docs

03 **OpenHands Enterprise organizations, SSO, and integrations**

Not checked yet NEW

60

OpenHands Enterprise adds organizations, roles, single sign-on, service accounts, and budgets, along with built-in integrations with GitHub, Jira Data Center, and Azure DevOps. It is deployable via Docker Compose or Kubernetes.

— Lists concrete integrations and deployment options, no config keys. product docs

04 **Local Git Sync activation toggle for automations** Not checked yet

NEW

60

Adds a Git Sync switch in automations so the activation status is stored locally rather than synced remotely; in the UI, navigate to Automations > Git Sync and toggle the

switch to save the state on your device without pushing it to the remote.

Navigate to the Git Sync automation settings to keep the toggle state saved locally without pushing it to the remote.

> 📍 In the OpenHands UI, go to Automations › Git Sync and toggle the Git Sync switch — the activation status is saved locally on your device.

— Exact UI navigation path given for the toggle. v1.46.0

THINNER COVERAGE BELOW

05 **Automation Server for scheduled and event-driven agents**

Not checked yet NEW 45

OpenHands adds an Automation Server to run agents on a schedule or in response to events, with public webhook URLs available on OpenHands Cloud.

— Describes behavior but no endpoint or config detail. product docs

06 **Legacy local GUI and backend repository deprecated**

Not checked yet DEPRECATED 40

The legacy local GUI is deprecated in favor of Agent Canvas for maintained local installation, and the previous backend repository is no longer actively maintained, preserved for reference only.

— States what's deprecated but no migration steps. product docs

07 **Team conversation sharing management** Not checked yet NEW

40

Adds the ability to list and unshare conversations that have been shared with a team organisation.

— Names the capability but no UI path or API. v1.46.0

08 **Skills, plugins, and MCP server extensibility** Not checked yet

NEW 25

Agents across all OpenHands products can now be extended with skills, plugins, and MCP servers.

— Bare mention with no mechanism or example. product docs

09 **Feedback submission flow** Not checked yet NEW 20

Adds a Feedback button and submission flow to the UI.

— Bare mention, no further detail provided. v1.46.0

└──▷ BREAKING ON UPGRADE

! The legacy local GUI is deprecated; use Agent Canvas for maintained local installation.

! The previous backend repository is no longer actively maintained and is preserved for reference only.

WAS THIS USEFUL? 👍 👎

Earendil Works Pi

Feedback ⓘ v1.6.1 NOTES ↗ RANK

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

> Pi's v1.6.1 release adds decision-model classification and rate-limit visibility for extensions, splits notice settings into granular controls with a breaking API removal, and makes several smaller config and tooling additions.

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

01 Granular transcript notice settings replace single toggle

Not checked yet BREAKING 78

Control transcript notices individually with `notices.cacheMisses`, `notices.cacheWarming`, `notices.compactionCosts`, and `notices.reasoningDrops` under Notices in `/settings`, replacing the single `showCacheMissNotices` setting (existing values migrated automatically). SettingsManager.getShowCacheMissNotices() and setShowCacheMissNotices() are removed; use getNotices() and setNotice() instead.

— Names old and new config keys and API methods, no example v1.6.1

02 blockedKind dialog status for extensions

Not checked yet NEW 78

Set the `blockedKind` option on ctx.ui.select(), ctx.ui.confirm(), and ctx.ui.input() so terminals that display program status can show a dialog as a permission request, question, or login.

Signal a permission-style dialog status when an extension prompts the user to approve a potentially destructive action.

typescript

```
const approved = await ctx.ui.select(
  'Allow pi to delete build artifacts?',
  ['Yes', 'No'],
  { blockedKind: 'permission' }
);
```

— Runnable code example naming the exact methods and option v1.6.1

03 Decision model support in codemode

Not checked yet NEW 75

Call GitHub Copilot's `gpt-6-luna-decisions` and Radius decision models from codemode models.classify(), with text and images; results now include a `responseId` field populated by the provider.

Call a decision model from codemode to classify content and capture the provider's response ID for billing or audit.

typescript

```
const result = await models.classify({
  model: 'gpt-6-luna-decisions',
  input: [{ type: 'text', text: 'Is this output safe to display?'
}]
});
console.log(result.responseId, result.label);
```

— Named models and field, with a runnable code example v1.6.1

04 **Rate-limit response details in provider response hook**

Not checked yet IMPROVED 65

Extensions can now read the HTTP status and `retry-after` header of rate-limited requests because `after_provider_response` now fires for 4xx/5xx error responses and each retried attempt on OpenAI, Azure OpenAI, Anthropic, and Google models.

— Names hook and providers but lacks a usage example v1.6.1

05 **Strict mode flag for Gemini models in models.json** Not checked yet

NEW 60

Add `compat.supportsStrictMode` in `models.json` for `google-generative-ai` and `google-vertex` models so Gemini 3+ models served under unversioned IDs can use strict tools.

— Names config key and file, no usage example v1.6.1

THINNER COVERAGE BELOW

06 **Bash tool truncation error field** Not checked yet IMPROVED 55

The bash tool's structured output now includes `full_output_error` (and `fullOutputError` in `BashResult` and `BashExecutionMessage`) explaining why truncated output could not be saved to a temp file.

— Names fields but no example or deeper mechanism v1.6.1

07 **Default Mistral model changed** Not checked yet BREAKING 50

The default Mistral model changed from `devstral-medium-latest` to `mistral-medium-latest` because Mistral retired Devstral.

— Names old and new model IDs, minimal further context v1.6.1

└──▷ **BREAKING ON UPGRADE**

! The `showCacheMissNotices` setting is removed and replaced by four individual settings — `notices.cacheMisses`, `notices.cacheWarming`, `notices.compactionCosts`, and `notices.reasoningDrops` — configured under Notices in `/settings`; existing `showCacheMissNotices` values are migrated automatically.

! SettingsManager.getShowCacheMissNotices() and setShowCacheMissNotices() are removed; use getNotices() and setNotice() instead.

! The default Mistral model changed from `devstral-medium-latest` to `mistral-medium-latest` because Mistral retired Devstral.

WAS THIS USEFUL? 👍 👎

Graphify Labs Graphify

Feedback ⓘ v0.9.120 NOTES ↗

RANK

Graphify converts code, docs, SQL schemas, configs, and PDFs into a queryable knowledge graph with explainable relationships.

> Graphify expanded its MCP server with new blast-radius and seeded-traversal tools plus caching-driven query speedups, shipped an official Cursor Marketplace plugin for in-editor graph queries, and bumped its AST cache schema to fix a Python import bug.

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

01 Official Cursor Marketplace plugin

Not checked yet NEW INTEGRATION

NEW

95

Install the Graphify Cursor plugin via `/add-plugin graphify` in Cursor chat, or by searching Graphify under Customize in the sidebar, to connect Cursor to Graphify's code knowledge graph over the MCP server at `https://api.graphify.com/mcp`, with OAuth sign-in and no API key required. The plugin bundles a rule telling Cursor's agent when to reach for Graphify — covering indexed code, symbols, dependencies, change impact, and repository memory — citing the files it found, and lets users query call graphs, symbol reach, test linkage, and change-impact across repositories directly from Cursor, with answers returning file and line citations.

Add the Graphify plugin to Cursor from the chat bar to immediately start querying your code knowledge graph without leaving the editor.

```
$ /add-plugin graphify
```

– Exact install command, endpoint, and auth model given launch-20261124-ff2c7bfd

02 Blast-radius analysis via `affected` MCP tool

Not checked yet

NEW

72

The new `affected` MCP tool computes the reverse-dependency blast radius of any symbol, resolving the target node by name, file path, or id.

– Names the tool and resolution modes but no sample call v0.9.120

03 Query performance caching

Not checked yet IMPROVED

65

Query performance is now faster because per-node search fields, degrees, hub threshold, integer-indexed adjacency for BFS/DFS, and rendered NODE lines are each computed once per loaded graph and reused across all subsequent queries.

— Mechanism detailed but nothing for reader to invoke v0.9.120

04 **Seeded traversal via `seeds` argument** Not checked yet IMPROVED

60

The `query_graph` MCP tool gains an optional `seeds` argument to seed a graph traversal from known nodes.

— Named argument but no usage example given v0.9.120

└──▷ **BREAKING ON UPGRADE**

! The AST cache schema bumps from version 55 to 56 due to a Python function-local-import fix; Python files re-extract automatically on the first build after upgrading.

announced · Nov 24, 2026 BLOG ↗

Graphify adds a Grok Bot plugin connecting Bots to the Graphify MCP server via OAuth, enabling code-graph queries in Slack with no API key.

Connect Bots to the Graphify MCP server at `https://api.graphify.com/mcp` as a custom connector, enabling code-graph queries — symbol lookup, call traces, change-impact checks, linked tests, and PR review status — directly from Grok Bot. Not checked yet

Authenticate via OAuth in the browser when adding the Graphify Grok Bot plugin, with no API key required. Not checked yet

WAS THIS USEFUL? 👍 👎

Yuval Sarel cones

Feedback ⓘ v0.4.35 NOTES ↗ RANK

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

> Cones improved its terminal dashboard with richer read-only previews, drag-and-drop session reordering, smarter idle detection for background Claude sessions, and the ability to take over a running session directly from the dashboard.

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

01 **Reorder child sessions via drag or keyboard** Not checked yet

IMPROVED 85

A child session (a launch or fork) can now be reordered among its siblings using `ctrl+b` or a mouse drag; moving a top row still carries all sessions beneath it. Dropping a row onto one at a different nesting level now shows a plain error message instead of mislabelling every nested row as a fork.

Reorder a child session up among its siblings from the keyboard without touching the mouse.

> 📍 # In the cones dashboard, highlight a launch or fork session, then press ctrl+b to move it up among sessions sharing the same parent.

– Exact key combo plus usage example provided v0.4.34

02 **Inline tool-call output in read-only previews** Not checked yet

NEW 80

Tool calls in a read-only preview now show their first three lines of output followed by a `+N lines` note, matching Claude Code's style, for Claude Code, Codex, and pi sessions. Pressing `ctrl+o` in a focused preview expands every tool call to its full input and result (up to 4 KiB each), or collapses them again.

– Names exact shortcut and byte limit, no screenshot v0.4.35

03 **Take over a session's own terminal from dashboard** Not checked yet

IMPROVED 65

Pressing Enter in the dashboard on a session running in its own terminal now offers to take it over (shown as `take over` in the row hint) instead of only opening a read-only preview; harnesses that cannot resume still show `watch`.

– Names exact key and hint labels, no walkthrough v0.4.32

THINNER COVERAGE BELOW

04 **Idle status for Artifact-watching background sessions**

Not checked yet IMPROVED 50

Finished Claude background sessions that are only watching Artifact comments now show as idle instead of working in the dashboard; a new comment starts a turn and shows the session as working again.

— Explains behavior but no UI path or command v0.4.33

WAS THIS USEFUL? 👍 👎

mvschwarz OpenRig

Feedback ⓘ v0.7.23 NOTES ↗ RANK

OpenRig is a network manager for AI coding agents: build persistent teams with roles, shared context and owned work.

> OpenRig's v0.7.23 release adds peer-to-peer network requests between hub seats, opt-in seat lifecycle controls like `cold_start` and `keep_alive`, and a new queue-transfer export format, alongside smaller additions to workflow handoffs, build-order scoping, and Codex seat restarts.

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

01 Cross-seat peer networking and relay

Not checked yet NEW

90

Adds cross-seat networking with `rig network peers <binding-or-seat>` to list hub peers and `rig network request <binding-or-seat> <peer> --id <request-id> --body-file <path>` to send requests; answers return as local queue items with tokens never exposed to the agent. The new `rig queue relay-answer` command relays a person's answer back to the originating agent and is always marked as relayed.

Send a request from your seat to a hub peer and receive the answer as a local queue item, without exposing the hub token to the agent.

```
$ rig network peers my-hub-binding
rig network request my-hub-binding peer-seat-id --id req-001 --
body-file ./request-payload.json
```

— Names full commands with flags and behaviour, runnable as-is v0.7.23

02 Seat lifecycle controls: cold_start and keep_alive

Not checked yet NEW

70

Set `cold_start: true` on a pod or member to start a dormant seat from queued work, and opt in `keep_alive` on a seat to relaunch it if it falls.

Let queued work automatically start only a dormant seat rather than the whole rig, to handle a backlogged task with minimal resource use.

```yaml
cold_start: true
```

— Named config keys with example, mechanism briefly explained v0.7.23

03 **In-place Codex seat restart** Not checked yet NEW 70

Restart a Codex seat in place on its own conversation on a Mac using `rig seat restart <codex-seat> --keep-conversation` or `rig provider switch`.

— Exact commands and platform scope named v0.7.23

04 **Daemon startup readiness wait control** Not checked yet NEW 60

Set `daemon.start_readiness_wait_seconds` to control how long the daemon waits for readiness at startup, with docs for running under a supervisor.

— Named config key with clear purpose, no example value given v0.7.23

05 **Build-order edges for scope slices** Not checked yet NEW 60

Define build-order edges that readiness reads with `rig scope slice create --execution-depends-on`.

— Exact runnable command and flag, thin on mechanism v0.7.23

THINNER COVERAGE BELOW

06 **Workflow step labels and handoff advice** Not checked yet IMPROVED 55

Workflow steps accept an optional label, same-role handoffs now include advice and a `next_owner: same_as_previous` field, and run titles appear on queue packets.

— Named field and behaviours but no usage example v0.7.23

07 **Terminal pane geometry query in browser viewer** Not checked yet NEW 50

Query a terminal pane's size from a browser viewer by appending `?geometry=1` to the viewer URL.

— Exact query parameter and usage given v0.7.23

08 **Rig start exclusion reporting** Not checked yet IMPROVED 50

`rig start --last` now names every rig it leaves out and the reason it was excluded.

— Named command and behaviour, no example output shown v0.7.23

09 **Queue state export and staged import** Not checked yet NEW 45

Export queue state in the `openrig.queue-transfer/v2` format and stage an import for review.

— Format named but no command or workflow shown v0.7.23

10 **Queue list page-cut reporting** Not checked yet IMPROVED

30

`rig queue list` now reports when it cut a page.

— Bare statement with no mechanism or example v0.7.23

WAS THIS USEFUL? 👍 👎

swttch

Feedback ⓘ v0.38.3 NOTES ↗ RANK

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

> Swttch's only update in this window adds a configurable scroll speed setting for the chat view.

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

01 **Scroll speed setting for chat view** Not checked yet NEW

25

Adds a scroll speed setting for the chat view, letting users adjust how fast the chat session scrolls.

— Names the setting but gives no location or mechanism v0.38.3

WAS THIS USEFUL? 👍 👎

aerovato Operator Memory

Feedback ⓘ deepseek@v0.5.0

NOTES ↗ RANK

Operator Memory is a self-improving context engine for coding agents.

> Operator Memory rolled out preamble v2.1 across all six harness adapters (retiring v1.1), while adding a per-project settings file and a content-neutral `/operator:project-init` scaffolding command.

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

01 Preamble v2.1 rollout across harnesses

Not checked yet BREAKING

65

Operator Memory upgraded its internal preamble to v2.1 across the DeepSeek, Pi, Claude Code, Helper, Codex and OpenCode V2 harness adapters. Preamble v1.1 is retired; v2.0 is now the only supported baseline (further upgraded to v2.1 here), so any configuration or workflow depending on v1.1 will no longer work. The OpenCode V2 adapter can be installed with `operator-helper install opencode` to get the fully supported preamble v2.1 behavior.

Install the Operator adapter for OpenCode V2 to gain preamble v2.1 and the new settings file support.

```
$ operator-helper install opencode
```

– Names versions and an install command but not the mechanism change deepseek@v0.5.0 pi@v1.4.0 claude-code@v0.4.0 helper@v1.14.0 codex@v0.6.0 opencode-v2@v2.5.0

THINNER COVERAGE BELOW

02 Content-neutral project-init command

Not checked yet IMPROVED

55

The `/operator:project-init` command now scaffolds a project without imposing opinionated content defaults or assuming any existing document structure.

– Names the exact command but not its options helper@v1.14.0 claude-code@v0.4.0 deepseek@v0.5.0

03 Settings file for per-project configuration

Not checked yet NEW

20

Operator Memory now supports a settings file for configuring the context engine on a per-project basis, available in the DeepSeek, Claude Code, Helper and OpenCode V2

adapters.

— No config key, file path, or schema given `helper@v1.14.0` `claude-code@v0.4.0` `deepseek@v0.5.0` `opencode-v2@v2.5.0`

└──▷ **BREAKING ON UPGRADE**

! Preamble v1.1 is retired; v2.0 is now the only supported preamble (upgraded to v2.1 in this release). Any configuration or workflow depending on preamble v1.1 will no longer work.

! Preamble v1.1 is retired; v2.0 is now the only preamble version. Any configuration or workflow that depended on v1.1 will no longer function.

! Preamble v1.1 has been retired; v2.0 is now the only supported preamble version.

! Preamble v1.1 is retired; v2.0 is now the only supported preamble (upgraded further to v2.1 in this release). Setups relying on preamble v1.1 will no longer work.

! Preamble v1.1 is retired; v2.0 is now the only supported preamble, so any configuration or workflow depending on v1.1 will no longer function.

WAS THIS USEFUL?  

agentbro

Feedback ⓘ v3.7.0 NOTES ↗ RANK

Agentbro is a macOS Dynamic Island for AI coding agents: Claude Code, Codex, Gemini CLI, Cursor, hooks, skills, and local agent management.

> AgentBro added support for managing Qoder Desktop, IDE, and CLI as agents, extended `CLAUDE_CONFIG_DIR` awareness across Hooks and Skill management, and shipped keyboard-driven Island question answering alongside a run of Skill Pack migration, SSH distribution, and MCP editor reliability fixes.

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

01 **Qoder Desktop, IDE, and CLI agent support** Not checked yet NEW

70

AgentBro now manages Qoder Desktop, Qoder IDE, and Qoder CLI as separate agents on macOS and Windows, with live task activity shown in the Hook and allow-once or deny approvals available in the Island for Qoder Desktop.

— Names the three new agent types and the approval mechanism v3.6.0

02 **Keyboard navigation for Island questions** Not checked yet NEW

65

Island questions can now be answered entirely by keyboard, using Tab, Space, and Enter to select options and submit answers; the question stays visible while the Island holds keyboard focus.

— Names the exact keys but no deeper mechanism v3.7.0

03 **SSH Claude Code Hooks follow CLAUDE_CONFIG_DIR** Not checked yet

IMPROVED

65

Claude Code Hooks over SSH now install, check, and remove files in the directory named by the remote login shell's `CLAUDE_CONFIG_DIR` environment variable.

— Names the env var and the three operations it affects v3.6.0

04 **Skill management follows CLAUDE_CONFIG_DIR** Not checked yet

IMPROVED

65

Local Skill discovery, distribution, and plugin settings now follow the `CLAUDE_CONFIG_DIR` environment variable on macOS and Windows; Agent Management reports active paths, and Skill Pack preview shows the destination for each change.

— Names the env var and the three affected surfaces v3.5.0

05 **Skill Pack export/import for Agent migration** Not checked yet

NEW

65

Skill Pack export now works offline via ZIP file; export for migration also records which built-in Agents use the pack, and on import you can edit suggested targets, review changes, and confirm distribution — on Windows, detected Agent launchers are selectable even before their Skills folders exist.

— Describes mechanism and format but limited numeric detail v3.5.0 v3.7.0

---

06 **OpenCode native MCP config display** Not checked yet NEW

65

Agent Management now displays OpenCode's native MCP configuration — entries from user-level JSON and JSONC files on local and SSH hosts — showing configuration source and enabled state, read-only.

Inspect which MCP servers OpenCode has enabled on a remote SSH host without leaving AgentBro — the read-only view shows each entry's source file and enabled state.

> 📍 In AgentBro, open Agent Management, select an SSH host, and open the MCP configuration panel to view OpenCode's JSON/JSONC entries with their source and enabled state.

— Has a usage example showing the exact UI path v3.3.2

---

07 **Island shortcut behavior refinements** Not checked yet IMPROVED

60

Island shortcuts now act once per key press, use the visible prompt when the Island is open, pick the latest pending request when the Island is collapsed, and open structured questions when you press Skip.

— Multiple concrete behaviors but no exact shortcut keys v3.6.0

THINNER COVERAGE BELOW

08 **SSH Skill distribution pre-write rechecking** Not checked yet

IMPROVED

55

AgentBro now rechecks SSH Skill distribution before writing: if pack members, source files, destinations, or existing copies change after preview, a fresh review is required, and quick-apply conflicts open a review before any files are written.

— Clear safety mechanism but no named flag or command v3.5.0

---

09 **Windows startup recovery and welcome screen flow** Not checked yet

IMPROVED

50

On Windows, when startup stalls AgentBro now offers to keep waiting or close the app, and the welcome screen closes automatically after the display choice is saved, with a retry option if saving fails.

— Describes the flow but no UI path or trigger details v3.7.0

10 **Windows Agent Management live PATH scanning** Not checked yet

IMPROVED 50

Agent Management on Windows now reads current saved PATH entries when scanning for agents, so CLI tools installed after AgentBro opened are discovered on the next refresh.

— Clear before/after but no exact refresh trigger v3.6.0

11 **GitHub source verification for Skill Pack export** Not checked yet

NEW 50

Skill Pack export now verifies GitHub sources, retaining a source only when the recorded commit and path reproduce the current Skill; locally edited Skills instead export with their current fingerprints.

— Names commit/path verification but no exact surface v3.4.0

12 **Skill restore conflict handling** Not checked yet NEW 45

Restoring a pinned Skill that conflicts with different content already held under the same ID now keeps both versions by restoring the pinned version under a new ID rather than overwriting.

— Explains the behavior but no UI path or command v3.4.0

13 **MCP editor focus and draft preservation fixes** Not checked yet

IMPROVED 45

The MCP editor now preserves keyboard focus and draft values when renaming environment variables or HTTP headers, switching transports, or saving with an empty replacement credential field.

— Lists triggering actions but no reproduction steps v3.3.2

14 **MCP changes scoped per Agent and host** Not checked yet IMPROVED 45

MCP changes now stay scoped to the reviewed Agent and host — switching between Agents, local machine, or SSH hosts clears outdated inventories and confirmations so a late response from a previous host cannot overwrite the current list.

— Explains the race condition fixed but no named surface v3.3.2

WAS THIS USEFUL? 👍 👎

coffee-cli

Feedback ⓘ v4.1.1 NOTES ↗ RANK

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

> Coffee CLI's latest releases add a dedicated Work layout, remote control over Codex, OpenCode, and DeepSeek Harness sessions, and desktop conversation support for Antigravity, Cursor, and Kiro, alongside composer attachment previews, usage-tracking start cards, and team-collaboration and update-flow improvements.

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

01 **Work and Code layout modes** Not checked yet NEW 75

A new Work layout shows only the conversation panel with an icon rail and no terminal windows, introduced under Settings → Appearance → Layout. This was later extended to let you switch between Work and Code layouts from the same menu, with mobile automatically following the desktop choice.

— Names exact settings path and both layout modes v4.1.0 v4.0.9

02 **Usage tracking on Code start cards** Not checked yet NEW 75

Adds Code start cards that show remaining usage, open windows, and balance — with a red edge when a window is used up — and let you set the launch permission for Claude Code or Codex directly on the card.

— Concrete UI elements and behaviors named, clear starting point v4.1.1

03 **Remote control for Codex, OpenCode, and DeepSeek Harness** Not checked yet NEW 70

Coffee CLI now lets Codex, OpenCode, and DeepSeek Harness operate remote computer windows, with a dedicated Remote Control page in app config and local-microphone voice input for remote sessions.

— Mechanism and named tools given, UI location noted v4.1.0

04 **Local and Remote team tabs in Collaboration** Not checked yet NEW 60

Splits the Collaboration page into Local team and Remote team tabs, letting a conversation hand work to an agent app on another computer with full or restricted access.

— Names tabs and access levels but not exact flow v4.1.1

05 **Shared step line for conversation turns** Not checked yet NEW

60

Adds a shared step line across all five conversation tools, collapsing finished turns into one row listing changed files with file-type icons, and showing the reason for any failed step.

— Explains mechanism but no interaction instructions v4.1.1

06 **Composer attachment previews and draft saving** Not checked yet

NEW

60

Attached images in the composer now display as thumbnails, and files and folders display as cards. You can also save a composer draft to pre-input with ⌘S or Ctrl+S.

— Named keyboard shortcut makes it directly actionable v4.0.9

THINNER COVERAGE BELOW

07 **Changelog preview before update** Not checked yet IMPROVED

55

The update flow now shows the new version's changelog before proceeding, letting you review changes before clicking Start update; on Windows the installer downloads first and installs on click.

— Describes new flow but limited mechanism detail v4.1.0

08 **Tool cards remember project folder and launch target**

Not checked yet IMPROVED

50

Each tool card on the start page now remembers its own last project folder and can open in the IDE desktop or the native CLI terminal.

— Clear behavior but no config detail v4.1.0

09 **Desktop conversations for Antigravity, Cursor, and Kiro**

Not checked yet NEW

45

Adds desktop conversation mode for Antigravity, Cursor, and Kiro ACP tools, each with its own particle welcome icon.

— Names three integrations but little mechanism v4.1.1

10 **Full conversation history loading** Not checked yet IMPROVED

40

Opening a conversation now loads the full history without a Load earlier messages button, and scrolling up no longer drops content below.

— Behavioral fix described, no further detail v4.1.0

11 **'My workbench' replaces 'New session'** Not checked yet IMPROVED

35

Renames New session to My workbench with a house icon, returning you to the tool picker in one click.

— Minor rename, thinly described v4.1.1

WAS THIS USEFUL? 👍 👎

Factory

Feedback ⓘ

CLI v1.21.0, Desktop v1.21.0: API key visibility, repository picker recovery, and fixes

BLOG ↗ RANK

Factory's Droids are AI software-development agents with a CLI, web app and API.

> Factory's v1.21.0 release gives organization owners visibility into personal API key usage across the org, and improves error handling and tool display in the CLI and app.

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

01 **Organization-wide API key inventory** Not checked yet NEW

50

Organization owners can now list every personal Factory API key across their organization via the API key inventory, and Settings shows when each personal Factory API key was last used.

— Names the feature and location but no exact UI path or endpoint

CLI v1.21.0, Desktop v1.21.0: API key v…

02 **Repository picker load-failure reporting** Not checked yet IMPROVED

40

Repository pickers now report when a code host's repositories could not be loaded, instead of silently showing nothing.

— Describes behavior change but no specific hosts or UI steps

CLI v1.21.0, Desktop v1.21.0: API key v…

03 **Generic tool result card** Not checked yet IMPROVED

30

Tools without their own view now show a card with their arguments and result.

— Brief description, no concrete examples given CLI v1.21.0, Desktop v1.21.0: API key v…

WAS THIS USEFUL? 👍 👎

agents-cli

Feedback ⓘ v1.16.0 NOTES ↗ RANK

The CLI and skills that turn any coding assistant into an expert at creating, evaluating, and deploying AI agents on Google Cloud.

> agents-cli (v1.16.0) adds dry-run visibility into Cloud Run IAM grants and organization policy failures, stricter eval metric validation and gating, GitLab nested-group template fetching, and tighter controls over extension trust and updates.

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

01 Dry-run IAM and policy diagnostics for Cloud Run deploys

Not checked yet IMPROVED 85

`agents-cli deploy --dry-run` for Cloud Run with Agent Identity now lists the IAM roles granted and the receiving principal before deploying. When a deploy to Cloud Run is rejected by an organization policy, it now names that policy and, for `constraints/run.allowedIngress`, suggests the `--ingress` values left to try.

— Names exact flag, constraint and behavior before/after v1.16.0

02 Eval metric validation and threshold gating Not checked yet

IMPROVED 85

`agents-cli eval run` now validates eval metrics before running inference, failing immediately when a metric is misconfigured. `agents-cli eval run` and `eval grade` without thresholds now end with `RESULT: NOT GATED` plus a warning on how to set a threshold, instead of previously ending with `RESULT: PASS`.

— Exact commands and output strings given with before/after v1.16.0

03 Session-type opt-in for extension templates Not checked yet NEW

85

Extension templates can opt into `--session-type` on `agent_runtime` by setting `settings.agent_runtime_session_types` in `.template/templateconfig.yaml`.

— Exact config key and file path to edit v1.16.0

04 GitLab nested-group template fetching Not checked yet IMPROVED

70

`agents-cli scaffold create --agent` can now fetch templates from GitLab repos located in nested groups, and accepts GitLab `/-/tree/<ref>/<path>` links directly.

— Names exact command and accepted link format v1.16.0

05 **Extension update origin and trust-prompt pinning** Not checked yet

IMPROVED 70

`agents-cli extension update <name>` now updates a first-party extension's origin to `google/agents-cli-extensions`. Extensions from pre-approved sources now install without a trust prompt only when pinned to a concrete release tag such as `v1.2.3`.

— Names command, origin repo and pinning condition v1.16.0

THINNER COVERAGE BELOW

06 **Template suggestions for unknown --agent value** Not checked yet

IMPROVED 45

An unknown `--agent` value in `agents-cli create` now lists the built-in templates you can use.

— Small UX tweak with no further mechanism given v1.16.0

WAS THIS USEFUL? 👍 👎

kane-cli

Feedback ⓘ 0.10.9 NOTES ↗ RANK

Kane CLI by TestMu AI (Formerly LambdaTest)

> Kane CLI shipped a new AI-agent integration that lets teammates' coding agents run real end-to-end tests automatically from CLAUDE.md/AGENTS.md, plus TUI improvements for selecting the right test device.

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

01 AI agent-driven E2E testing setup

Not checked yet NEW

89

Embed a `kane-cli:start` / `kane-cli:end` block in `CLAUDE.md` or `AGENTS.md` at the repo root so AI coding agents (Claude Code, Codex, Cursor, and others) automatically run `kane-cli run "<objective>" --agent` before committing UI or user-facing changes. Mobile apps can be targeted in these agent-driven runs with `--target android` or `--target ios` added to `kane-cli run`. The instruction block can be installed interactively via the Kane CLI skill with `npx @testmuai/kane-cli-skill`.

Paste this block into CLAUDE.md or AGENTS.md so every AI agent in the repo runs a real-browser E2E check before committing a UI change.

text

```
<!-- kane-cli:start -->
## End-to-end testing
Use Kane CLI (`kane-cli`) for end-to-end, regression and smoke
testing of web and mobile apps in this repository.
- After changing UI or user-facing behaviour, verify the flow in
a real browser or mobile app before committing: `kane-cli run "
<objective>" --agent`
<!-- kane-cli:end -->
```

Run an agent-driven E2E test against an Android app after a UI change, so CI catches regressions on mobile before merge.

```
$ kane-cli run "Complete checkout flow" --agent --target android
```

Bootstrap a new teammate's agent setup so it can automatically add the Kane CLI instruction block to CLAUDE.md on request.

```
$ npm install -g @testmuai/kane-cli && npx @testmuai/kane-cli-skill
```

— Names exact block markers, files, flags and install command. product docs

THINNER COVERAGE BELOW

02 **TUI device selection improvements** Not checked yet IMPROVED

52

The TUI now labels the tunnel as real-device only so the limitation is visible before a run starts, shows private (organization-dedicated) devices alongside the shared device pool, and marks recommended targets in a curated device catalog with a TIP badge.

— Describes three TUI changes but no exact UI path or flag. 0.10.9

WAS THIS USEFUL? 👍 👎

ai-rulez

Feedback ⓘ v5.15.0 NOTES ↗ RANK

Ai-rulez is an author, validate and govern AI agent rules, skills, and MCP configuration across 52 coding assistants with model evidence and bundle publishing.

> ai-rulez v5.15.0 adds a configurable fallback for the Cursor hook root in its Cursor plugin.

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

01 **Configurable Cursor hook root fallback** Not checked yet IMPROVED

35

The Cursor plugin now supports a configurable fallback for the Cursor hook root, letting users control where hooks are anchored when no explicit root is set.

— Describes behavior but no config key or flag name given v5.15.0

WAS THIS USEFUL? 👍 👎

prflow

Feedback ⓘ v2.66.0 NOTES ↗ RANK

prflow is a tool that turns feature requests into review-ready pull requests with implementation, tests, review loop, and docs.

> prflow's v2.66.0 release sharpens `/prflow:spec` audit rounds with filing recommendations and distributed acceptance-criteria checks, fixes the weekly retrospective to stay on the checked-out branch, and locks workflow tokens to read-only, which breaks any step reading the removed `health_summary` output.

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

01 **Read-only Actions tokens for workflow jobs** Not checked yet

85

BREAKING

The `command` job in `devflow.yml` and the `retrospective` job in `devflow-retrospective.yml` now use a read-only Actions token, removing the ability to dispatch, re-run, cancel or delete workflow runs; the `health_summary` output of the `setup-project-env` action is also removed. Any workflow step reading that output will break; re-run the installer to receive updated `devflow.yml` and `devflow-retrospective.yml` with read-only job tokens.

— Names exact files, jobs, output, and gives migration step. v2.66.0

02 **Filing and coverage improvements in /prflow:spec audits**

Not checked yet IMPROVED

70

The `/prflow:spec` command can now recommend filing instead of scheduling another audit round, and can show which draft sections changed since the latest audit round that returned a verdict. An audit round can also return FILE for a draft whose only fault is a false statement nothing rests on, reporting that statement as a note rather than blocking filing. Additionally, each auditor in a three-part `/prflow:spec` first round now checks a share of the acceptance criteria, rather than one auditor covering only descriptive sections.

— Describes multiple concrete behavior changes to a named command. v2.66.0

THINNER COVERAGE BELOW

03 **Retrospective stays on checked-out branch** Not checked yet

IMPROVED

45

The weekly retrospective now runs on the branch that is checked out rather than switching your checkout to `main`, both locally and in the scheduled workflow.

— Clear behavior change but no further mechanism detail given. v2.66.0

└──▷ **BREAKING ON UPGRADE**

! The `health_summary` output of the `setup-project-env` action is removed; any workflow step reading that output will break. Re-run the installer to receive updated `devflow.yml` and `devflow-retrospective.yml` with read-only job tokens.

WAS THIS USEFUL?  

◆ **AI App Builders**

Lovable

Feedback ⓘ Nov 25, 2026 (2026-11-25) BLOG ↗

Lovable builds web apps and sites from natural-language prompts.

Lovable adds ElevenLabs speech and music via four models (no API key needed) and expands Automations to Enterprise with 50-automation and 5-minute limits.

NEW INTEGRATION Enable ElevenLabs speech and music generation in a project — workspace admins turn it on in Connectors or when Lovable prompts in project chat — supporting `elevenlabs/eleven_v3`, `elevenlabs/eleven_multilingual_v2`, `elevenlabs/eleven_flash_v2_5`, and `elevenlabs/music_v2`, with no ElevenLabs account or API key required. Not checked yet

Automations are now available on Enterprise plans, with up to 50 automations per person per workspace and a minimum run interval of every 5 minutes. Not checked yet

---

WAS THIS USEFUL?  

◆ Coding Agent Add-ons

Paradev

Feedback ⓘ cli-v2.0.12 NOTES ↗ RANK

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

> Paradev rebranded its CLI from Superset to Paradev (with backward-compatible `SUPERSET_*` aliases), expanded cloud workspace management with per-environment machine sizing, box pressure views, and `[install]` `-block` environment builds, and rolled out Android and web support alongside AWS Marketplace fulfillment, new PR workflow actions, and Gemini 4 models.

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

01 Cloud workspace management and environment configuration

Not checked yet IMPROVED 95

Paradev adds a ports button beside the Share button in the cloud workspace tab bar, lets you view box pressure for every cloud workspace, and lets you change machine size from the desktop UI or via the `environments.machine_size` config key, managed from an environment page in Settings or set up directly from chat. Workspaces can be moved to the cloud from any computer, carrying Claude terminal sessions, and cloud workspaces can now create and build environments from an `[install]` block, removing the need for a local setup step.

Check cloud workspace resource pressure and switch to a larger machine size when a box is under load.

> 📍 In the desktop UI, open a cloud workspace, then go to Settings › Environments, select the environment, and choose a new machine size under the machine size option.

Move a local workspace to the cloud so teammates can access it without keeping your computer on.

> 📍 In the desktop sidebar, right-click a workspace and choose Move › Cloud from the workspace menu.

— Names the config key and gives exact Settings navigation steps. cli-v2.0.12 desktop-v2.0.12

02 CLI rebrand from Superset to Paradev

Not checked yet IMPROVED 80

The CLI is rebranded from Superset to Paradev: it sets `PARADEV_*` terminal variables alongside the existing `SUPERSET_*` variables, runs bundled skills as `paradev:<skill>` (with `superset:<skill>` still working as an alias), and uses `paradev` as the CLI name while keeping `superset` as an alias for existing scripts.

— Full rename detailed with backward-compatible aliases and variable names.

desktop-v2.0.12

---

03 **Structured observability for cloud boxes** Not checked yet NEW

65

Adds structured observability for cloud boxes: sends a cloud box's console lines to Sentry Logs as structured templates and parameters, and emits boot phases and memguard actions as typed box events for structured host-service monitoring.

— Detailed mechanism but no reader-facing action. desktop-v2.0.12

THINNER COVERAGE BELOW

05 **Mobile and web platform expansion** Not checked yet NEW

55

Paradev now supports Android and web in addition to existing platforms, with an Android app available on Google Play and a mobile Home list that combines several hosts and Cloud in one view, including a Needs me section.

— Lacks detail on feature parity across the new platforms. cli-v2.0.12 desktop-v2.0.12

---

06 **Plugin routes via tRPC for cloud workspaces** Not checked yet NEW

45

Lets a cloud workspace call every plugin route via tRPC.

— Names the mechanism but no usage example. desktop-v2.0.12

---

08 **AWS Marketplace SaaS contract fulfillment** Not checked yet NEW

40

Adds AWS Marketplace SaaS contract fulfillment, letting organizations subscribe to Paradev directly through AWS Marketplace.

— States the integration exists without setup steps. desktop-v2.0.12

---

09 **Gemini 4 Carbon and Hydrogen in Antigravity picker**

Not checked yet NEW

40

Adds Gemini 4 Carbon and Hydrogen to the Antigravity model picker.

— Names the models but gives no usage guidance. desktop-v2.0.12

---

10 **MCP plugin session persistence** Not checked yet IMPROVED

40

Keeps a plugin MCP session alive between an agent's successive call-tool runs in the CLI, avoiding repeated session setup overhead.

— Explains the benefit without commands or config. desktop-v2.0.12

11 **Sparkline UI component for billing data** Not checked yet NEW

35

Adds a Sparkline UI component for displaying time-series data, used for GB-hours per billing day for organizations or users.

— Describes the use case but not interaction. desktop-v2.0.12

WAS THIS USEFUL? 👍 👎

Diagram Design

Feedback ⓘ changes since 2026-10-27 CODE ↗

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

Diagram Design adds box plot charts, an arc diagram variant, PlantUML and Mermaid quadrantChart import, SBGN process description, requirement traceability, event modeling, st

Diagram Design adds a box plot variant of Scatter for five-number summaries, with Tukey whiskers to the last inlier and outliers outside the fence, and an arc diagram variant for skip-span imports. Not checked yet

Diagram Design adds import support for PlantUML sequence and class diagrams, redrawing them as native diagrams, and for Mermaid `quadrantChart` blocks, redrawing them as quadrant diagrams. Not checked yet

Diagram Design adds three new diagram types: a steps diagram (an ordered stage rail), a unit grid diagram, and verified schedule network diagrams. Not checked yet

Diagram Design adds a geometry verifier inside the skill package, a self-check that flags boxed text overflowing the width budget, and a builder that packages a Claude Desktop skill ZIP under the upload file limit. Not checked yet

WAS THIS USEFUL? 👍 👎

Token Optimizer

Feedback ⓘ v5.13.79 NOTES ↗ RANK

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

> Token Optimizer's release adds a 300K-token compact trigger for 1M-context models with an undoable approval flow, while also hardening background dashboard rebuilds and paused-session refresh timing.

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

01 **300K compact threshold for 1M-context models** Not checked yet

NEW 85

The `/token-optimizer:next` command now surfaces a 300K-token compact trigger for 1M-context models, offering it as a one-approval safe saving action with undo, ranked against your history to show when it pays off.

On a 1M-context model, surface the 300K compact threshold as a reviewable, undoable saving action ranked against your last 30 days of history.

```
$ /token-optimizer:next
```

— Names exact command, threshold and approval/undo mechanism. v5.13.79

THINNER COVERAGE BELOW

02 **Background dashboard rebuild resource caps** Not checked yet

IMPROVED 50

Background dashboard rebuilds now cap themselves at 10 minutes or 1.5 GB of memory and wait a day before retrying, preventing swap pressure on large histories.

— Concrete limits given but no user-facing action. v5.13.79

03 **Paused session refresh timing reliability** Not checked yet

IMPROVED 30

Paused sessions now refresh close to cache expiry even during an upgrade or while another session is saving, keeping keep-warm timing reliable.

— Describes behavior change but no concrete surface or command. v5.13.79

WAS THIS USEFUL? 👍 👎

Meta Astryx

Feedback ⓘ v0.7.17 NOTES ↗ RANK

Astryx is Meta's open-source React design system with 150+ accessible components, page templates, seven themes and a CLI that coding agents can read and drive.

> Astryx's run centers on performance and native-feeling touch interaction: Table gained column/row virtualization, windowing stability, and infinite scroll, a new family of low-level gesture and animation primitives (pan gestures, drag-to-snap-points, direct animators) landed, and dozens of components picked up platform-matching touch, drag, and keyboard-hint behavior. The CLI added icon fallback chains with doctor diagnostics, MCP tools for every command, and discovery scaffolding, alongside several breaking changes (anchored-layer entrance animations, menu press timing, Dialog's default backdrop blur, InfoTip's package move, and locale-driven hour formats).

└──▷ **WHAT SHIPPED · 57 FEATURES** most completely described first ? what's the number?

01 **Structured entry-error reporting for inputs** Not checked yet NEW

96

`NumberInput`, `DateInput`, `DateTimeInput`, and `TimeInput` take `entryErrorBehavior` (with FormLayout's `defaultEntryErrorBehavior` and the `EntryErrorBehavior` type naming `'complete'` and `'blur-or-enter'`), firing `onEntryErrorChange` with structured error objects: `NumberInputEntryError` / `TimeInputEntryError` carry `kind` (`unreadable`, `fractional`, or `outOfRange` for NumberInput; `unreadable` or `outOfRange` for TimeInput) plus the typed `text` and, where applicable, the number or ISO time as `value`; `DateInputEntryError` carries `kind` (`unreadable`, `incomplete`, `nonexistent`, `outOfRange`, or `unavailable`), the typed `text`, and for `outOfRange` / `unavailable` an ISO `date` — pass an `error` status with a message to override the field's built-in rejection copy.

Show a custom rejection message in a NumberInput by reading the structured error — including the specific kind and the out-of-range value — before it reaches the user.

```javascript
import { NumberInput } from '@astryxdesign/core';
import type { NumberInputEntryError } from '@astryxdesign/core';

<NumberInput
  label="Quantity"
  min={1}
  max={99}
  onEntryErrorChange={(error: NumberInputEntryError | null) => {
    if (error?.kind === 'outOfRange') {
      setErrorMessage(`${error.value} is outside the allowed
range 1–99.`);
    } else if (error?.kind === 'fractional') {
      setErrorMessage('Whole numbers only.');
    } else {
      setErrorMessage(null);
    }
  }}
/>
```

Show a custom rejection message when a user types an out-of-range or unavailable date, using the structured error kind to pick your own copy.

```typescript
import { DateInput, DateInputEntryError } from
'@astryxdesign/core';

function MyDateField() {
  const [errorMsg, setErrorMsg] = React.useState(null);

  function handleEntryErrorChange(err: DateInputEntryError |
null) {
    if (err === null) { setErrorMsg(null); return; }
    if (err.kind === 'outOfRange') setErrorMsg(`${err.date} is
outside the allowed range.`);
    else if (err.kind === 'unavailable') setErrorMsg(`${err.date}
is not available.`);
    else if (err.kind === 'nonexistent') setErrorMsg('That date
does not exist.');
    else setErrorMsg('Please enter a complete, valid date.');
  }

  return (
    <DateInput
      onEntryErrorChange={handleEntryErrorChange}
      status={errorMsg ? { kind: 'error', message: errorMsg } :
undefined}
      min="2024-01-01"
      max="2024-12-31"
    />
  );
}
```

— Full error-kind taxonomy and callback shape with runnable examples. v0.7.16 v0.7.15

02 **Table column and row virtualization** Not checked yet NEW

95

Table gains independent column windowing via useTableColumnsVirtualization({initialViewportWidth}) and combined row-and-column windowing via useTableRowsVirtualization({rowHeight, initialViewportHeight}) passed together as a list to `Table.virtualization`, mounting only unpinned columns near the viewport (plus two overscan columns per side) or cells near the viewport on both axes while preserving full `aria-colindex` / `aria-colcount`. `TableRowContext` (carrying `logicalIndex`, row/header counts, row model) and `TableRenderContext` (carrying a stable `viewport` reference) are passed to body and scroll-wrapper transforms for table plugins, and useTableRowIndex() numbers rows from context by default. Column windowing keeps focus and scroll position stable, windowed tables keep row selection and row numbers active together, and the

`useTableRowsVirtualization` option `viewportHeight` is renamed `initialViewportHeight` (now only sizing the first server-side window, no longer setting the Table's height).

Virtualise a large table's columns to avoid mounting thousands of cells — only columns near the viewport are rendered, while pinned columns and ARIA attributes stay correct.

```javascript
import { Table, useTableColumnsVirtualization } from
'@astryxdesign/core';

const columnsVirtualization = useTableColumnsVirtualization({
initialViewportWidth: 1200 });

<Table
  columns={columns}
  rows={rows}
  virtualization={columnsVirtualization}
/>
```

Virtualize a large flat table so only visible rows mount, reducing DOM size for high-row-count datasets.

```javascript
import { Table, useTableRowsVirtualization } from
'@astryxdesign/core';

const virtualization = useTableRowsVirtualization({
  rowHeight: 48,
  initialViewportHeight: 600,
});

<div style={{ height: 600, overflow: 'auto' }}>
  <Table columns={columns} data={rows} virtualization=
{virtualization} />
</div>
```

— Full mechanism, hook signatures, and a breaking rename named. v0.7.16 v0.7.12

---

03 **Icon set fallback chains and diagnostics** Not checked yet NEW

95

Pick a main icon set and ordered backup sets with `astryx experimental icon use <main> --backup <set>`, writing `experimental.iconSet` as a prioritized list; icons missing from every set become a `MissingIcon` placeholder and are reported in `icons.placeholder`. The `icon-sets` and `icon-placeholders` doctor checks

validate health — `icon-sets` fails when a listed set is not installed, withdrawn, or duplicated, and `icon-placeholders` warns while copied code still uses the placeholder, naming each file and meaning; the active icon-set order, each set's `role` and `rank`, and a prompt to run `experimental icon --list` are inspectable when multiple sets are installed with no main set designated. `integration check` now detects undeclared template icon imports as `template_icon_undeclared`, failing with `--strict --next`.

Register a primary icon set with one ordered fallback so templates always resolve icons, even when the main set is missing a glyph.

```
$ astryx experimental icon use material --backup phosphor
```

Catch template icons that are imported but not declared before merging, so strict CI pipelines fail on undeclared icon usage.

```
$ astryx integration check --strict --next
```

— Exact commands, flags, and doctor checks with runnable examples. v0.7.13

04

CodeBlock copy feedback, custom languages, and line numbers

Not checked yet NEW 95

`CodeBlock` gains an onCopyError(error) prop that fires once per failed clipboard write (showing a cross icon and announcing `@astryx.codeBlock.copyFailed` for 2 seconds), a `startLineNumber` prop so excerpts can display real file line numbers, an `endContent` prop for extra buttons beside Copy, and `createCodeLanguage` (from `@astryxdesign/core/CodeBlock`) for custom syntax highlighting. It now highlights C, C++, SQL, Rust, Kotlin, Swift, and Go with no setup, with nine more languages available as on-demand imports from `@astryxdesign/core/CodeBlock/languages/`.

Log clipboard errors silently when a user's browser blocks the copy action, without exposing the raw error in the UI.

javascript

```
import { CodeBlock } from '@astryxdesign/core';

<CodeBlock
  code={snippet}
  onCopyError={(error) => console.error('Copy failed:', error)}
/>
```

Extend CodeBlock with a custom language tokenizer — useful when your project uses a DSL or niche language not in the built-in set.

```javascript
import {createCodeLanguage} from '@astryxdesign/core/CodeBlock';

const zig = createCodeLanguage({name: 'zig', identifiers:
['zig'], tokenize});

<CodeBlock code={source} language={zig} />
```

— Every prop and import path named with runnable examples. v0.7.11 v0.7.8 v0.7.5

05 **Low-level gesture and animation primitive APIs** Not checked yet

NEW 95

`createDragToSnapPoints` (from `@astryxdesign/core` root) and `useDragToSnapPoints` (from `@astryxdesign/core/hooks`) let any element drag along one axis and settle on named snap points, with `flickThreshold` (default 1008 px/s), settleTo(index), and onSettle(index) callbacks. createPanGesture(element, {axis?, onPan}) and usePanGesture(ref, {axis?, onPan}) report `{state, translation, velocity}` across touch, mouse, and pen, with tracking limited to one axis via `axis: 'x'` or `axis: 'y'`. createAnimator(element) and useAnimator(ref) (from the package root and `@astryxdesign/core/hooks` respectively) support `animate`, `decay`, `set`, `get`, `getVelocity`, and `stop` for direct element animation.

Track a user's drag on a card element and update its position, limiting the gesture to horizontal movement only.

```javascript
import { createPanGesture } from '@astryxdesign/core';

const card = document.getElementById('card');
const pan = createPanGesture(card, {
  axis: 'x',
  onPan({ state, translation, velocity }) {
    if (state === 'changed') {
      card.style.transform = `translateX(${translation.x}px)`;
    }
    if (state === 'ended') {
      console.log('release velocity:', velocity.x);
    }
  },
});
// later: pan.destroy();
```

Animate a draggable card's position and opacity with spring physics, then coast it on release — sharing one element ref between the animator and a gesture hook.

```javascript
import { createAnimator } from '@astryxdesign/core';
import { useAnimator } from '@astryxdesign/core/hooks';
import { useRef } from 'react';

function DraggableCard() {
  const ref = useRef(null);
  const animator = useAnimator(ref);

  function onRelease(velocityX, velocityY) {
    animator.decay({ x: velocityX, y: velocityY });
  }

  function onReset() {
    animator.animate({ x: 0, y: 0, opacity: 1 });
  }

  return <div ref={ref} onPointerUp={() => onReset()}>Drag
me</div>;
}
```

— Three full API signatures with runnable examples. v0.7.10 v0.7.9 v0.7.8

06 **Drag-and-drop row reordering for List** Not checked yet NEW

92

Adds drag-and-drop row reordering to List via createReorderable({onReorder, isReorderable}) passed as `reorderable` — each committed move calls onReorder({key, to}), and start(key) enters keyboard reorder mode from app UI such as a context menu — plus a tap-only reorder hint anchored to the drag handle in keyboard-hint layer and theme roles, showing move buttons and a Done button so rows can be reordered without a drag.

Let users drag rows into a new order in a List, with keyboard reorder mode available via a context menu.

```javascript
import { List, ListItem, createReorderable } from
'@astryxdesign/core';

const reorderable = createReorderable({
  onReorder: ({ key, to }) => moveItem(key, to),
  isReorderable: (key) => !lockedKeys.has(key),
});

<List reorderable={reorderable}>
  {items.map(item => (
    <ListItem key={item.id} itemKey={item.id}>{item.label}
</ListItem>
  ))}
</List>
```

— Full API signature, keyboard mode, and a runnable example. v0.7.12

07

Touch-keyboard hint props across text inputs

Not checked yet

NEW

92

`TextInput` and `TextArea` accept `enterKeyHint`, `inputMode`, `autoCapitalize`, `hasAutoCorrect`, and (TextInput only) `hasSpellCheck` to control touch-keyboard behaviour, with `TextInput`'s `type` now also accepting `'search'`, `'tel'`, and `'url'` (value types `EnterKeyHint`, `InputMode`, `AutoCapitalize`). `ChatComposerInput` gains its own `enterKeyHint` prop (typed by `EnterKeyHint`) — set `enter` to insert a line break instead of sending. Search inputs in `Typeahead`, `Tokenizer`, `PowerSearch`, `CommandPaletteInput`, `Selector`, and `MultiSelector` now carry `enterkeyhint="search"` and `autocorrect="off"` to show a Search key on touch keyboards and suppress autocorrection.

Show a `Send` key on iOS touch keyboards in a standard ChatComposerInput, or override to `Enter` for a newline-on-Enter composer.

```javascript
import { ChatComposerInput } from '@astryxdesign/core';

// Default: shows Send key (enterkeyhint="send" is applied
automatically)
<ChatComposerInput />

// Override: show Return/Enter key for a newline-inserting
composer
<ChatComposerInput enterKeyHint="enter" />
```

Tell a mobile keyboard to show a search action key and a URL keypad for a search field, so users get the right keyboard layout and return-key label automatically.

```tsx
import { TextInput } from '@astryxdesign/core';

<TextInput
  label="Search"
  value={query}
  onChange={setQuery}
  enterKeyHint="search"
  inputMode="search"
/>
```

— Every prop and type union named with runnable examples. v0.7.7 v0.7.6

---

08 **CLI component and hook discovery additions** Not checked yet NEW

90

`astryx discover` now lists the experimental `TabBarMoreList` component from `@astryxdesign/lab`. Copy example blocks — `useCalendarDays — Month Grid`, `useIcon — Status Row`, `useLinkComponent — Article Cards`, and `useOverlay — Product Quick View` — via `astryx template useCalendarDaysHookUsage <path>`, `astryx template useIconHookUsage <path>`, `astryx template useLinkComponentHookUsage <path>`, and `astryx template useOverlayHookUsage <path>`; `astryx hook useCalendarDays`, `astryx hook useIcon`, `astryx hook useLinkComponent`, `astryx hook useOverlay`, and `astryx search` now return these hooks.

Scaffold the useOverlay quick-view example into your project to start building hover-revealed overlays on card components.

```
$ astryx template useOverlayHookUsage
./src/components/ProductQuickView
```

— Every command and hook name given, with a runnable example. v0.7.13 v0.7.9

---

09 **Per-component icon slots** Not checked yet NEW

90

Define per-component icon slots in a theme with defineTheme({componentIcons}); the first slot is `code-block-copy`, accepting a shared icon name, `null`, a partial `{icon, variant, size}` request, or a state function, readable back via `theme.componentIcons` (`ComponentIconSlotMap` and slot value types export from `@astryxdesign/core/Icon`). `astryx theme build` carries `componentIcons` slot functions through, writing static slot values as JSON beside each slot import, with `-`

`-icons-specifier` rewriting only the icon registry import and failing the build if a slot function is defined directly in the theme file; `astryx component` also lists a component's icon slots via its new `slots` output.

Override the icon shown in CodeBlock's copy button — hiding the glyph after copy — using the new `componentIcons` theme slot.

javascript

```
import { defineTheme } from '@astryxdesign/core';

export default defineTheme({
  componentIcons: {
    'code-block-copy': ({ copied }) => copied ? null : { icon:
'clipboard' },
  },
});
```

— Named theme key, CLI flag, and an example. v0.7.9

10 **Theme motion tempo and character knobs** Not checked yet NEW

90

A theme's `motion` object accepts `tempo` (0.75–1.5) and `character` (`'crisp'`, `'standard'`, or `'expressive'`) knobs, setting the `--ease-entrance` and `--ease-move` CSS tokens (`ThemeMotionConfig`, `ThemeMotionKnobs`, `MotionCharacter` are the exported types); spring-based touch motion timing also scales with `tempo` in addition to CSS durations, so each spring keeps its shape and runs `tempo` times as long, as does the animated-layout glide.

Give a theme a faster, crisper motion feel without touching individual duration tokens.

tsx

```
import { ThemeProvider } from '@astryxdesign/core';
import type { ThemeMotionConfig } from '@astryxdesign/core';

const motion: ThemeMotionConfig = {
  tempo: 0.8,
  character: 'crisp',
};

<ThemeProvider motion={motion}>
  <App />
</ThemeProvider>
```

— Named config object, range, and CSS tokens with an example. v0.7.9 v0.7.6

11 **Tokenizer inside InputGroup** Not checked yet NEW 88

`Tokenizer` is now InputGroup-compatible: inside an `InputGroup` it consumes `InputGroupContext`, merges `aria-describedby` via `getInputARIA`, adopts `groupStyles.inGroup` border treatment, and caps or summarizes overflow tokens with `maxEntries` or `tokenOverflowBehavior`.

Place a Tokenizer inside an InputGroup so the group owns the label, description, size, and validation — cap visible tokens with `maxEntries` to keep the row fixed-height.

jsx

```jsx
import { InputGroup, Tokenizer, TextField } from
'@astryxdesign/core';

<InputGroup label="Recipients" size="medium">
  <Tokenizer
    label="Add email"
    items={suggestions}
    maxEntries={5}
    onTokenAdd={handleAdd}
  />
  <TextField label="Subject" />
</InputGroup>
```

— Names every integration point and config prop with an example. v0.7.14

12 **Animated layout transitions** Not checked yet NEW 86

Pass useAnimatedLayout({containerRef}) as `animatedLayout` to `List` (with `itemKey` set on each `ListItem`) so inserted, removed, and reordered rows animate in place; the same handle can be passed to `Collapsible`, `TreeList`, and `Table` row expansion to animate their expand and collapse transitions.

Add animated insert/remove/reorder transitions to a list so rows glide to their new positions on data changes.

```javascript
import { List, ListItem, useAnimatedLayout } from
'@astryxdesign/core';

const containerRef = useRef(null);
const animatedLayout = useAnimatedLayout({ containerRef });

<div ref={containerRef}>
  <List animatedLayout={animatedLayout}>
    {items.map(item => (
      <ListItem key={item.id} itemKey={item.id}>{item.label}
</ListItem>
    ))}
  </List>
</div>
```

— Hook signature and three consuming components, with an example. v0.7.12 v0.7.5

13 **Popover touch presentation and entrance animation** Not checked yet

NEW 86

Popover grows out of its trigger (wide before tall) under a touch pointer, animating with the theme's `--duration-medium-min` and cross-fading at 200 ms or less under reduced motion; it also takes a `presentation` prop accepting `'popover'`, `'bottom-sheet'`, or `'adaptive'`, where `'adaptive'` switches to a sheet on viewports narrower than the theme's `md` breakpoint (768px by default) or shorter than 500px.

Make a Popover drop into a bottom sheet automatically on small screens — useful for menus and selectors that should feel native on phones without a separate mobile component.

```javascript
<Popover label="Settings" content={<SettingsContent />}
presentation="adaptive">
  <Button label="Settings" />
</Popover>
```

— Named prop values, breakpoint default, and an example. v0.7.11 v0.7.5

14 **Adaptive bottom-sheet presentation for PowerSearch and Tokenizer**

Not checked yet NEW 86

Give `PowerSearch` a touch-first sheet flow by setting its `presentation` prop to `'popover'`, `'bottom-sheet'`, or `'adaptive'` (default), typed by the exported `PowerSearchPresentation`, where `'adaptive'` activates the bottom-sheet flow on compact coarse-pointer screens. `Tokenizer`'s `presentation` prop is stabilised with the same values (default `'adaptive'`), typed by the exported `TokenizerPresentation`, switching to a bottom-sheet draft on phones.

Pin PowerSearch to the popover surface on every device to preserve pre-0.7.8 behaviour while you evaluate the new bottom-sheet flow.

```javascript
import { PowerSearch } from '@astryxdesign/core';

<PowerSearch
  presentation="popover"
  filters={filters}
  onChange={setFilters}
/>
```

— Exact prop values and defaults for both components. `v0.7.8`

---

15 **Dialog backdrop blur removed by default** `Not checked yet` `BREAKING`

86

`Dialog`, `AlertDialog`, `CommandPalette`, `Drawer`, and `MobileNav` modals no longer apply a 2px backdrop blur by default — themes relying on it must add `backdropFilter: 'blur(2px)'` explicitly on `components.dialog.base['::backdrop']`, `components['mobile-nav'].base['::backdrop']`, or `components.drawer.base[':modal::backdrop']`.

Restore the backdrop blur on Dialog modals after it was removed as a default, using a theme override.

```javascript
const myTheme = {
  components: {
    dialog: {
      base: {
        '::backdrop': {
          backdropFilter: 'blur(2px)'
        }
      }
    },
    'mobile-nav': {
      base: {
        '::backdrop': {
          backdropFilter: 'blur(2px)'
        }
      }
    },
    drawer: {
      base: {
        ':modal::backdrop': {
          backdropFilter: 'blur(2px)'
        }
      }
    }
  }
};
```

— Exact theme keys and a runnable restoration example. v0.7.7

16 **Incremental search results via SearchContext** Not checked yet

NEW 86

`SearchSource` now supports incremental results via a new context.update(items) call inside search(query, context), using the new exported `SearchContext` type, so `BaseTypeahead`, `Typeahead`, and `Tokenizer` can show local matches immediately while a server request is in flight.

Show cached results instantly while a server search is still in flight, so the typeahead feels responsive on slow connections.

```tsx
import { Typeahead } from '@astryxdesign/core';
import type { SearchContext } from '@astryxdesign/core';

const source = {
  async search(query: string, context: SearchContext) {
    context.update(cachedPeople.filter(p =>
p.name.includes(query)));
    const serverResults = await fetchPeople(query);
    return serverResults;
  },
};

<Typeahead label="Assignee" source={source} debounceMs={0} />
```

— Named callback, type, and consuming components with example. v0.7.6

17 **disabledMessage tooltip for ComplexSelector** Not checked yet NEW

84

`ComplexSelector` gains a `disabledMessage` prop to explain why the selector is disabled, surfaced as a tooltip on hover and keyboard focus via `aria-describedby`, compatible with both the built-in trigger and `renderTrigger`.

Show users why a ComplexSelector is disabled, with a tooltip that appears on hover and focus and is linked via `aria-describedby`.

```jsx
import { ComplexSelector } from '@astryxdesign/core';

<ComplexSelector
  isDisabled
  disabledMessage="You need edit permission to change this
filter."
  items={items}
  onSelectionChange={handleChange}
/>
```

— Named prop, aria wiring, and a runnable example. v0.7.14

18 **Menus open on press instead of long-press** Not checked yet

84

BREAKING

An anchored `DropdownMenu` or `MoreMenu` and `useMenuPress-driven` menus now open on a finger or pen press instead of after a 500 ms hold. Pass

`longPressDelayMs: 500` to `useMenuPress` to keep the previous hold behaviour, or pass `presentation="popover"` to a DropdownMenu/MoreMenu to keep the previous tap-on-lift and 500 ms hold behaviour.

— Exact flags to restore old behaviour, fully actionable. v0.7.14

---

19 **Timezone-pinned Timestamp** Not checked yet NEW 84

Render absolute times in a named IANA zone with the new `timezoneID` prop on Timestamp (for example `'UTC'` or `'America/Los_Angeles'`), eliminating server/client hydration mismatches; pass `'local'` or omit it to keep the viewer's zone.

Pin a server-rendered event time to UTC so it hydrates identically regardless of the server's local timezone.

javascript

```
import { Timestamp } from '@astryxdesign/core';

<Timestamp value={eventDate} presentation="absolute"
timezoneID="UTC" />
```

— Exact prop, values, and a runnable example. v0.7.11

20 **Carousel snap paging** Not checked yet NEW 84

Enable iOS-style paging on `Carousel` by passing useSnapPaging() (imported from `@astryxdesign/core/SnapPaging`) to the `hasSnap` prop; set useSnapPaging({pages: 'item'}) to flick one item at a time.

Add item-by-item flick paging to a Carousel so each swipe advances exactly one card on iOS and trackpad.

```javascript
import { Carousel } from '@astryxdesign/core';
import { useSnapPaging } from '@astryxdesign/core/SnapPaging';

function CardCarousel({ cards }) {
  const snapPaging = useSnapPaging({ pages: 'item' });
  return (
    <Carousel hasSnap={snapPaging}>
      {cards.map(card => <Carousel.Item key={card.id}>
{card.content}</Carousel.Item>)}
    </Carousel>
  );
}
```

— Exact import, prop, and option with a runnable example. v0.7.10

21 **LayoutPanel collapse animation and resize resistance**

Not checked yet IMPROVED 82

`LayoutPanel` now slides its pane closed and open from the edge facing its `ResizeHandle` on collapse and expand (triggered via collapse(), expand(), Enter, double-click, the grow arrow, controlled `isCollapsed`, or a lift past the collapse threshold), and `ResizeHandle` now shows spring resistance of up to 8 px when dragged past a region's minimum or maximum, collapsing a region only on pointer lift past `collapsedSize` rather than mid-drag.

— Named methods and thresholds, no standalone example. v0.7.12

22 **useTooltip show() and hide() controls** Not checked yet NEW 81

Control tooltip visibility programmatically via new show() and hide() methods returned by `useTooltip`, enabling toggle-style triggers with `aria-expanded` and `aria-controls`.

Build a toggle-style More info button that opens and closes a tooltip on successive presses, with correct ARIA attributes.

```javascript
const { tooltipProps, triggerProps, describedBy, show, hide } =
useTooltip({ content: 'This field is required by policy.' });

function InfoToggle() {
  const [open, setOpen] = useState(false);
  return (
    <>
      <button
        {...triggerProps}
        aria-expanded={open}
        aria-controls={describedBy}
        onClick={() => { open ? hide() : show(); setOpen(o =>
!o); }}
      >
        More info
      </button>
      <Tooltip {...tooltipProps} />
    </>
  );
}
```

— Named methods and ARIA wiring with a runnable example. v0.7.8

23 **InfoTip promoted to core** Not checked yet BREAKING 81

`InfoTip` is promoted from `@astryxdesign/lab` to `@astryxdesign/core`, importable from `@astryxdesign/core/InfoTip` or the package root — an info-icon button that pins help on click, tap, Enter, or Space and dismisses on Escape or outside press. The `@astryxdesign/lab` export is now a deprecated re-export of the Core component and will be removed in the next release.

Use InfoTip from core (not lab) to pin contextual help next to a form field.

```javascript
import { InfoTip } from '@astryxdesign/core';
// or: import InfoTip from '@astryxdesign/core/InfoTip';

<InfoTip>Your username must be unique across all workspaces.
</InfoTip>
```

— Exact import paths and deprecation timeline, with example. v0.7.7

24 **Static page render command** Not checked yet NEW 81

Render a page as a self-contained HTML file with no scripts or network calls using `astryx experimental render <page>`, and reverse it back to source with `--reverse`.

— Exact command and flag named, little on output format. v0.7.5

25 **DropdownMenuSubMenu hover timing** Not checked yet NEW

78

Control hover-open and hover-close timing on `DropdownMenuSubMenu` with the new `delay` and `hideDelay` props (in milliseconds), matching the same options `Tooltip`, `HoverCard` and `TopNavMenu` already accept.

Show a submenu after a longer hover pause and keep it open briefly after the pointer leaves — useful for dense navigation menus where accidental hover is common.

```javascript
import { DropdownMenuSubMenu } from '@astryxdesign/core';

<DropdownMenuSubMenu delay={300} hideDelay={400}>
  {/* submenu items */}
</DropdownMenuSubMenu>
```

— Exact props and a runnable example. v0.7.11

26 **Keyboard hint no longer auto-dismisses by default** Not checked yet

75

BREAKING

`useKeyboardHint`'s `dismissAfterMs` no longer defaults to 3000; leave it unset so the arrow-key hint in `Toolbar`, `TabList`, and `SegmentedControl` stays until the user acts, or pass `dismissAfterMs: 3000` explicitly to restore the previous auto-dismiss.

— Exact option and migration value named. v0.7.14

27 **Lightbox swipe-to-dismiss and paging** Not checked yet NEW

75

On touch devices, Lightbox adds a physics-driven swipe-down-to-dismiss gesture (the photo follows the finger, the backdrop fades, and releasing while moving down carries the photo off screen) and pages between items the way the iOS Photos viewer does, with a 40 px gap between items, edge-aware zoom panning, per-item captions, and `onIndexChange` reporting swipes.

— Detailed mechanism and callback named, no code example. v0.7.12

28 **Locale-aware hour format default** Not checked yet BREAKING

75

`TimeInput` and `DateTimeInput`'s `hourFormat` now defaults to the `InternationalizationProvider` locale's hour cycle instead of always `'12h'`; apps that relied on the `'12h'` default must now pass `hourFormat="12h"` explicitly.

— Exact prop and migration value to restore old default. v0.7.6

---

29 **CheckboxList optional and required states** Not checked yet NEW

73

`CheckboxList` now accepts `isOptional`, showing the optional indicator after the group label, and `isRequired`, showing the required indicator on the group label and setting native `required` on the first enabled, non-read-only checkbox until any box is checked — consistent with `RadioList` and `CheckboxInput`.

— Named props and exact native behaviour, no example. v0.7.8

---

30 **Infinite scroll for Table** Not checked yet NEW

72

Table gains infinite loading via useTableInfiniteScroll({hasMore, isLoading, onLoadMore}), coordinating automatic and manual loading with a single Promise request latch.

— Names the hook and options but no usage example. v0.7.16

---

31 **Anchored layer entrance animation rework** Not checked yet BREAKING

71

Anchored layers (`Popover`, `DropdownMenu`, `ContextMenu`, `Selector`, `ComplexSelector`, `Tooltip`, `HoverCard`, `Breadcrumbs`, `PowerSearch` menus) now slide and scale in from the side the layer lands on, with `layerAnimations` styles also setting `transform-origin`; there is no option to keep the old entrance animation, so it replaces the previous behaviour unconditionally.

— Names every affected component but offers no migration path. v0.7.14

---

32 **Readable Tailwind status color shades** Not checked yet BREAKING

69

In Tailwind, `text-<status>` and `border-<status>` now use readable status shades; write `text-[var(--color-warning)]` to keep the old fill color on text.

.
.
.
.

— Exact utility classes and override, with before/after screenshots. v0.7.5

---

33 **AppShell safe-area handling** Not checked yet NEW

68

`AppShell` now keeps header items and content inside the device safe area when a page opts into `viewport-fit=cover`, handling the status bar, Dynamic Island,

landscape notch, and home indicator via `contentPadding` plus the bottom inset.

— Named config and device cases, no standalone example. v0.7.7

34 **useAnnounce replace-only mode** Not checked yet NEW 67

`useAnnounce` now accepts an optional `{hasReplace: true}` option to create an instance that replaces its own unspoken messages, so only the latest announcement is delivered for high-frequency updates such as barcode scanning or dictation.

— Named option and use cases, no code example. v0.7.6

35 **FieldStatus tooltip variant deprecated** Not checked yet DEPRECATED 67

`FieldStatus variant="tooltip"` is deprecated and will be removed in the next minor release; replace with `statusVariant="tooltip"` on the field control instead.

— Exact replacement prop named for migration. v0.7.6

36 **Typeahead token clearing with hasClear** Not checked yet NEW 65

Typeahead now clears the selected token's value on Backspace or Delete (calling onChange(null), announcing the cleared value, and moving focus to the empty input) when `hasClear` is active; without `hasClear`, the keys start editing as before.

— Names the flag and exact keyboard behaviour. v0.7.13

37 **Deprecated component doc metadata** Not checked yet NEW 65

Mark components as deprecated in doc metadata with `deprecated: {replacement, description}`; `astryx search` results, `astryx build` kit entries, and the docsite all surface the `deprecated` field, badge, and replacement link.

— Names the metadata shape and surfaces but no example. v0.7.8

38 **Touch and pointer interaction parity across form controls**

Not checked yet IMPROVED 64

`SegmentedControl` lifts the selected segment's surface under a finger or pen press matching UIKit behaviour and settles on release; commits the segment under the pointer when a drag drifts off the control (clamped to the end segments, Escape or a browser cancel cancels the press); and tracks touch and pointer drags so a drag from the selected segment carries the selection and commits on release, firing `onChange` once at lift, with UIKit's spring under touch and theme duration/easing under mouse or pen. Switch now tracks touch and pointer drags — a horizontal drag past the flip point toggles at release, with the thumb lifting under a soft shadow on touch screens, mirroring iOS. `ToggleButtonGroup` tracks a finger's or pen's press across its toggles under a coarse primary pointer, changing only the toggle where the user lifts, with a 44 px touch target

drawn by the group. `CheckboxInput` and `RadioList` options now paint immediately on `pointerdown` and fade on release for mouse users on fine-pointer devices, matching the existing finger-press behaviour. `FieldLabel` also keeps a tap on the info icon (`labelTooltip`) from toggling or focusing the labelled control on touch screens.

— Mechanism described per control but no props to act on. v0.7.15 v0.7.14 v0.7.12 v0.7.7 v0.7.6

39 **Icon variants per theme** Not checked yet NEW 64

Themes can now select icon artwork by variant and set glyph size per Icon size using `iconVariants` in `defineTheme` and the `variant` prop on `<Icon>`.

— Names the config key and prop, no example. v0.7.5

40 **Lightbox pinch and trackpad zoom** Not checked yet NEW 62

Lightbox now supports pinch-to-zoom and trackpad-pinch zoom, trackpad-scroll pan of zoomed photos, and double-tap zoom to fill the screen — controlled by the existing `hasZoom` prop.

— Names the controlling prop but no standalone example. v0.7.10

41 **Search option IDs keyed by item id** Not checked yet BREAKING 62

`SearchSource` / Typeahead / Tokenizer option element ids now follow the item's `id` instead of its row index; tests that select options by an index-based id must be updated. A grouped Tokenizer with `hasCreate` now also shows its Create row after the groups instead of first.

— Names the behaviour change but no migration helper. v0.7.6

42 **Native-feeling picker wheels for date/time inputs** Not checked yet IMPROVED 60

`TimeInput` and `DateTimeInput` now loop the minute wheel under a finger, so scrolling past 59 wraps to 00 without changing the hour, matching UIDatePicker behaviour. Picker wheels in `TimeInput`, `DateTimeInput`, and `DateInput` bottom sheets also now behave like UIPickerView on coarse-pointer devices, with UIKit-style deceleration, rubber-band edges, haptic ticks, and a semibold accent row on the band, while a mouse still uses the scrolling column.

— Behavior detailed but no prop to configure it. v0.7.14 v0.7.8

THINNER COVERAGE BELOW

43 **Native stacking bottom sheets** Not checked yet IMPROVED 57

`BottomSheet` now stacks natively — a sheet opened over another stacks on top, maintains the underlying sheet's state, and shows only one backdrop; `BottomSheetSwitcher` is deprecated but still works.

— Behaviour and deprecation named, no API surface. v0.7.5

44 **TextArea native character limit** Not checked yet NEW 55

`TextArea` now takes `htmlMaxLength` to set the native `maxlength` attribute in UTF-16 code units, independent of the soft `maxLength` counter.

— Names the prop but no usage example. v0.7.16

45 **ProgressBar touch animation** Not checked yet IMPROVED 55

`ProgressBar` now animates value changes linearly over one second per full track on touch devices, running the transition on the compositor, and pulses opacity (full to 84% and back over two seconds) under reduced motion for indeterminate bars.

— Precise timing values but no prop to configure. v0.7.7

46 **Kbd modifier display by platform** Not checked yet IMPROVED 54

`Kbd` now shows `meta` as ⌘ on Apple platforms and as `Win` off Apple platforms, and shows `ctrl`, `alt`, and `shift` as text (`Ctrl` , `Alt` , `Shift`) off Apple platforms, with values sourced from the catalog so other locales can translate them.

— Exact label mapping named, no usage example. v0.7.12

47 **ChatLayout keyboard avoidance** Not checked yet NEW 52

`ChatLayout` now keeps its composer above the software keyboard on Android (Chrome), iOS Safari, and Firefox for Android, tracking keyboard slide animations via the visual viewport API.

— Platforms named but no config surface. v0.7.12

48 **RichTextEditor list markers and link placeholder** Not checked yet IMPROVED 50

`RichTextEditor` can now re-mark a list item with a Markdown block marker, and adds the locale message `@astryx.richTextEditor.link.urlPlaceholder` (`example.com`) as the placeholder for the RichText toolbar's link field.

— Names the locale key but mechanism stays thin. v0.7.17 v0.7.16

49 **MCP tools for every CLI command** Not checked yet NEW 50

The Astryx CLI now exposes MCP tools for every command, letting MCP-aware coding agents invoke the CLI's full command set.

— States the capability but no command list or example. v0.7.17

50 **BaseTypeaheadMenu CLI block** Not checked yet NEW 50

Adds a `BaseTypeaheadMenu` block to the CLI: a BaseTypeahead whose listbox menu shows information around the result list, and whose pick or filter button opens a dialog-role menu on the same layer.

— Names the block but no scaffolding command shown. v0.7.13

---

51 **Lightbox gallery captions** Not checked yet NEW 48

Lightbox gallery captions now sit in a persistent two-line band under every photo, with a `MediaTheme-driven` dark surface and an inline `more` /'less' link for longer captions.

— Describes behavior but no prop names or example. v0.7.16

---

52 **Swipe actions for Item and ListItem** Not checked yet NEW 48

Astryx's `Item` and `ListItem` components now track swipe gestures through the full open, confirm, fire, and cancel motion, matching UIKit-like swipe actions.

— Names the motion states but no API detail. v0.7.17

---

53 **Locale-abbreviated calendar weekday headers** Not checked yet IMPROVED 46

Calendar weekday headers now show locale-abbreviated names (`Sun Mon Tue` in English) where all seven fit in four characters, falling back to short names otherwise; `DateInput` , `DateRangeInput` , `DateTimeInput` , and their touch bottom sheets follow automatically.

— Behaviour described, no config to act on. v0.7.8

---

54 **Focus return after selection in menus and selectors** Not checked yet IMPROVED 44

`TopNavMenu` and `TopNavMegaMenu` now close once after a touch or keyboard selection and return focus to the trigger; `Selector` and `MultiSelector` now return focus after a selection and keep the selected row aligned with the trigger as option lists grow.

— Behavior described across four components, no config surface. v0.7.17

---

55 **DiffViewer large-change fold localisation** Not checked yet NEW 44

Adds three `@astryx.diffViewer.*` localisation keys to the `en` catalog for the lab `DiffViewer` component's large-change fold: its `Large change: N lines` label and its Show full diff / Hide full diff button.

— Names the keys but no component usage shown. v0.7.6

56 **Truncation state while expanded** Not checked yet IMPROVED

40

`useTruncation` now takes `isExpanded` to keep reporting whether text is clamped while it is shown in full.

— One-line description of a single hook option. v0.7.16

---

57 **CLI routing, catalog, and diagnostics improvements**

Not checked yet IMPROVED

27

The CLI gains one-base routing, faster kept catalogs and upgrade scans, clearer theme and integration diagnostics, safer rendering, and page templates whose demo controls work.

— Vague prose with no named flags or commands. v0.7.17

└──▷ BREAKING ON UPGRADE

! `useKeyboardHint`'s `dismissAfterMs` no longer defaults to 3000 — callers that relied on the hint auto-dismissing after 3 seconds must now pass `dismissAfterMs: 3000` explicitly.

! Built-in anchored layers (Popover, DropdownMenu, ContextMenu, Selector, ComplexSelector, Tooltip, HoverCard, Breadcrumbs, PowerSearch menus) have no option that keeps the old entrance animation — the new slide-and-scale entrance from the landing side replaces it unconditionally.

! `useMenuPress` without `longPressDelayMs` now opens on press instead of after a 500 ms hold — pass `longPressDelayMs: 500` to restore the previous behavior.

! An anchored DropdownMenu or MoreMenu now opens on finger press instead of after a 500 ms hold — pass `presentation="popover"` to restore the tap-on-lift and hold behavior.

! The `viewportHeight` option on `useTableRowsVirtualization` is renamed to `initialViewportHeight`; it now only sizes the first server-side window and no longer sets the Table's height — callers must rename the prop and give the Table's parent a height instead.

! The `InfoTip` export from `@astryxdesign/lab` is now a deprecated re-export of the Core component and will be removed in the next release: import it from `@astryxdesign/core` instead.

! Dialog, AlertDialog, CommandPalette, Drawer, and MobileNav modals no longer apply a 2px backdrop blur by default — themes relying on the blur must add `backdropFilter: 'blur(2px)'` explicitly on `components.dialog.base['::backdrop']`, `components['mobile-nav'].base['::backdrop']`, or `components.drawer.base[':modal::backdrop']`.

! TimeInput and DateTimeInput: `hourFormat` now defaults to the `InternationalizationProvider` locale's hour cycle instead of always `'12h'`; apps that relied on the `'12h'` default must pass `hourFormat="12h"` explicitly.

! `SearchSource` / `Typeahead` / `Tokenizer`: option element ids now follow the item's `id` instead of its row index — tests that select options by an index-based id must be updated.

! A grouped `Tokenizer` with `hasCreate` now shows its Create row after the groups instead of first.

! `FieldStatus variant="tooltip"` is deprecated and will be removed in the next minor release; replace with `statusVariant="tooltip"` on the field control instead.

WAS THIS USEFUL? 👍 👎

◆ Speech & Voice AI

LiveKit Agents

Feedback ⓘ livekit-agents@1.9.0 NOTES ↗

RANK

LiveKit Agents builds realtime voice and video AI agents on LiveKit's WebRTC infrastructure, with turn detection, interruption handling, and telephony.

> LiveKit Agents 1.9.0 ships multi-turn answering-machine detection, cross-provider decision-model fallback, and a batch of new STT/TTS plugin integrations spanning SILMA AI, Reson8, Gnani, Zoom, ElevenLabs, AssemblyAI, Sarvam, and Phonic.

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

01 **Multi-turn answering-machine detection (AMD)** Not checked yet

NEW 82

Track calls across voicemail, IVR menus, and hold queues with multi-turn `AMD`; disable any stage's reply by setting its instructions to `None`. Breaking: use opening instructions instead of say() or generate_reply() during detection, do not enter AMD from an active `AgentTask`, and `send_dtmf_events` no longer requests a reply.

Follow a call through voicemail without replying during the screener stage, then act when the stage that needs a response is reached.

```python
async with AMD(session, voicemail_instructions=None) as detector:
    result = await detector.execute()
```

— Names API, stage control, and migration caveats clearly. livekit-agents@1.9.0

02 **Decision-model fallback across LLM providers** Not checked yet

NEW 81

Fall back across LLM providers on decision-model failures using `inference.DecisionModel` with a `fallback` list, or chain adapters with `llm.DecisionModelFallbackAdapter`. Adds `microsoft-eu/decision-1` to the available decision model IDs.

Avoid dropped decision requests by falling back to a secondary model when the primary fails, with a cross-provider chain as a final backstop.

```python
model = inference.DecisionModel(
    'openai/gpt-6-luna-decisions', fallback=['google/gemini-decision-1']
)
model = llm.DecisionModelFallbackAdapter([model, microsoft_ai.DecisionModel()])
```

— Runnable code sample with named classes and model ID. `livekit-agents@1.9.0`

03 **Reply handle from commit_user_turn** Not checked yet IMPROVED

65

Get the transcript and reply handle from commit_user_turn(return_handle=True), then await playout without generating a second reply.

Retrieve the reply handle for a committed user turn so you can await its playout without triggering a second generation.

```python
transcript, handle = await session.commit_user_turn(return_handle=True)
if handle is not None:
    await handle
```

— Exact flag and runnable snippet, but limited mechanism detail. `livekit-agents@1.9.0`

04 **New STT/TTS provider integrations** Not checked yet NEW

65

Adds a SILMA AI TTS plugin supporting English, Modern Standard Arabic, and Saudi Najdi TTS v2 models; a Reson8 `rumble-2` STT model; Gnani Prisma v3.0 STT and Timbre v2.5/v3.0 TTS models; a Zoom Voice Agent speech-to-speech model; ElevenLabs Scribe v3 Live STT with turn prediction; AssemblyAI `mode='manual'` for client-side turn detection; Sarvam keyterm prompting on `saaras:v4`; and Phonic text-only conversations via `modalities`.

— Enumerates every named model and provider, no usage example. `livekit-agents@1.9.0`

THINNER COVERAGE BELOW

05 **Tavus DEFAULT_FACE_ID removed** Not checked yet BREAKING

55

Tavus `DEFAULT_FACE_ID` is removed; set a face explicitly if needed.

— Clear migration note, minimal surrounding context. livekit-agents@1.9.0

└──▷ **BREAKING ON UPGRADE**

! AMD: use opening instructions instead of say() or generate_reply() during detection; do not enter AMD from an active `AgentTask`; `send_dtmf_events` no longer requests a reply.

! Tavus: `DEFAULT_FACE_ID` is removed; set a face explicitly if needed.

WAS THIS USEFUL? 👍 👎

OpenWhispr

Feedback ⓘ v1.12.2 NOTES ↗ RANK

OpenWhispr is a desktop dictation and meeting-transcription app that supports local speech models and user-configured cloud providers.

> OpenWhispr added ElevenLabs Scribe and AssemblyAI as new speech-to-text providers with custom-header support for self-hosted servers, alongside a richer note editor with slash commands and `[[note links]]`, snippet date/time variables, mic mute for note recordings, and new IT audit-log and policy-indicator controls.

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

01 ElevenLabs Scribe and AssemblyAI dictation providers

Not checked yet NEW 80

Adds `ElevenLabs Scribe` and `AssemblyAI Dictation` as speech-to-text providers, with `ElevenLabs Scribe` supporting dictation, uploads, and Note Recording, and adds support for custom headers (such as an `Authorization` header) for self-hosted speech-to-text servers.

Switch dictation to ElevenLabs Scribe when you want higher-accuracy transcription without OpenAI.

📍 In Settings, go to Dictation, open the `Speech-to-text provider` dropdown, and select `ElevenLabs Scribe`.

Point dictation at your own self-hosted speech-to-text server and pass a required auth header.

📍 In Settings, go to Dictation, select your self-hosted provider, then add a custom header (for example `Authorization: Bearer <token>`) in the `Custom headers` field.

— Names exact providers and header mechanism with runnable UI steps v1.12.1

02 Date/time/locale snippet variables Not checked yet NEW

75

Snippets can now insert `{{date}}`, `{{time}}`, `{{datetime}}`, or `{{locale}}` variables, which are filled in each time the snippet is expanded, so archived notes can be self-dated without manual entry.

Stamp every meeting snippet with the current date automatically, so archived notes are self-dated without manual entry.

📍 In a snippet body, type `{{date}}` where you want the insertion date to appear — for example: `Meeting notes — {{date}}`. The variable is replaced with the current date each time the snippet is expanded.

— Exact variable syntax plus a runnable usage example v1.12.2

03 **Rich note editor with slash commands and note links**

Not checked yet IMPROVED 65

The note editor now supports toggles, callouts, a table of contents, code blocks, math, embeds, and text colours accessible via a `/` menu, plus `[[note links]]` with backlink visibility showing where a note is linked from, and find and replace.

— Names many elements and the entry syntax but no step-by-step usage v1.12.1

04 **Mic mute during note recording** Not checked yet NEW 60

A mic mute control lets users silence their microphone during a note recording to keep side conversations out of the transcript, while system audio continues transcribing uninterrupted.

Keep a private side conversation off the record while still capturing all system audio during a meeting recording.

📍 While a note recording is active, click the mic mute button in the recording controls to silence your microphone input. System audio transcription continues uninterrupted.

— Clear mechanism and exact UI steps, but no named control/flag v1.12.2

05 **IT audit log and policy-managed indicators** Not checked yet NEW 60

A local audit log of meeting recordings is now available for IT review, and API keys and MCP are shown as managed when policy disables them.

Review a local record of which meetings were recorded before an IT compliance audit.

📍 In Settings, go to IT & Admin, then open Meeting Recording Audit Log to see the local history of recordings on this machine.

— Names audit surface and exact navigation path v1.12.1

THINNER COVERAGE BELOW

06 **Plans and Billing view changes** Not checked yet IMPROVED 40

The Plans and Billing view now shows your personal plan first, a separate view per workspace, and a preview of the new seat total before purchase.

— Describes changes but no exact navigation or figures v1.12.1

07 **New local speech models** Not checked yet NEW 40

New local models `LFM3 1.7B`, `VL 2B`, and Omni 2B are now available.

— Names exact models but no usage guidance v1.12.1

08 **Subfolders in notes sidebar** Not checked yet NEW 35

Notes can now be organized with one level of subfolders nested in the notes sidebar, synced across devices.

— Describes the capability but no navigation detail v1.12.2

09 **Skills sync via Cloud** Not checked yet NEW 30

Skills now sync on and off states to other computers via Cloud.

— States behaviour but no mechanism or setup steps v1.12.1

10 **Outside guest transcription blocking** Not checked yet NEW 25

Organizations can now block transcription of meetings that include outside guests.

— One-line policy description with no setup detail v1.12.2

11 **Workspace icons in settings and switcher** Not checked yet NEW 25

Settings and the workspace switcher now show workspace icons.

— Minimal UI note with no further detail v1.12.1

WAS THIS USEFUL? 👍 👎

ElevenLabs Python SDK

Feedback ⓘ v2.80.0 NOTES ↗

RANK

The official Python SDK for the ElevenLabs API.

> ElevenLabs' Python SDK adds a decorator-based client tool registration, agents-as-code workflows with push/pull, and speaker diarization support for real-time speech-to-text.

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

01 Diarization in realtime speech-to-text

Not checked yet NEW

81

Enable speaker diarization and automatic speaker-name inference in real-time speech-to-text by passing `diarize` and `infer_speaker_names` to the realtime speech-to-text wrapper, e.g. elevenlabs.speech_to_text.realtime(model='scribe_v2_realtime', diarize=True, infer_speaker_names=True).

Capture who is speaking and resolve speaker names automatically in a real-time transcription session.

python

```
session = elevenlabs.speech_to_text.realtime(
    model='scribe_v2_realtime',
    diarize=True,
    infer_speaker_names=True,
)
```

— Includes exact parameters and a runnable code example v2.80.0

02 Agents-as-code with push and pull

Not checked yet NEW

62

Define, push, and pull ElevenLabs agents as versioned code using `agents.push` and `agents.pull`, with a structured CLI project layout for managing agent definitions.

— Names commands but no CLI usage example given v2.80.0

THINNER COVERAGE BELOW

03 **`@client_tool` decorator for client tools** Not checked yet NEW

56

Decorate a function with `@client_tool` to register it as a client tool directly, without manual `ClientTools` wiring.

— Names decorator but no code sample shown v2.80.0

04 **Conditional workflow edges in agent graphs** Not checked yet

IMPROVED

36

Model workflow edges now direction by condition presence, letting conditional and unconditional transitions coexist in agent graphs.

— Describes behaviour change but no concrete API shown v2.80.0

WAS THIS USEFUL?  

streamcore-server

Feedback ⓘ v0.17 NOTES ↗

Streamcore-server is an open-source realtime voice agent server in Go with WebRTC (WHIP), barge-in, streaming STT/LLM/TTS pipelines, plugin system, multi-language SDKs, SIP telephony, ESP32 support & fully local mode.

streamcore-server adds an EmitsPartials capability flag for finals-only STT providers and improves turn-taking logic.

└──▷ GET THIS VERSION ▾

Docker

```
$ docker pull ghcr.io/streamcoreai/streamcore-server:v0.17
```

git

```
$ git clone --branch v0.17 https://github.com/streamcoreai/streamcore-server.git
# already have the repo? check out this version:
$ git checkout v0.17
```

published under the project's GitHub account · checked 26 Nov Wrong command?

Adds `EmitsPartials` capability flag to correctly handle STT providers that emit only final transcripts, not partial results. Not checked yet

NEW INTEGRATION Telnyx streaming STT provider

NEW INTEGRATION Moonshine STT provider

WAS THIS USEFUL? 👍 👎

ElevenLabs

Feedback ⓘ November 23, 2026 (2026-11-23) BLOG ↗

RANK

ElevenLabs is a voice AI platform with text-to-speech, speech-to-text, voice agents, dubbing and music generation, available in the app and through an API.

> ElevenLabs shipped agents-as-code support for defining and syncing ElevenAgents from Python and JavaScript, a new Swift SDK v4 with a unified conversation client, four additional LLMs for ElevenAgents, a 10x increase in Text to Speech request length, and new union/array tool parameter schemas.

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

01 **Agents-as-code for ElevenAgents in Python/JS SDKs** Not checked yet

NEW 95

Define ElevenAgents agents in code using the `Agent` builder in the Python and JavaScript SDKs from v2.80.0, then push or update them — including every webhook and client tool — with agents.push(), pull the project layout with agents.pull(), and validate the definition in CI without an API key using validate().

Push a fully defined agent — settings, tools, and webhooks — to ElevenAgents from code, so your agent config lives in version control alongside your application.

python

```python
from elevenlabs.agents import Agent

agent = Agent(
    name='support-bot',
    llm='gpt-6.2-sol',
    tools=[...],
)
agents.push(agent)
```

Validate an agent definition in CI without an API key to catch configuration errors before deploying.

python

```python
validate(agent)
```

— Named methods with runnable push/pull/validate code examples

November 23, 2026 (2026-11-23)

02 **Swift SDK v4 unified ConversationClient** Not checked yet NEW

90

Swift SDK v4.0.0 introduces a single `ConversationClient` for starting voice conversations over WebRTC with `startVoiceConversation` and text-only conversations over WebSocket with `startTextOnlyConversation`, adding compile-time credential matching, `clientTools` registration, `chatHistory` streaming, and `onDisconnect` reporting.

— Named methods and class but no code example given November 23, 2026 (2026-11-23)

---

03 **100,000-character Text to Speech requests** Not checked yet

IMPROVED

85

Text to Speech requests to `eleven_v4` and `eleven_v4_turbo` now accept up to 100,000 characters per request, up from 10,000, enabling roughly 100 minutes of audio from a single call across every Text to Speech endpoint that accepts these models.

Send a long-form document to TTS in one request now that `eleven_v4` accepts up to 100,000 characters, eliminating manual chunking for content up to ~100 minutes of audio.

```
$ curl -X POST https://api.elevenlabs.io/v1/text-to-
speech/<voice_id> \
  -H 'xi-api-key: <api_key>' \
  -H 'Content-Type: application/json' \
  -d '{"text": "<up_to_100000_chars>", "model_id": "eleven_v4"}'
```

— Exact limits with a runnable curl example November 23, 2026 (2026-11-23)

---

04 **Union and array tool parameter schemas** Not checked yet NEW

75

Declare union tool parameters with `AnyOfJsonSchemaProperty` using `any_of` shapes and a `discriminator` property, and declare array parameters with `QueryArrayJsonSchemaProperty` sent as repeated keys such as `?status=open&status=pending`.

— Named schema types and syntax, no full code sample November 23, 2026 (2026-11-23)

THINNER COVERAGE BELOW

---

05 **Four new LLMs for ElevenAgents** Not checked yet NEW

55

ElevenAgents now supports four new LLMs — `gpt-6.2-sol`, `gpt-6.2-astra`, `gemini-4-hydrogen`, and `gemini-4-neon` — selectable via the agent LLM picker in the app or the `llm` field through the API.

— Names models and API field but no example November 23, 2026 (2026-11-23)

WAS THIS USEFUL? 👍 👎

AGENT

◆ Web Search & Scraping for Agents

Perplexity API

Feedback ⓘ updated · Nov 26, 2026 NOTES ↗

RANK

Perplexity API provides language models with real-time web search and retrieval for building applications that research and answer questions.

> Perplexity API shipped a new page-retrieval tool and endpoint for MCP, and added historical time-travel support across its Search, Context, and Agent APIs.

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

01 perplexity_context MCP tool and /v1/context endpoint

Not checked yet NEW 95

The new `perplexity_context` tool, backed by `POST /v1/context`, retrieves page content as Markdown by default with optional HTML, links, query-relevant highlights, yes/no criteria scores, and page metadata. It supports freshness controls `if_modified_since`, `max_age_seconds`, `cache_only`, and `snapshot_as_of`, plus a `timeout_ms` parameter, and is reachable on the hosted MCP server at `https://api.perplexity.ai/mcp` via OAuth or API key, or locally via `@perplexity-ai/mcp-server` version 1.5.0 in stdio mode.

Pull Markdown content and freshness-filtered highlights for a known URL without hitting the live page — useful for auditing what a site said at a fixed point in time.

```
$ curl -X POST https://api.perplexity.ai/v1/context \
  -H 'Authorization: Bearer <API_KEY>' \
  -H 'Content-Type: application/json' \
  -d '{
    "urls": ["https://example.com/report"],
    "snapshot_as_of": "2026-11-01T00:00:00Z",
    "cache_only": true,
    "highlights": true
  }'
```

— Full endpoint, parameters, and connection methods named. snapshot-20261126

02 snapshot_as_of time-travel across APIs

Not checked yet NEW 90

Passing `snapshot_as_of` (RFC 3339 with time zone, e.g. `2026-11-01T00:00:00Z`) to `POST /search`, `POST /v1/context`, or the Agent API `web_search` and `fetch_url` tools via `filters` fetches a historical snapshot of any page up to 365

days in the past; pages with no stored copy at that time return `status.type` of `no_content`. The parameter is available as a `datetime` object in Perplexity Python library 0.50.0 and as a `Date` object in TypeScript SDK 0.45.0.

Run a historical web search scoped to pages as they existed 90 days ago, to compare search results over time.

```
$ curl -X POST https://api.perplexity.ai/search \
  -H 'Authorization: Bearer <API_KEY>' \
  -H 'Content-Type: application/json' \
  -d '{
    "query": "critical infrastructure vulnerabilities",
    "snapshot_as_of": "2026-08-26T00:00:00Z"
  }'
```

— Names endpoints, SDK versions, limit, and error behavior. snapshot-20261126

WAS THIS USEFUL? 👍 👎

Parallel CLI

Feedback ⓘ v0.10.4 NOTES ↗ RANK

Parallel CLI is a command-line tool and Python package for Parallel's web search and research APIs.

> Parallel CLI v0.10.4 switches authenticated sessions to OAuth credentials rather than raw API keys.

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

01 **OAuth-based authentication after login** Not checked yet IMPROVED

53

After running `parallel-cli login`, the CLI now uses product API OAuth credentials instead of raw API keys, exchanging short-lived tokens rather than long-lived secrets for authenticated sessions.

— Names the command and mechanism but no further detail v0.10.4

WAS THIS USEFUL? 👍 👎

brightdata-mcp

Feedback ⓘ v4.9.0 NOTES ↗

Brightdata-mcp is a powerful Model Context Protocol (MCP) server that provides an all-in-one solution for public web access.

brightdata-mcp adds a `business_search` tool for LinkedIn company and people lookups, with three query modes and multiple result views.

└──▷ GET THIS VERSION ▾

```
$ git clone --branch v4.9.0 https://github.com/brightdata/brightdata-mcp.git
# already have the repo? check out this version:
$ git checkout v4.9.0
```

NEW INTEGRATION Find companies or people from LinkedIn data using the new `business_search` tool, with `category` set to `company` or `people`; choose from `instant` (default, plain text, up to 200 characters), `smart` (plain text with AI reranking), or `ludicrous` (exact JSON filters) modes; views are `summary` (default), `full` (limit 5 or fewer), or a field list; returns up to 10 records per call, with `instant` and `smart` reaching the first 100 matches and `ludicrous` paging to offset 1000.

Not checked yet

Accounts without Business Search access now receive a `Business Search not enabled` message that names `search_dataset` and `run_scraper` as alternatives.

Not checked yet

└──▷ BREAKING ON UPGRADE

! The `web_scraping_strategy` and `diagnose_scraping_approach` prompts have been removed; the server instructions now give the same tool order to every client.

WAS THIS USEFUL? 👍 👎

◆ Browser & Computer Use

agent-s

Feedback ⓘ v0.3.5 NOTES ↗ RANK

Agent S: the state-of-the-art open source computer use agents

> Agent S's latest releases focus on desktop app polish and local sandbox agent reliability: VPN-aware sandbox networking, in-app update checks, bundled DeepSeek Harness runtime, guided first-run setup, searchable model catalogs, and recovery after interrupted turns.

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

01 **VPN-aware sandbox networking on macOS** Not checked yet IMPROVED

80

The sandbox now follows the Mac host's IPv4 routes and DNS, including VPN routes, so agents can reach VPN-protected resources without extra configuration. Networking can be switched back to isolated mode via Settings → Sandbox Network → Standard sandbox network, which takes effect after a sandbox restart.

— Names exact setting path and restart behavior v0.3.5

02 **In-app update checks** Not checked yet NEW

65

Agent S now checks for updates automatically, with manual checks available via Agent S → Check for Updates... ; the sandbox is saved before the app restarts to apply updates.

— Exact menu path given, clear action v0.3.4

03 **Searchable Ollama and OpenRouter model catalogs** Not checked yet

NEW

60

Agent S adds searchable Ollama and OpenRouter model catalogs that surface input capabilities, memory guidance, and Zero Data Retention eligibility.

— Names providers and specific surfaced attributes v0.3.3

THINNER COVERAGE BELOW

04 **Guided first-run setup for local sandbox agents** Not checked yet

NEW

45

Agent S adds guided first-run setup for local sandbox agents, including configurable memory and setup progress tracking.

— Names components but no concrete steps v0.3.3

05 **Recovery after interrupted turns** Not checked yet IMPROVED

45

Agent S now recovers sandbox sign-in, desktop controls, and agent state after interrupted turns for local sandbox agents.

— Explains what recovers but not the trigger mechanism v0.3.3

06 **Sidebar agent reordering** Not checked yet NEW 40

Agents within sidebar groups can now be reordered by dragging or using move actions in the Agent S desktop app.

— Describes action but no detail on mechanics v0.3.4

07 **Bundled DeepSeek Harness runtime** Not checked yet NEW 40

Agent S now bundles the DeepSeek Harness runtime for Apple Silicon Macs and maintains a separate runtime per agent.

— States what shipped, little on how to use it v0.3.4

WAS THIS USEFUL? 👍 👎

◆ MCP Servers & Tooling

mcp-server

Feedback ⓘ v2.7.2 NOTES ↗ RANK

BrowserStack's Official MCP Server

> BrowserStack's MCP server added a read-only capability registry mode, migrated Accessibility features to a dashboard API with token auth, improved test-case filtering, and added App Live app listing, while dropping the deprecated accessibilityExpert tool.

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

01 **Accessibility reads and actions via dashboard API** Not checked yet

IMPROVED 60

App Accessibility reads, issue actions, and Screen Reader Automation now route through the dashboard API using a minted token, reaching the Accessibility dashboard's own views instead of the External API.

– Names the APIs and mechanism but no endpoints or token format. v2.7.2

02 **Filter listTestCases by ID** Not checked yet IMPROVED 60

`listTestCases` in test management can now be filtered by ID to narrow case retrieval.

– Names the exact method but no usage example. v2.7.2

THINNER COVERAGE BELOW

03 **accessibilityExpert tool removed** Not checked yet BREAKING 45

The deprecated `accessibilityExpert` tool has been removed from the server in this release.

– Names the tool but gives no migration path. v2.7.2

04 **List apps uploaded to App Live** Not checked yet NEW 45

The server can now list apps already uploaded to App Live directly.

– States the capability without naming a tool or command. v2.7.0

05 **Read-only capability registry mode** Not checked yet NEW

35

A new read-only capability registry mode restricts which tools are exposed by the server.

— No config key or flag given to enable it. v2.7.2

└──▷ **BREAKING ON UPGRADE**

! The `accessibilityExpert` tool is removed (deprecated tool dropped in this release).

WAS THIS USEFUL? 👍 👎

gram

Feedback ⓘ server@2.44.0 NOTES ↗ RANK

Gram provides MCP servers and AI-agent tools from APIs with access controls and observability.

> Gram's latest releases add a full agent access-request workflow and an agent feedback system for MCP servers, alongside intent-based policy guardrails, read-only tool execution in gateways, OAuth secret migration tooling, and a string of admin and telemetry refinements.

└──▷ **WHAT SHIPPED · 17 FEATURES** most completely described first ? what's the number?

01 Device agent access requests with approval

Not checked yet NEW

95

Members can file temporary access requests from the device agent with `agent.requestAccess` (optional message and either a duration of 1, 2, or 4 hours or a deadline), see what they can still request with `agent.listRequestableAccess`, and administrators can review pending requests with `agent.listPendingAccessRequests`, which lists requests newest-first with dashboard approve/deny links. When a delivered MCP server reports no access, `agent.getMyConnections` now names the exact access to request. Requests carry a `requested_until` field (duration or deadline, stored in UTC) surfaced on the Slack card, and the dashboard's Access Requests tab shows a new Wants it column with the requested duration (1 hour, 2 hours, 4 hours, or end of day) and optional notes. All device-agent endpoints require a per-user device agent key and are off unless access requests are enabled.

As a member, list what temporary access you can still request and then file a two-hour request for a specific resource from the device agent.

```text
# List requestable access
agent.listRequestableAccess()

# File a request for two hours with an optional message
agent.requestAccess({ duration: 2, message: 'Needed for incident triage' })
```

As an org administrator, check which access requests are pending so you can approve or deny them from the dashboard.

```text
agent.listPendingAccessRequests()
```

— Names every endpoint, field, and UI element with runnable examples `server@2.44.0` `dashboard@0.157.0`

---

02 **Agent feedback system for MCP servers** Not checked yet NEW

95

An MCP server's new Agent Feedback tab lets you configure a `submit_feedback` tool (name and instructions configurable) that agents call to report what blocked or confused them; reports are stored with secrets and personal data masked, and the tab shows 30-day counts by severity, most-named tools, and every report. Gateways expose each member server's feedback tool under the member's qualified name (for example `orders--submit_feedback`) in direct, drill-down, and code mode. Reports can be exported as OTLP logs via the new MCP feedback data export or received via the new `mcp_feedback.submitted_v1` webhook event.

Subscribe to agent feedback events so your SIEM or alerting pipeline receives a payload each time an agent calls `submit_feedback` on an MCP server.

```
$ POST /webhooks
{
  "events": ["mcp_feedback.submitted_v1"],
  "url": "https://your-siem.example.com/hooks/gram"
}
```

— Full mechanism, named tool pattern, export, and webhook event with example `server@2.43.0`

---

03 **tool_calls_only filter for observability overview** Not checked yet

NEW

80

The `telemetry.getObservabilityOverview` endpoint now accepts a `tool_calls_only` parameter; when set, chat, token, cost, and session duration figures return as 0, and overviews of two hours or more read from pre-summed tool call counts aligned to 15-minute boundaries.

Fetch a lightweight telemetry overview covering only tool-call activity for the last two hours, using pre-summed counts instead of raw logs.

```text
telemetry.getObservabilityOverview({ tool_calls_only: true, window_hours: 2 })
```

— Named endpoint and parameter with a runnable example `server@2.44.0`

---

04 **RemoteSessionClients RPCs renamed to RemoteSessions**

Not checked yet

BREAKING

70

The internal RPCs `remoteSessionClients.attachUserSessionIssuer`, `remoteSessionClients.detachUserSessionIssuer`, and `organizationUserSessionIssuers.listRemoteSessionClientBindings` are removed; use `remoteSessions.attachServerClientAssociation`, `remoteSessions.detachServerClientAssociation`, and `remoteSessions.listServerClientAssociations` instead.

— Every old and new RPC name given verbatim `server@2.44.0`

---

05 **OAuth secrets migration to org-owned keys** Not checked yet NEW

65

Staff can migrate an organization's OAuth client secrets and session tokens from the platform key to their own encryption keys from the admin dashboard, one at a time or in batches, via enroll, backfill, finalize, and rollback steps. A dedicated step-by-step Tailscale OAuth migration page moves an organization's saved Tailscale OAuth clients from platform-key copies to an organization key through enroll, backfill and verify, finalize, then clear.

— Clear migration steps and UI path, no API names given `server@2.43.0` `admin@0.14.1`

---

06 **Read-only tool execution in gateways** Not checked yet NEW

65

Progressive gateways now offer `execute_read_only_tool` alongside `execute_tool`; it runs only member tools annotated read-only, so MCP clients that skip approval for read-only tools can execute those calls without a prompt.

— Named endpoint and behaviour, no runnable example given `server@2.43.0`

---

07 **Platform MCP core tools removed** Not checked yet BREAKING

65

Platform MCP no longer offers `list_my_servers`, `search_my_tools`, `describe_tools`, or `execute_tool`; MCP clients that call these must switch to connecting the Organization Gateway, retrieved via `get_my_organization_gateway`.

— Names removed tools and the replacement migration path `server@2.43.0`

---

08 **Intent-based policy primitives and guardrail replay**

Not checked yet NEW

60

Intent-based policies now use six reusable primitive labels and ten composed rules, adding protections against broadcasting personal data and making unreviewed public releases. A new Intent-based policies page (rolled out to selected users) replays a stored or uploaded session and shows which tool calls the intent guardrail would permit, ask about, or block.

– Explains mechanism but primitives and rules aren't individually named `server@2.43.0` `server@2.44.0`

09 **Tailscale Kubernetes operator provider removed** Not checked yet

60

BREAKING

The Tailscale Kubernetes operator provider, its per-connection attestor, and the `private-network-gateway` rollout check are removed; the reconciler no longer reads `GRAM_NETWORK_INGRESS_OPERATOR_*`, `GRAM_NETWORK_INGRESS_GATEWAY_ENABLED`, or the backend, attestor, tag, CIDR, and token audience settings, so these can be removed from deployments.

– Names env vars and a concrete deployment cleanup step `server@2.43.0`

THINNER COVERAGE BELOW

10 **Organization Gateway connection retrieval** Not checked yet NEW

55

Retrieve the Organization Gateway connection with `get_my_organization_gateway` to use your own MCP servers from an MCP client.

– Named endpoint but thin on behaviour beyond its purpose `server@2.43.0`

11 **Admin encryption transitions list filtering and sorting**

Not checked yet IMPROVED

55

The admin encryption transitions list now shows only organizations with values for the selected use case by default, including value counts and last-written timestamps, sorted most-recently-written first, with options to switch to all organizations or sort oldest first.

– Clear UI behaviour described but no config or API named `admin@0.14.2`

12 **Hooks rollout progress tool for staff admin** Not checked yet NEW

55

The staff Admin MCP gains a read-only `get_hooks_release_progress` tool showing how many organizations clear the pin, how many marketplace repositories have republished, and how many installs report the binary version for a given hooks release.

– Named tool and its three metrics, no usage example `server@2.44.0`

13 **Legacy managed-assistant platform toolset removed** Not checked yet

50

BREAKING

The legacy managed-assistant platform toolset and the `assistant-platform-mcp` rollout flag are removed; the project's managed assistant now always uses the Platform

MCP read toolset, and the `platform_*` tools named in the system prompt no longer exist.

— Names the flag and tool prefix but no migration steps `server@2.44.0`

---

14 **MCP connection health API** Not checked yet NEW 45

The MCP connection health API lists suspected problems for an MCP sign-in and returns one problem's people, servers, messages, and failed attempts.

— Describes output fields but doesn't name the exact endpoint `server@2.43.0`

---

15 **Model completion span telemetry enrichment** Not checked yet IMPROVED 45

Model completion spans now record which upstream provider served the call, the time to the first streamed chunk, and the reasoning and cached token counts.

— Names the new fields but no way to query them `server@2.44.0`

---

16 **EU-only AI model routing via OpenRouter** Not checked yet IMPROVED 40

In the EU partition, AI model requests are now routed through OpenRouter in the EU, so prompts and responses stay with in-region model providers and only EU-available models can be used.

— Explains data residency change but no action for readers `server@2.43.0`

---

17 **AI client catalog served from database** Not checked yet IMPROVED 40

Gram now reads the AI client catalog from the database instead of the server binary, refreshing it every minute and falling back to the built-in catalog if the database catalog is empty or invalid.

— Describes mechanism and fallback but nothing for readers to act on `server@2.43.0`

└──▷ **BREAKING ON UPGRADE**

! The internal RPCs `remoteSessionClients.attachUserSessionIssuer`, `remoteSessionClients.detachUserSessionIssuer`, and `organizationUserSessionIssuers.listRemoteSessionClientBindings` are removed; use `remoteSessions.attachServerClientAssociation`, `remoteSessions.detachServerClientAssociation`, and `remoteSessions.listServerClientAssociations` instead.

! The legacy managed-assistant platform toolset and the `assistant-platform-mcp` rollout flag are removed; the project's managed assistant now always uses the Platform MCP read toolset, and the `platform_*` tools named in the system prompt no longer exist.

! Platform MCP no longer offers `list_my_servers`, `search_my_tools`, `describe_tools`, or `execute_tool`; MCP clients that call these must switch to connecting the Organization

Gateway (returned by `get_my_organization_gateway`).

! The Tailscale Kubernetes operator provider, its per-connection attestor, and the `private-network-gateway` rollout check are removed; the reconciler no longer reads `GRAM_NETWORK_INGRESS_OPERATOR_*`, `GRAM_NETWORK_INGRESS_GATEWAY_ENABLED`, or the backend, attestor, tag, CIDR, and token audience settings, so these can be removed from deployments.

WAS THIS USEFUL? 👍 👎

claudexor

Feedback ⓘ v6.4.1 NOTES ↗ RANK

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

> Claudexor's latest releases add shell-command support for headless Antigravity workspace_write runs with version gating, new `claudexor remote preflight/install/bootstrap` commands for setting up the SSH runtime, and server-enforced run deadlines for Linux.

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

01 SSH remote runtime setup commands

Not checked yet NEW

88

`claudexor remote preflight` previews SSH runtime details — the login, host key fingerprints and signed build — before writing anything. `claudexor remote install` installs the SSH runtime without sudo or a running daemon, accepting the digest from `claudexor remote preflight` and re-verifying the recipient before writing. `claudexor remote bootstrap` completes SSH setup as a separate step after installation.

Inspect a remote server's SSH runtime fingerprints and signed build before committing to an install.

```
$ claudexor remote preflight
```

Install the SSH runtime on a fresh server using the digest from preflight, with no sudo required.

```
$ claudexor remote install <digest>
```

Connect to the newly installed SSH runtime as a final, separate bootstrap step.

```
$ claudexor remote bootstrap
```

— Three runnable subcommands with their exact order of use v6.4.0

02 Shell commands in headless Antigravity workspace_write runs

Not checked yet NEW

83

Headless Antigravity runs in `workspace_write` mode now pass session-only settings that allow shell commands while keeping the profile's deny and ask rules, so commands complete in the same turn instead of ending with an empty result. Runs on a known agy version older than 1.5.7 now refuse before starting with error code `agy_workspace_write_version_unsupported` and prompt an update; runs where the version can't be determined start without the grant and report that. `claudexor doctor` now reports the minimum required agy version for workspace_write runs as `workspace_write_min_version`, and native permission denials in workspace_write runs are now reported as `permission_denied` instead of the previous cause unknown empty-result error.

Check whether your installed agy meets the 1.5.7 minimum before attempting a workspace_write run — look for the `workspace_write_min_version` field in the output.

```
$ claudexor doctor
```

— Names exact error codes and doctor field, gives one check command v6.4.1

THINNER COVERAGE BELOW

03 **Server-enforced run deadlines for Linux** Not checked yet NEW

50

Adds server-enforced deadlines for Linux runs: a controller can set an absolute deadline at run start, renew it while armed, and receive a report distinguishing a confirmed stop from an unresolved one.

— Explains mechanism but no command or flag given v6.4.0

WAS THIS USEFUL? 👍 👎

robloxstudio-mcp Feedback ⓘ v3.6.0 NOTES ↗ RANK

Robloxstudio-mcp is a MCP server for Roblox Studio runtime debugging, playtest control, screenshots/input, multiplayer testing, and per-peer server/client eval from AI agents.

> robloxstudio-mcp's v3.6.0 release adds asynchronous background execution with status polling, richer playtest readiness controls, multi-frame screenshot capture, Luau wait conditions for runtime eval, and Linux/Wine screen capture support.

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

01 Async background calls with status polling Not checked yet NEW

88

Long-running calls that answer `state: "pending"` can now be polled via `get_request_status` using the `requestId` returned with the pending response. The response budget before a call is treated as pending is configurable via the `ROBLOX_STUDIO_MCP_RESPONSE_BUDGET_MS` environment variable, defaulting to 25 seconds.

Tune the async response budget so long-running playtest calls don't time out your HTTP client prematurely.

```
$ ROBLOX_STUDIO_MCP_RESPONSE_BUDGET_MS=40000 npx -y
@chrrxs/robloxstudio-mcp@latest --auto-install-plugin
```

– Names env var, flag, and default with runnable example v3.6.0

02 Playtest readiness and targeting controls Not checked yet

IMPROVED

66

`solo_playtest` and `multiplayer_playtest` now accept a `ready_timeout_s` parameter (default 90s) controlling how long they wait for all runtime peers to be ready, and the response includes a `targets` field for addressing each peer directly. `solo_playtest` start now defaults to `mode=play`.

– Names parameters and default but no usage example v3.6.0

03 Multi-frame screenshot capture with contact sheet Not checked yet

NEW

62

`capture_screenshot` can now take up to 16 frames in one call using the `frames` and `interval_ms` parameters, and passing `grid: true` returns them as a single numbered contact sheet.

— Names exact params and limit but lacks worked example v3.6.0

THINNER COVERAGE BELOW

04 **Luau wait conditions for runtime eval** Not checked yet NEW

55

`eval_server_runtime` and `eval_client_runtime` now accept a `wait_until` Luau condition that is polled inside the running game until it becomes true or `timeout_s` expires.

— Names the two functions and parameters, no example given v3.6.0

05 **Linux capture support under Wine** Not checked yet NEW

40

`capture_screenshot` and `capture_video` now capture the Studio window on Linux under Wine, cropped to the viewport via the X display.

— States mechanism briefly but no setup steps v3.6.0

WAS THIS USEFUL? 👍 👎

gflow-cli

Feedback ⓘ v0.107.0 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 expanded its Google Flow CLI and MCP server with Ultra 4K video upscaling, video-to-video editing, an upgraded video-extend MCP tool, a new OpenAI-compatible image generations endpoint, and an experimental direct-HTTP image transport.

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

01 **OpenAI-compatible image generations endpoint** Not checked yet

NEW 90

`gflow serve` now exposes an OpenAI-compatible `POST /v1/images/generations` endpoint that returns `b64_json` responses using Flow image models, letting any OpenAI-compatible client or script generate images without changing client code.

Call a locally running gflow server from any OpenAI-compatible client or script to generate images without changing client code.

```
$ curl http://127.0.0.1:8000/v1/images/generations \
  -H 'Content-Type: application/json' \
  -d '{"model": "nano2", "prompt": "a quiet mountain lake at
dawn, cinematic photography", "n": 2, "response_format":
"b64_json"}'
```

— Named endpoint, method and response format with a runnable curl example. v0.107.0

02 **Video extend upgrades and MCP tool** Not checked yet NEW

75

Adds a `gflow_extend_video` MCP tool and upgrades video extend with local upload support, 3–10 second duration control, 0–3 reference images, and a credit cap.

— Concrete limits and a named MCP tool, but no usage example. v0.107.0

03 **Direct HTTP image generation transport** Not checked yet NEW

70

Adds an experimental `direct_http` image transport that sends images via a single direct POST instead of the browser-driven path, and `gflow serve` now supports generating 2–4 images per request with a reference or character through this transport.

— Mechanism and scope named but no exact command shown. v0.107.0 v0.106.0

THINNER COVERAGE BELOW

04 **Video caption styling options** Not checked yet NEW 55

Video captions gain `position`, `margin`, `font-size`, and `box-opacity` options for styling control.

— Flags named but no example of applying them. v0.107.0

05 **`--thinking-level` flag for image generation** Not checked yet NEW 55

`gflow image` now accepts a `--thinking-level` option for deeper reasoning during image generation.

— Flag and command named, no concrete values or example. v0.107.0

06 **Video-to-video edit** Not checked yet NEW 40

New video-to-video edit capability via `flow.google.com`, available both as a CLI command and as an MCP tool.

— Surface named but no specific command or flag given. v0.107.0

07 **Ultra 4K video upscale** Not checked yet NEW 20

Enables Ultra 4K video upscale for supported Flow items.

— Bare description with no mechanism, command, or scope. v0.107.0

WAS THIS USEFUL?  

gograph

Feedback ⓘ v1.7.25 NOTES ↗ RANK

Gograph is a local Go static analysis, call-graph, blast-radius, and CI architecture-gate CLI—with a built-in stdio MCP server

> gograph's latest releases add historical call-graph queries at any Git ref via `--rev`, mutex lock-scope tracing via `--locks`, automatic stale-graph rebuilds with an opt-out, and broader selector syntax across the CLI — while discontinuing Windows binaries.

└──▷ **WHAT SHIPPED · 13 FEATURES** most completely described first ? what's the number?

01 Historical call-graph queries via --rev

Not checked yet NEW

93

Query the call graph at any past Git commit, branch, or tag with the global `--rev <ref>` flag, which extracts the tree via `git archive` into `.gograph/revs/<commit>/` and caches it so later queries against the same ref are answered instantly.

Inspect callers of a function as it existed on the main branch, without touching your working tree.

```
$ gograph callers Run --rev origin/main
```

Inspect what the call graph looked like at a past commit or release tag without checking it out — useful when triaging a regression or auditing a past deployment.

```
$ gograph --rev origin/main callers Store.Save --json
```

— Names flag, cache path, and git mechanism with runnable examples. v1.7.24

02 Mutex lock-scope analysis via --locks

Not checked yet NEW

93

Trace mutex usage in a function body with `gograph concurrency --locks <function>` (also available as a global `--locks <function>` flag), which lists each `Lock` / `RLock` call with its release line, marks `held until return` for deferred unlocks, and shows the calls made between lock and unlock — surfacing lock-acquisition scopes for the function.

Audit mutex discipline in a specific method — see every lock, its paired unlock, and the calls sandwiched between them.

```
$ gograph concurrency --locks Manager.start
```

Audit which code paths hold a mutex and where they release it, to find potential lock-scope bugs in a concurrent function.

```
$ gograph --locks ProcessQueue
```

— Full mechanism and example commands for lock tracing. v1.7.24

---

03 **Strings command JSON, file-only, and incremental search**

Not checked yet NEW 88

The `strings` command gains `--json` for machine-readable output and `--files-only` for deduplicated file paths; it now also searches files added since the last build and reports why indexed files were skipped.

Get a JSON-structured strings report for scripting or piping into other tools

```
$ gograph strings Store.Save --json
```

Emit only the file paths where strings appear — useful for feeding downstream grep or editors

```
$ gograph strings Store.Save --files-only
```

— Flags named with runnable example commands. v1.7.23

---

04 **Line-range reading on gograph source** Not checked yet NEW 85

Print a slice of a declaration or file with `--lines START-END` on `gograph source`, clamped to the declaration boundary, to read regions larger than the 64 KiB budget.

Read only the lines of a large function body relevant to a specific logic block, rather than pulling the entire file.

```
$ gograph source Store.Save --lines 42-67
```

— Flag, byte limit, and example command given. v1.7.24

---

05 **Test selection via gograph tests** Not checked yet IMPROVED

80

Select a specific test by name with `gograph tests <TestName> --format go-test` (or by function name when using test selectors) to run only that test. Test selection now also follows type, variable, and constant uses, as well as binary tests, widening the set of tests surfaced by `gograph tests`.

— Command and flag given, widened matching described. v1.7.25 v1.7.24

---

06 **Stale-graph auto-rebuild, skip flag, and warnings** Not checked yet

IMPROVED

75

CLI queries now automatically rebuild a stale graph before answering so results are never silently out of date. That rebuild can be skipped with `--no-refresh` (or `GOGRAPH_NO_REFRESH=1`) to answer immediately from the persisted graph, surfacing a stale warning on stderr or in the `warning` JSON field; commands also emit a CLI warning whenever they answer from a stale graph.

— Names flag, env var, JSON field but no example command. v1.7.25 v1.7.23

---

07 **Broader selector syntax across commands** Not checked yet IMPROVED

75

`path` now accepts file, package-path, and ID selectors; `source`, `callers`, and `callees` accept several selectors at once; selectors across commands accept repository-relative package paths and `pkg::Name` syntax; `globals` accepts a package path such as `internal/app`; and `callers` now matches whole identifiers instead of partial names, finding callers and tests via exact selectors in AST graphs.

— Enumerates every selector surface but lacks runnable example. v1.7.24 v1.7.23

---

08 **Path filtering on graph-backed commands** Not checked yet NEW

60

Filter any graph-backed command's results to a file or subdirectory with `--path DIR|FILE`, including scoping `query/strings` results to a single file.

— Flag named but no example command shown. v1.7.24

---

09 **Globals command for package-level state** Not checked yet NEW

60

Find all package-level variables and constants — and the functions that mutate them — with `gograph globals <pkg|dir>`.

— Command and behavior named, no example given. product docs

---

10 **Since-based review diffing** Not checked yet NEW

60

Review all declarations changed between a Git ref and the working tree with `gograph review --since <ref>`, enabling MCP-driven since-based diffing.

— Flag named, mechanism described, no example. v1.7.24

THINNER COVERAGE BELOW

11 **Build-tag filtering for queries** Not checked yet NEW 55

Filter queries by build tags with `--tags`, defaulting freshness checks to the graph's recorded tags.

— Flag named, minimal mechanism, no example. v1.7.24

12 **Richer node and skeleton output** Not checked yet IMPROVED 40

`node` output now shows a constant's declared value, and `skeleton` output now shows the file's import block.

— Two thin one-line output additions, no examples. v1.7.24

13 **Windows binaries discontinued** Not checked yet BREAKING 40

Windows binaries are no longer built or published; macOS and Linux targets remain unchanged.

— States removal and unaffected platforms, no migration detail. v1.7.25 v1.7.24 v1.7.23

└──▷ **BREAKING ON UPGRADE**

! Windows binaries are no longer built or published; macOS and Linux targets are unchanged.

! Windows binaries are no longer built or published; only macOS and Linux targets are available from v1.7.24 onward.

WAS THIS USEFUL? 👍 👎

xiaohongshu-mcp

Feedback ⓘ v2.7.6 NOTES ↗ RANK

MCP for xiaohongshu.com

> xiaohongshu-mcp added an optional parameter to flag AI-synthesised content when publishing image and video notes.

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

01 AI content disclosure flag for publishing

Not checked yet NEW

60

Image and video note publishing now accepts an optional `is_ai_generated` parameter to declare AI-synthesised content, e.g. `is_ai_generated: true`.

When publishing an AI-generated image post, flag the note as containing AI-synthesised content to meet platform disclosure requirements.

```text
is_ai_generated: true
```

— Names exact parameter and value but no further mechanism detail v2.7.6

WAS THIS USEFUL? 👍 👎

novamira

Feedback ⓘ v2.0.0 NOTES ↗ RANK

Novamira is a MCP server that gives AI agents full access to WordPress through PHP execution and filesystem operations

> Novamira v2.0.0 adds a set of production-ready AI oversight controls to WordPress, including adjustable approval thresholds, a 30-day reversible change history, an email-hold queue, an extensible Abilities Hub, and a safer draft-first default for new content.

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

01 **30-day AI change history** Not checked yet NEW 65

Review and reverse database changes your AI makes with AI change history, which records previous values and retains them for 30 days; accessed via the AI change history tab in AI Activity.

Review what your AI changed in the database and ask it to roll back an unintended edit.

📍 In Novamira, go to AI Activity and open the AI change history tab to see previous values, then ask your AI to restore the entry.

— Clear retention window and restore workflow given v2.0.0

02 **Adjustable AI approval levels** Not checked yet NEW 60

Control how often your AI asks for approval before acting by setting Ask more, Balanced, or Ask less under `Novamira > AI Activity`.

Raise the approval bar so your AI pauses before more consequential actions — useful when deploying to production for the first time.

📍 In Novamira, go to AI Activity and set the approval level to Ask more.

— Names exact UI path and setting values, no deeper mechanism v2.0.0

03 **Abilities Hub for custom approval prompts** Not checked yet NEW 60

Extend approval prompts to any ability, including those from other plugins, via the Abilities Hub, or list additional site actions that require approval under `AI Activity > Settings`.

Require your AI to ask before running a specific third-party plugin action that is not covered by default.

📍 In Novamira, go to AI Activity › Settings and add the action to the list of actions your AI should ask before running.

— Names exact settings path and extension mechanism v2.0.0

THINNER COVERAGE BELOW

04 **Email-hold queue for AI-sent emails** Not checked yet NEW

45

Emails sent through WordPress while your AI works now wait in AI Activity until you send or discard them, preventing unintended sends.

— Describes behavior but no usage example or deeper detail v2.0.0

05 **Draft-first default for new content** Not checked yet BREAKING

35

Novamira now keeps new posts and pages as drafts by default unless explicitly told to publish, and tells you where to preview them.

— States the default change but no config key to adjust it v2.0.0

WAS THIS USEFUL? 👍 👎

concord-mcp

Feedback ⓘ v0.25.0 NOTES ↗ RANK

Concord-mcp is a live messaging for coding agents

> Concord-mcp's sole release in this window adds integration with the OpenClaw client, bringing workspace context and live Concord message delivery into OpenClaw sessions.

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

01 **OpenClaw client integration with workspace instructions and live messaging**

Not checked yet NEW 40

concord-mcp now works with the OpenClaw client: it adds OpenClaw integration, workspace instructions delivered automatically in global mode so global-mode users get workspace context, and live delivery of Concord messages inside OpenClaw sessions.

— Names the integration and its pieces but no commands or config to act on. v0.25.0

NEW INTEGRATION DeepSeek Harness integration via native MCP

NEW INTEGRATION Kiro integration

> v0.24.0 NOTES ↗ STABLE
>
> concord-mcp adds a `concord work` CLI command to run workflow tools directly from the shell.
>
> └──▷ **GET THIS VERSION** ▾
>
> ```
> $ git clone --branch v0.24.0 https://github.com/Get-Concord-AI/concord-mcp.git
> # already have the repo? check out this version:
> $ git checkout v0.24.0
> ```

WAS THIS USEFUL? 👍 👎

eda-agent

Feedback ⓘ v0.8.2 NOTES ↗ RANK

Eda-agent is an open-source MCP server that lets an AI drive a live session of Altium Designer, and optionally KiCad or EasyEDA Pro, editing the open design in place while you watch. 560+ tools: schematic, PCB, library and project automation, design review, audits, rendering, auto-placement, panelization, autonomous design-plan executor.

> eda-agent's v0.8.2 release adds a dedicated board router, sandboxes SPICE simulation decks, locks KiCad conversion to local tooling, and expands editable surface across design rules, sheet options, dimensions, exports, and symbol parts.

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

01 **Board router with priced turns and dogleg entries** Not checked yet

NEW 75

One board router now handles priced turns, ordered connections, and dogleg pad entries, with round-by-round routing under a brief, job cancellation, and time limits.

— Mechanism and scope given, but no exact tool/command name v0.8.2

02 **SPICE deck sandboxing** Not checked yet NEW 75

`.control` blocks in `design_simulate_deck`, `sim_run_deck`, and `kicad_run_spice` are now limited to simulation, vector, and output commands; `shell`, `source`, `exec`, `codemodel`, `osdi`, `cd`, `edit`, and `.ferret` are refused, and file writes are confined to the run folder.

— Names exact tools and refused commands, mechanism clear v0.8.2

03 **KiCad conversion restricted to local kicad-cli.exe**

Not checked yet 65

BREAKING

`proj_import_from_kicad` and `proj_convert_to_kicad` are now restricted to a local-drive `kicad-cli.exe` only — UNC paths and mapped network drives are refused.

— Named endpoints and exact restriction, but no migration guidance v0.8.2

04 **Design rule editing: Matched Lengths, Clearance, Silk/Solder Mask, pair rules**

Not checked yet NEW 65

Matched Lengths rules can now be created and edited, keeping their position when edited; a Clearance rule's minimum clearances can be read and written by object kind; the Silk To Solder Mask choice can be read and edited; and a pair rule's uncoupled length can be read and edited.

— Names several rule types but no tool/command identifiers v0.8.2

THINNER COVERAGE BELOW

05 **Sheet options editing** Not checked yet NEW 55

Sheet parameters, sheet templates, title block, border, zone origin, fonts, and colours can now be edited through sheet options.

— Lists editable elements but no tool names given v0.8.2

06 **Board and schematic export to PDF and 3D renders** Not checked yet NEW 50

Boards and schematics can be exported to PDF or picture format for a named sheet or board, including photographic 3D renders.

— Describes output formats but not the invoking tool/command v0.8.2

07 **Symbol part editing** Not checked yet NEW 50

Symbol parts can be removed from the middle of a component, a chosen part can be shown, and counts per part can be displayed.

— Clear actions described but no tool name v0.8.2

08 **Dimension object editing** Not checked yet NEW 40

Dimension objects can be read, created, and edited, including radial dimensions.

— Thin description, no tool or command named v0.8.2

09 **Units in mils or millimetres across tools** Not checked yet IMPROVED 30

Lengths can now be specified in mils or millimetres on every tool.

— Bare statement with no mechanism or tool names v0.8.2

WAS THIS USEFUL?  

kinocut

Feedback ⓘ v2.2.0 NOTES ↗ RANK

Kinocut is a guardrailed video editing MCP server for AI agents. FFmpeg, Hyperframes, repurposing tools, Python client, and CLI.

> kinocut shipped a Chrome Manifest V3 extension with a Native Messaging bridge to FFmpeg, plus a cluster of editing features in v2.1.0 — beat-synced and onset-synced cutting, incremental shot rendering, spoken-word edit planning, full offline mode, and Hermes integrations for live render progress and TV-based review.

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

01 **Chrome extension and Native Messaging bridge to FFmpeg**

Not checked yet NEW 95

A new Chrome Manifest V3 Companion extension (source under `extensions/chrome/`) connects the browser to FFmpeg via a Native Messaging host. The host can be run directly with `python -m kinocut.native_messaging`, a binary 4-byte length-prefixed JSON stdio bridge (defined in `kinocut/native_messaging.py`) with a strict 1 MiB frame cap and a 7-action allowlist: `ping`, `doctor`, `info`, `quality_check`, `direct_plan`, `direct_apply` (gated on `confirm=True`), and `review_card`. The OS manifest is registered via `kinocut/native_messaging_install.py` with `--install --extension-id <32-char-[a-p]-id>` for Chrome, Chromium, Brave, or Edge on Linux, macOS, and Windows, and the extension uses least-privilege permissions (`nativeMessaging`, `downloads`, `storage` — no `<all_urls>`, no `externally_connectable`) with a popup offering a recent-download picker, live Host/FFmpeg/Offline readiness pills, one-click Inspect, Quality Check with a `--profile dark-grade` toggle, Review Card, and a spoken-direction plan preview requiring explicit Confirm & Render approval before any render is applied.

Register kinocut's Native Messaging host with your Brave browser profile so the Chrome MV3 Companion extension can reach the local FFmpeg bridge.

```
$ python -m kinocut.native_messaging_install --install --
extension-id abcdefghijklmnopabcdefghijklmnop
```

Smoke-test the stdio bridge directly from a terminal to confirm the host is reachable and FFmpeg is detected before pairing the browser extension.

```
$ python -m kinocut.native_messaging
```

— Names exact flags, commands, module paths, and allowlist actions. v2.2.0

02 **Full offline mode for air-gapped use** Not checked yet NEW

95

Setting `KINOCUT_OFFLINE=1` restricts model providers to loopback and adds LFM3-VL-2B vision, LFM3-Audio-2B transcription, FSRCNN upscaling and staged BiRefNet mattes for local-only processing; `kino doctor` reports `offline_readiness` from local files only.

Run kinocut in a fully air-gapped environment and check which local models satisfy all offline requirements.

```
$ KINOCUT_OFFLINE=1 kino doctor
```

— Names the env var, local models, and a verification command. v2.1.0

03 **Beat-synced cutting with music_grid** Not checked yet NEW

90

`kinocut.music_grid` exposes a `bar:beat:sub` grid with BPM, song offset, groove templates, microtiming, and keyframe lint so cuts land on the beat. `kino trim` gains `--bpm`, `--fps` and `--offset` flags, and timeline JSON accepts a `"grid": {"bpm": …, "offset_s": …}` block.

Trim a clip to a 120 BPM music grid so every cut lands on the beat, starting at 4 seconds into the song.

```
$ kino trim input.mp4 --bpm 120 --fps 24 --offset 4
```

— Exact flags and a runnable command example are given. v2.1.0

04 **Spoken-word edit planning with video_direct** Not checked yet NEW

85

Turn spoken words into a reviewable edit plan via `video_direct`, then execute it with video_direct_apply(confirm=true) or `kino direct "WORDS" -i INPUT --apply`; both tools live in the `intent` profile and `video_plan_ops` family.

Translate a spoken edit direction into a plan and immediately execute it on the source footage.

```
$ kino direct "cut to the action at 00:30 and add a j-cut before
the dialogue" -i rushes.mp4 --apply
```

— Exact tool names and a runnable command are provided. v2.1.0

05 **Incremental shot rendering with shot_cache** Not checked yet NEW

75

`kinocut.shot_cache` renders only changed shots by keying each shot on source contents, argv, parameters, frame range and upstream digests, splicing unchanged shots with `-c copy`, verifying the decoded frame count, and returning a receipt with per-shot digests.

— Mechanism fully described but no usage command shown. v2.1.0

06 **Opt-in local request logging** Not checked yet NEW

75

Setting `KINOCUT_REQUEST_LOG=1` or a path captures (request → chosen operation) pairs as redacted JSONL; records can be harvested with `python -m kinocut.request_log`.

— Env var and harvest command named, but no sample output shown. v2.1.0

07 **Dark-grade quality check profile** Not checked yet NEW

75

A new `dark-grade` quality profile, enabled via `--profile dark-grade` on `video-quality-check` or `release-checkpoint`, prevents intentionally dark footage such as night, noir and low-key interiors from being flagged as underexposed.

— Exact flag and target commands named, behavior explained. v2.1.0

08 **Cut review on paired TV via Hermes skill** Not checked yet NEW

75

A Hermes skill prepares a contact sheet and square-pixel proxy of a cut, sends both to a paired TV, and accepts `APPROVE <id>` or `REJECT <id>` commands after verifying the cut, proxy and sheet digests.

— Named commands and verification steps give a clear start point. v2.1.0

09 **Default review cards for trim, merge and shorts** Not checked yet

IMPROVED

75

`video_trim`, `video_merge` and `video_shorts` now write a review card next to the output by default, returning it as `review_card` plus a `preview_directive`; pass `review=False` to skip it.

— Exact fields and opt-out parameter named. v2.1.0

10 **Sound effect placement on audio onsets** Not checked yet NEW

70

`kinocut.audio_sync` detects onsets in kick, snare and hat bands and places sound effects using `pocket_lock`, `both` or `avoid_masking` placement modes, with sample-phase-stacked hits and A/B proxy exports of premixed alternatives.

— Mechanism is detailed but no CLI/API usage example given. v2.1.0

---

11 **Live render progress in Hermes Desktop** Not checked yet NEW

60

An opt-in Hermes Desktop plugin shows percent complete, estimated time left, elapsed time, output size and failure cause for renders in a Kinocut renders pane and status bar.

— UI location named but no config key or command given. v2.1.0

WAS THIS USEFUL? 👍 👎

Prefect FastMCP

Feedback ⓘ v4.2.6 NOTES ↗ RANK

FastMCP is a Python framework for building MCP servers and clients, handling transport, schema generation, authentication, and server composition.

> FastMCP v4.2.6 adds a client-side CodeMode client for cross-server tool execution, introduces a new typed event system with optional Redis persistence, adds OpenTelemetry transport attributes to server spans, and removes the experimental events module and bundled Authlib dependency.

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

01 **Typed event system via fastmcp-events package** Not checked yet

95

BREAKING

Declare event types as Pydantic models in the new `fastmcp-events` package (install with `pip install 'fastmcp[events]'`, import from `fastmcp_events`), register with @events.event('name'), emit with await events.emit(...), and subscribe via @events.reader(Model) with an optional `for_subscription` method. Events can be persisted across restarts and replicas via `DocketEventLog`, installed with `pip install 'fastmcp-events[docket]'`. This replaces the removed `fastmcp.experimental.events` module, whose check-function-based event types must now be ported to Pydantic models with `@events.event` and `@events.reader`.

Define a typed event, emit it from a tool, and consume it with a reader in the new fastmcp-events package.

```python
# pip install 'fastmcp[events]'
import fastmcp_events as events
from pydantic import BaseModel

class ToolInvoked(BaseModel):
    tool_name: str
    user: str

@events.event('tool_invoked')
class ToolInvoked(ToolInvoked):
    pass

@events.reader(ToolInvoked)
async def on_tool_invoked(event: ToolInvoked):
    print(f'{event.user} called {event.tool_name}')

# Inside a FastMCP tool:
await events.emit('tool_invoked', tool_name='search',
user='alice')
```

— Names every decorator, install extra, and migration path, with a runnable example

v4.2.6

02 **Client-side CodeMode via CodeModeClient** Not checked yet NEW

75

Wrap any Client or `ClientGroup` with the new `CodeModeClient` class to run CodeMode on the client side, giving agents search, schema, and execute meta-tools even over servers that do not run CodeMode themselves; a single `execute` block can chain tool calls across multiple servers.

— Explains mechanism and cross-server chaining but no runnable snippet given v4.2.6

THINNER COVERAGE BELOW

03 **OpenTelemetry network transport attributes on server spans**

Not checked yet NEW

50

Server spans now carry `network.transport` and, over HTTP, `network.protocol.name` and `network.protocol.version` OpenTelemetry attributes.

— Names exact attribute fields but no usage or config detail v4.2.6

04 **Authlib no longer bundled as dependency** Not checked yet BREAKING

35

Authlib is no longer installed with fastmcp; code that imports `authlib` directly must now declare it as its own dependency.

— Clear migration note but minimal explanation v4.2.6

└──▷ **BREAKING ON UPGRADE**

! Authlib is no longer installed with fastmcp — code that imports `authlib` directly must now declare it as its own dependency.

! The `fastmcp.experimental.events` module is removed; install the `events` extra (`pip install 'fastmcp[events]'`) and import from `fastmcp_events` instead. Each event type, formerly a check function, must be ported to a Pydantic model registered with @events.event('name') and a reader registered with @events.reader(Model).

WAS THIS USEFUL? 👍 👎

davinci-resolve-mcp

Feedback ⓘ v5.5.0 NOTES ↗ RANK

davinci-resolve-mcp is an MCP server integration for DaVinci Resolve Studio.

> davinci-resolve-mcp's v5.5.0 release expands `set_subtitle_preset` with a full set of Word Highlight caption controls, covering per-word outlines, spoken-word highlight styling, background boxes, and caption layout.

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

01

Word Highlight caption controls in set_subtitle_preset

Not checked yet NEW 92

`set_subtitle_preset` gains keys to control Word Highlight captions: `allWordsOutlineEnabled`, `allWordsOutlineRed`, `allWordsOutlineGreen`, `allWordsOutlineBlue`, and `allWordsOutlineThickness` draw an outline on every word by writing Text+ shading element 4 pinned to shape `ElementShape4 = 1`; `highlightFillEnabled`, `highlightSoftness`, `highlightGlow`, and `highlightBlend` set the spoken-word highlight appearance; `highlightBackgroundEnabled`, `highlightBackgroundRed/Green/Blue`, and `highlightBackgroundOffsetX/OffsetY/Width/Height/Round/Softness/Glow/Blend` add a coloured box behind the spoken word; and `wordsPerLine`, `writeOn`, and `showSpeaker` control caption layout.

Apply a full Word Highlight preset with a white outline on every word, a coloured background box behind the spoken word, and two words per line — useful for social-caption reels.

```
$ set_subtitle_preset(
  timeline_ai=True,
  allWordsOutlineEnabled=True,
  allWordsOutlineRed=1.0,
  allWordsOutlineGreen=1.0,
  allWordsOutlineBlue=1.0,
  allWordsOutlineThickness=2,
  highlightBackgroundEnabled=True,
  highlightBackgroundRed=0.0,
  highlightBackgroundGreen=0.4,
  highlightBackgroundBlue=1.0,
  highlightBackgroundWidth=10,
  highlightBackgroundRound=5,
  wordsPerLine=2
)
```

— Names every config key with a runnable example command v5.5.0

WAS THIS USEFUL? 👍 👎

mcp

Feedback ⓘ v1.3.0 NOTES ↗ RANK

Rapidly build MCP servers for your Laravel applications.

> Laravel MCP v1.3.0 adds support for the MCP Skills extension and exposes client capabilities to handlers and tests, alongside OAuth authorize flow improvements.

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

01 **Client capabilities exposed to handlers and tests** Not checked yet

NEW

34

Client capabilities are now exposed to handlers and tests, allowing inspection and conditional logic based on what the connecting client supports.

– Describes behaviour but no concrete field or API name v1.3.0

02 **OAuth authorize flow improvements** Not checked yet IMPROVED

34

The OAuth authorize view is now translatable so applications can localise the consent screen, and applications can resolve the resource metadata URL of the OAuth challenge.

– Two related OAuth changes described without specific config keys v1.3.0

03 **MCP Skills extension support** Not checked yet NEW

33

Adds support for the MCP Skills extension, enabling handlers to declare and consume skills.

– Names the feature but no API or usage detail given v1.3.0

WAS THIS USEFUL? 👍 👎

context7 Feedback ⓘ @upstash/context7-pi@0.2.0 NOTES ↗ RANK

Context7 maintains up-to-date documentation and context for AI agents through an MCP server.

> Context7 shipped version-pinned documentation retrieval across its CLI, MCP and pi tools, along with MCP skills extension support, VS Code Insiders setup, and a safer skill-removal behavior.

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

01 Version-pinned documentation retrieval Not checked yet BREAKING

95

Fetch docs for a specific library version using `--tag` on `ctx7 docs` (e.g. `ctx7 docs react --tag 18.2.0`), or pass a `version` input to the `query-docs` MCP/pi tool so agents retrieve docs matching their exact dependency version. Requests now send `version` as a dedicated parameter to `/v2/context` instead of appending it as a suffix of the library ID, and context7 serves that version or the nearest available one. As a consequence, `resolve-library-id` no longer lists versions in its response, so any caller or agent that parsed version suffixes from the library ID must now pass `version` separately.

Fetch docs pinned to a specific library version so your agent sees the exact API surface your project depends on.

```
$ ctx7 docs react --tag 18.2.0
```

Fetch docs pinned to a specific library version from the CLI rather than getting the default latest.

```
$ ctx7 docs --tag 4.7.0 upstash-redis
```

Fetch docs for a specific pinned version of a library from the CLI, so the returned content matches your project's exact dependency.

```
$ ctx7 docs <library-id> --tag 4.2.1
```

— Names flags, config input and endpoint, with runnable commands ctx7@0.13.0 @upstash/context7-mcp@4.8.0 @upstash/context7-pi@0.2.0

02 **VS Code Insiders setup target** Not checked yet NEW 85

Add VS Code Insiders as a setup target with `--vscode-insiders` on `ctx7 setup` and `ctx7 remove`; combine it with `--vscode` in one command, e.g. `ctx7 setup --vscode --vscode-insiders`, to configure both editors at once.

Register context7 as an MCP server in both VS Code and VS Code Insiders at the same time during initial setup.

```
$ ctx7 setup --vscode --vscode-insiders
```

— Exact flags and a runnable command given ctx7@0.13.0

THINNER COVERAGE BELOW

03 **Shared skill folder removal safety** Not checked yet BREAKING 50

`ctx7 remove` no longer deletes a skill from a shared skills folder (such as `~/.agents/skills`) while another agent still uses it.

— Behavior change described but no migration guidance ctx7@0.13.0

04 **MCP skills extension support (SEP-2640)** Not checked yet NEW 40

The context7-mcp skill is now served through the MCP skills extension per SEP-2640.

— No usage example or mechanism beyond the spec name @upstash/context7-mcp@4.8.0

└──▷ **BREAKING ON UPGRADE**

! `resolve-library-id` no longer lists versions in its response — callers that relied on version listings from this tool will no longer receive them.

! The library version is now sent as a standalone `version` parameter to `/v2/context` rather than as a suffix of the library ID — any agent or integration that appended the version to the library ID string will need to be updated.

! `resolve-library-id` no longer lists versions — callers that relied on versions appearing as suffixes of the library ID must now pass version separately via the `version` input to `query-docs`.

! `resolve-library-id` no longer lists versions — agents that parsed version suffixes from the library ID must now pass version as a separate value.

! `ctx7 remove` no longer deletes a skill from a shared skills folder (such as `~/.agents/skills`) while another agent still uses it.

WAS THIS USEFUL? 👍 👎

mcp-grafana

Feedback ⓘ v2.4.0 NOTES ↗

MCP server for Grafana

mcp-grafana v2.4.0 adds YAML-file tool selection, per-user credential enforcement, per-call audit logging, enforced Prometheus label matchers, dashboard schema v2 guardrails, and native Loki response format.

└──▷ GET THIS VERSION ▾

Homebrew

```
$ brew install mcp-grafana
# already installed? upgrade to this version:
$ brew upgrade mcp-grafana
```

git

```
$ git clone --branch v2.4.0 https://github.com/grafana/mcp-grafana.git
# already have the repo? check out this version:
$ git checkout v2.4.0
```

Homebrew lists this version · checked 26 Nov Wrong command?

Load tool selection from a YAML config file with `--config` or the `GRAFANA_MCP_CONFIG` environment variable, supporting categories, mode, query, allow and deny lists — unknown keys and names that would have no effect fail at startup, and explicit flags override the file. Not checked yet

Enforce per-user credentials on shared streamable HTTP servers with `--require-user-credentials`, which rejects requests carrying no credential with a 401, and name extra auth-proxy headers that count as a credential with `--user-identity-headers`. Not checked yet

Log one line per tool call — including the caller's Grafana user, tool, org, result and duration — with `--log-tool-calls` (arguments are never logged). Not checked yet

Read the server auth token from a header other than Authorization with `--server-auth-header`, leaving Authorization free for a user's own Grafana credential behind a proxy. Not checked yet

AND operator-set label matchers into every Prometheus query and discovery call with `--prometheus-enforced-matchers`, mirroring `--loki-enforced-matchers`, and failing closed on queries that cannot be parsed. Not checked yet

+6 more

NEW INTEGRATION AWS X-Ray integration: search_xray_traces, get_xray_trace, xray

NEW INTEGRATION Azure Monitor integration: query_azure_log_analytics, query_azure_resource_graph, azuremonitor

NEW SUPPORTED PLATFORM Grafana 13.3 Notebooks support: list_notebooks, search_notebooks, get_notebook_by_uid, render_notebook_pdf, create_notebook and 2 more

└──▷ **BREAKING ON UPGRADE**

- ! `alerting_rules_read` now returns `{rules, truncated, next_cursor}` instead of a plain array; code that parses the output directly must be updated.
- ! `query_cloudwatch` now always returns results under a `series` key; code that parses the output directly must be updated.
- ! The default `--grafana-timeout` for the main Grafana client rises from 10s to 30s.
- ! Go embedders must call tools.Resolve(tools.Policy{...}).Register(s) instead of the former `Add*Tools` functions.

WAS THIS USEFUL? 👍 👎

elementor-mcp

Feedback ⓘ v3.22.1 NOTES ↗ RANK

Elementor-mcp wordpress MCP server with 234 Free tools and 547 total with Pro for building pages, managing content, and operating your site. Availability depends on installed integrations and enabled tools.

> Elementor-mcp's latest release adds a full-field code snippet editor and new filesystem backup management tools, alongside hardened storage for existing file-operation backups.

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

01 **Hardened filesystem backup storage under wp-content**

Not checked yet IMPROVED 75

Filesystem backups taken by Write File, Edit File, and Delete File now live under `wp-content` in a site-key-derived folder with `.bak` extensions, deny files for Apache and IIS, and automatic daily cleanup after 30 days.

– Full mechanism and retention policy described, but automatic with no user command.

v3.22.1

02 **Filesystem backup listing and purge tools** Not checked yet NEW 70

`list-fs-backups` lists filesystem backups, and `purge-fs-backups` removes old or named ones (switched off by default); purge output also names the History entries that lose their undo capability when a backup is removed.

– Names both tools and default-off behavior, but no example given. v3.22.1

└──▷ **ALSO FROM THESE RELEASES**

Permanently delete a stuck or failed backup that a host timed out, then confirm the removal.

```
$ delete-backup id=7 permanent=true confirm=true
```

Run a full backup in blocking mode from WP-CLI so the script waits for completion rather than polling.

```
$ create-backup wait=true
```

WAS THIS USEFUL? 👍 👎

api-mcp-server

Feedback ⓘ v4.17.0 NOTES ↗ RANK

MCP server that exposes Hostinger's platform API for hosting deployment, CDN control, domain management, and WordPress operations.

> api-mcp-server added a single new operation for checking website build task status.

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

01 **Build task status operation for Horizons websites** Not checked yet

NEW 55

Adds the `horizons_websites_build-task` operation to get the status of a website build task.

— Names exact operation but no further mechanism detail. v4.17.0

WAS THIS USEFUL? 👍 👎

fablecut

Feedback ⓘ v1.21.0 NOTES ↗ RANK

Fablecut is a zero-dependency browser video editor that AI agents can drive — JSON timeline, MCP + REST, live-reloading UI

> FableCut added support for importing openly licensed music and sound effects from Openverse, with automatic licence credit tracking for video descriptions.

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

01 **Openverse audio import with licence tracking** Not checked yet

NEW INTEGRATION NEW 45

FableCut can import openly licensed music and sound effects from Openverse. The tool automatically retains each imported track's licence credit so it can be included in the video's description.

– Describes behaviour and source but no UI path or command product docs

NEW INTEGRATION Freesound SFX search

v1.21.0 NOTES ↗ LATEST STABLE

FableCut adds openly licensed music and SFX search and import with per-track licence display, credit copy, Export dialog credit listing, and agent tools `fablecut_music` and `fablecut_export`.

└──▷ **GET THIS VERSION** ▾

npm

```
$ npm install fablecut@1.21.0
```

git

```
$ git clone --branch v1.21.0 https://github.com/ronak-create/FableCut.git
# already have the repo? check out this version:
$ git checkout v1.21.0
```

signed build provenance · checked 26 Nov Wrong command?

Search and import openly licensed music from Openverse (Jamendo, ccMixter, Wikimedia Commons) via the new `fablecut_music` agent tool, which also lists credits; `fablecut_export` now appends those credits to its reply. Not checked yet

New Music tab in the bin searches openly licensed tracks by mood, genre or instrument, with a length filter and OK to monetize flag, while the existing SFX tab now searches Freesound sound effects the same way. Not checked yet

Every music search result shows its licence terms (free to use, credit required, non-commercial, share alike, samples only, or no edits), and imported tracks carry that licence into Media info with a Copy credit action. Not checked yet

The Export dialog now lists the edit's music credits for use in a video description. Not checked yet

The bin's tabs are now grouped into Project, Audio (Music, SFX) and Graphics (Elements, SVG, Shapes, Templates) rows so the tab row no longer crowds a narrow panel. Not checked yet

---

WAS THIS USEFUL? 👍 👎

◆ Agent Memory

claude-mem

Feedback ⓘ v17.0.0 NOTES ↗ RANK

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

> claude-mem added a macOS iMessage integration that turns Mac Messages and group chats into searchable, per-project notes, introduced a Review Stack in My Work for human review of agent submissions, and rolled out a 30-day Cloud Sync trial alongside new environment variables for tuning observer reasoning effort and My Work API limits.

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

01 **iMessage integration for macOS** Not checked yet NEW 92

claude-mem can now observe iMessage group chats via a Photon iMessage line (`npx claude-mem imessage setup`), giving each group its own Claude-Mem project, and can observe selected Mac Messages app chats from Terminal using `claude-mem imessage mac` subcommands: `check` , `chats` , `pick` (with `--focus` to set what to watch for), `observe` , and `contacts` (also reachable via `npx claude-mem imessage mac` for checking Full Disk Access, listing chats, saving picks, listening for new messages, and importing contact names). A new Claude-Mem Messages Mac app lists iMessage chats and group chats, lets you tick which to observe with a per-chat focus, and turns new messages into notes in a per-chat project; it is built with `integrations/imessage-mac/scripts/build-app.sh` .

Start observing an iMessage group chat so every decision, plan, and shared link is captured as notes in a dedicated Claude-Mem project.

```
$ npx claude-mem imessage setup
```

Pick a specific Mac Messages chat to observe and tell claude-mem to watch only for financial details like who paid and who owes.

```
$ claude-mem imessage mac pick --focus 'money: who paid, who owes'
```

Start listening to a Mac Messages chat so new messages are saved as Claude-Mem notes — useful for capturing decisions made in group chats without leaving your terminal.

```
$ npx claude-mem imessage mac
```

Set up a Photon iMessage line so claude-mem can observe group conversations via your Photon credentials.

```
$ npx claude-mem imessage setup
```

— Full CLI command set, flags and build script named `v16.2.0` `v17.0.0`

---

02 **Configurable reasoning effort for OpenAI-compatible observers**

`Not checked yet` NEW 75

Set reasoning effort for OpenAI-compatible observers and API agents with `CLAUDE_MEM_OPENAI_COMPAT_REASONING_EFFORT`, accepting values `none`, `low`, `medium`, `high`, `xhigh`, or `max`, configurable via the viewer, settings file, or environment.

Tell OpenAI-compatible observers to use low reasoning effort to reduce latency and cost on routine observation calls.

```
$ export CLAUDE_MEM_OPENAI_COMPAT_REASONING_EFFORT=low
```

— Env var and all accepted values named with command `v16.1.3`

---

03 **My Work API timeout and output token limits** `Not checked yet` NEW 75

Cap My Work API refresh time and answer length with `CLAUDE_MEM_WORK_VIEW_API_TIMEOUT_MS` (default 600000 ms) and `CLAUDE_MEM_WORK_VIEW_MAX_OUTPUT_TOKENS` (default 16000), both active when `CLAUDE_MEM_WORK_VIEW_AGENT=api`.

Tighten the My Work API agent's answer budget to avoid runaway long answers in constrained environments.

```
$ export CLAUDE_MEM_WORK_VIEW_AGENT=api
export CLAUDE_MEM_WORK_VIEW_API_TIMEOUT_MS=120000
export CLAUDE_MEM_WORK_VIEW_MAX_OUTPUT_TOKENS=4000
```

— Two env vars with defaults and activation condition named `v16.1.3`

---

04 **Acknowledge in-tool messages** `Not checked yet` NEW 65

A new Messages from Claude-Mem queue delivers in-tool notifications such as Cloud Sync offers at interactive startup, which can be confirmed with `claude-mem messages acknowledge <message-id>`.

Acknowledge a Cloud Sync offer or other notification delivered to the Messages from Claude-Mem queue at interactive startup.

```
$ claude-mem messages acknowledge <message-id>
```

— Exact command with argument given v16.1.5

THINNER COVERAGE BELOW

05 **Extended default release monitoring window** Not checked yet

IMPROVED

55

Release monitoring now runs for 24 hours by default instead of 30 minutes, using graduated check intervals: frequent in the first hour, every 30 minutes until hour 6, every 2 hours after that, with a final sweep at the 24-hour mark.

— Clear before/after but no user-facing control v16.2.1

06 **Review Stack in My Work** Not checked yet NEW

53

A new Review Stack in My Work lets agents submit files, images, and links for human review. Each card supports actions to mark it Seen, save it for later, or leave a follow-up, and a hold-for-three-seconds shortcut marks an entire Focus Area seen across paired devices.

— UI behaviour described but no exact command or path v17.0.0

07 **Cloud Sync trial and My Work account handling** Not checked yet

NEW

50

Setup now offers a 30-day Cloud Sync trial without requiring a credit card, and direct or agent installs can choose a private, signed-in My Work view while preserving existing local and custom sync settings. My Work shows optional guidance for trying cross-agent sync during the trial and surfaces when to add a payment method to continue, while the cloud viewer now verifies the signed-in account when returning to a tab before showing its contents and retries a failed save by reusing the same action ID.

— Behaviour described but no named config or command v16.1.4

WAS THIS USEFUL? 👍 👎

EverMind AI EverOS

Feedback ⓘ v1.8.0 NOTES ↗ RANK

EverOS provides a local-first memory layer that stores and evolves context for AI agents.

> EverOS's v1.8.0 release delivered major search throughput gains (4× hybrid, 2× keyword) driven by a jemalloc-preloading CLI, added a per-request raw input log and optional MemTrace execution tracing for `/memory/add`, and introduced several new config limits alongside breaking index-rebuild and settings requirements for upgraders.

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

01 Raw input log for /memory/add requests

Not checked yet BREAKING

95

Every accepted `/memory/add` request is now logged to a per-day JSONL file at `<root>/<app>/<project>/.input_log/YYYY-MM-DD.jsonl` before extraction starts, controlled by `[memorize.input_log]` with keys `enabled` (default `true`), `retention_days` (default 30), and `max_inline_bytes` (default 1 MiB); attachments above the limit are replaced with their length and SHA-256. The log is on by default, so operators who do not want it must explicitly set `enabled = false`.

Tune the input log retention and inline-size limit to avoid bloating disk with large attachment payloads.

toml

```toml
[memorize.input_log]
enabled = true
retention_days = 14
max_inline_bytes = 524288
```

— Exact config path, keys, defaults and a runnable override example v1.8.0

02 Optional MemTrace execution tracing

Not checked yet NEW

90

Execution tracing can now record boundary decisions, extraction evidence, buffer changes, and retrieval-stage candidates as MemTraceBench-compatible JSON under `<root>/memtrace/`, enabled by installing `pip install 'everos[memtrace]'` and setting `EVEROS_MEMTRACE_ENABLED=1`; the output directory can be overridden with `EVEROS_MEMTRACE_DIR`.

Capture a full execution trace for a troublesome memory-add call to inspect how extraction decisions were made.

```
$ EVEROS_MEMTRACE_ENABLED=1
EVEROS_MEMTRACE_DIR=/var/everos/traces everos server start
```

— Named install command and env vars with a runnable example v1.8.0

03 **Hybrid and keyword search throughput gains** Not checked yet

IMPROVED

80

EverOS now sustains 14–16 `hybrid` searches per second (up from under 4) and 120–140 `keyword` searches per second (up from 70) on a 4-core, 8 GB server with 169k memories. Much of this comes from the CLI preloading jemalloc on Linux when `libjemalloc2` is present, re-executing with `LD_PRELOAD=libjemalloc.so.2` and `MALLOC_CONF=oversize_threshold:0,dirty_decay_ms:1000,narenas:4`, which lifts keyword search to 140 vs 120 req/s and hybrid to 16 vs 14 req/s compared with glibc alone.

— Numbers and mechanism given but tuning requires manual jemalloc install v1.8.0

04 **Configurable truncated LLM reply retries** Not checked yet NEW

75

Control how many times a truncated LLM reply is retried before failing with `[llm] truncated_reply_retries` (default `1`), overridable per task via `[llm.<task>]` and in `[multimodal]`.

Allow one extra retry for tasks that frequently hit the output token cap, without affecting other LLM tasks.

```toml
[llm]
truncated_reply_retries = 1

[llm.episode_extraction]
truncated_reply_retries = 3
```

— Named config key with default and per-task override example v1.8.0

05 **Disk cap for failed async queue payloads** Not checked yet NEW

70

Cap disk used by failed `async_mode` queue bodies with `memorize.async_queue.max_failed_payload_bytes` (default 64 MiB), keeping only the newest failed payloads while retaining the row and its `last_error`.

— Named config key and default but no usage example v1.8.0

THINNER COVERAGE BELOW

06 **Index rebuilds required after upgrade** Not checked yet BREAKING

40

The vector index is rebuilt once automatically after upgrading, as the index format changes from `IVF_FLAT` to `IVF_RQ`; separately, a store indexed with the `nltk` tokenizer must be rebuilt manually after upgrading.

— Names the formats but gives no rebuild command or procedure v1.8.0

07 **New setting required for cross-origin and file:// clients**

Not checked yet BREAKING

30

A web UI served from a different origin, a reverse proxy that forwards its own domain, and any client sending `file://` content items each now require a new setting to continue working after this upgrade.

— Describes the impact but the required setting itself is unnamed v1.8.0

└──▷ **BREAKING ON UPGRADE**

! A web UI served from a different origin, a reverse proxy that forwards its own domain, and any client sending `file://` content items each require a new setting to continue working.

! The vector index is rebuilt once automatically after upgrading (the index format changes from `IVF_FLAT` to `IVF_RQ`).

! The raw input log (`[memorize.input_log]`) is on by default (`enabled = true`); operators who do not want it must explicitly set `enabled = false`.

! A store indexed with the `nltk` tokenizer must be rebuilt after upgrading.

WAS THIS USEFUL?  

Zep Graphiti

Feedback ⓘ v0.32.1 NOTES ↗ RANK

Graphiti builds temporally-aware knowledge graphs from conversations and documents so agents can query facts that change over time.

> Graphiti v0.32.1 adds concurrent ingestion for a single graph and an embedded Relate graph driver, while tightening UUID handling so cross-group lookups and invalid episode references now raise errors instead of silently succeeding.

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

01 Stricter group-scoped UUID enforcement

Not checked yet **BREAKING**

83

UUID lookups for episodes, facts (`EntityEdge.get_by_uuid` and `get_by_uuids`), sagas (`summarize_saga`), and `add_triplet` now enforce group boundaries — a UUID from a different `group_id` raises `NodeNotFoundError` or `EdgeNotFoundError` instead of silently reading or modifying data. An unknown predecessor episode UUID now also raises `NodeNotFoundError` where it was previously silently skipped, and calling add_episode(uuid=...) without `group_id` now searches only the default group instead of across all groups.

— Names every affected method and the exact exceptions raised v0.32.1

THINNER COVERAGE BELOW

02 Embedded Relate graph driver backend

Not checked yet **NEW**

40

Graphiti adds an embedded Relate graph driver as a supported backend alongside Neo4j, FalkorDB, and Amazon Neptune.

— Names the new backend but no setup detail v0.32.1

03 Concurrent ingestion for a single graph

Not checked yet **NEW**

35

Graphiti now supports concurrent ingestion for a single graph, allowing multiple episodes to be ingested in parallel without needing separate graph instances.

— Describes the capability but no API or flag given v0.32.1

04 Kuzu driver removed

Not checked yet **BREAKING**

35

The Kuzu graph driver has been removed from Graphiti; any code that instantiated a Kuzu graph backend will break on upgrade.

— Clear removal notice but no migration path given v0.32.1

└──▷ BREAKING ON UPGRADE

! The Kuzu driver has been removed; any code that instantiated a Kuzu graph backend will break on upgrade.

! Calling `EntityEdge.get_by_uuid`, `get_by_uuids`, `summarize_saga`, or `add_triplet` with a UUID that belongs to a different `group_id` now raises `NodeNotFoundError` or `EdgeNotFoundError` instead of silently succeeding.

! An unknown predecessor episode UUID now raises `NodeNotFoundError`; previously Graphiti silently skipped it.

! add_episode(uuid=...) called without `group_id` now searches only the default group instead of across all groups.

WAS THIS USEFUL?  

◆ AI Agent Frameworks

HeyGen HyperFrames

Feedback ⓘ v0.9.97 NOTES ↗ RANK

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

> HyperFrames' latest releases focus on partial-film rendering across the CLI, Producer, and Studio Server, alongside deeper sealed-preview inspection and more memory-efficient long-audio handling in Studio.

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

01 **Partial film rendering via CLI, Producer, and Server route**

Not checked yet NEW 75

HyperFrames can now render just a segment of a film instead of the whole thing, exposed consistently across three surfaces: the CLI's render command accepts a start and end point with control over format, frame rate, and HDR; the Producer can render a part of the film with HDR video or shader transitions in a single pass; and the Studio Server's render route now accepts a start and end to render that segment.

– Names CLI, Producer, and Server route but not exact flags v0.9.97 v0.9.94

02 **Sealed preview inspection gains display and text-style reports**

Not checked yet IMPROVED 60

HyperFrames Studio's sealed preview, which lets you inspect a film without unsealing it, now reports the computed display value of every element a selector matches, and separately now answers a text toolbar's authored styles and fonts so rich-text formatting is visible without unsealing.

– Describes two sealed-preview reports without exact API names v0.9.95 v0.9.92

THINNER COVERAGE BELOW

03 **Streamed audio effects and beat data for long audio**

Not checked yet IMPROVED 45

Long audio in Studio now streams effects and beat data from the server instead of loading it all up front, avoiding a page out-of-memory limit.

– Explains behavior change but no concrete limits or config named v0.9.92

WAS THIS USEFUL? 👍 👎

octop-harness

Feedback ⓘ v1.0.12 NOTES ↗ RANK

Octop-harness is a production-grade agent runtime engineered from the Harness Engineering philosophy.

> Octop-harness v1.0.12 adds a CubeSandbox remote sandbox backend with object-storage workspace mirroring, plus per-turn tool-call limits, read-only memory recall metadata, and filesystem checkpointing.

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

01 CubeSandbox remote backend support

Not checked yet NEW

95

Add the CubeSandbox backend by installing `octop-harness[cubesandbox-native,remote-backends]` or `octop-harness[cubesandbox,remote-backends]`, then resolve it with a backend spec like `{"type": "cubesandbox", "template": "...", "octop_user": "...", "octop_agent": "..."}`, setting `CUBE_API_URL`, the SDK API key, and `OCTOP_CUBESANDBOX_LOCK_DSN` on the harness host. Optionally add `workspace_backup` with a nested backend spec such as `{"type": "cos", "bucket": "..."}` to mirror the sandbox workspace into object storage. Set `octop_scope` to `agent` (default) to reuse a sandbox per agent, or `user` to share the workspace across all of a user's agents; call destroy() to terminate the sandbox explicitly or archive_workspace() to push the workspace back to object storage before the sandbox closes.

Use the CubeSandbox backend to run an agent in an isolated remote sandbox with object-storage workspace mirroring.

python

```python
config = HarnessAgentConfig.from_env(
    workspace_dir=root,
    name="sandboxed-agent",
    backend={
        "type": "cubesandbox",
        "template": "default",
        "octop_user": "alice",
        "octop_agent": "my-agent",
        "octop_scope": "agent",
        "workspace_backup": {"type": "cos", "bucket": "my-backup-bucket"},
    },
)
```

— Names every config key, env var, method, with a runnable example v1.0.12

THINNER COVERAGE BELOW

02 **Per-turn allowed_tools limit** Not checked yet NEW 55

Agents can now be capped on which tools they may call each turn via a new per-turn `allowed_tools` limit, enforced at the tool execution entry point.

— Names the flag and enforcement point but gives no example v1.0.12

03 **Per-turn filesystem checkpoints** Not checked yet NEW 45

The filesystem middleware now captures per-turn file write checkpoints, enabling change-record generation and workspace recovery.

— Explains behavior but no config key or usage shown v1.0.12

04 **Read-only memory recall candidate metadata** Not checked yet NEW 35

Host middlewares can now access read-only memory recall candidates through a new candidate metadata interface.

— Describes purpose but names no concrete interface or API v1.0.12

WAS THIS USEFUL?  

claw-orchestrator

Feedback ⓘ v7.35.0 NOTES ↗ RANK

Claw-orchestrator is a run Claude Code, Codex, Antigravity, Grok Build and OpenCode behind one runtime, and a run only counts as done when checks it didn't write pass. Cross-engine fan-out, councils, session handoff, spend caps, MCP server, ACP agent.

> claw-orchestrator re-enabled Cursor as a supported engine with sandboxing, auto-review and live model switching, while shipping per-run spend caps, git worktree isolation, metric-based contract checks, fan-out timeout and output-file handling, automatic Claude account rotation on usage limits, a universal plugin tool command, and a breaking change that strips secrets from sandboxed check environments.

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

01 Cursor engine re-enabled with sandboxing and auto-review

Not checked yet NEW 95

Cursor returns as a supported engine, enabled via `engine: 'cursor'` in config or `--cursor` on the CLI, targeting Cursor Agent `cursor-agent` 2026.11.22; goals run turn by turn with `clawo goal --cursor`. Workspace-only edits are enforced with `sandboxMode: 'workspace-write'` (blocks writes outside the workspace and all network access), Cursor's Auto-review is enabled via `permissionMode: 'auto'`, and extra directories map into sessions with `addDir` (-> `--add-dir`) and `pluginDir` (-> `--plugin-dir`).

Run a goal against Cursor with workspace-sandboxing and auto-review so the agent can edit files but cannot reach the network or write outside the repo.

```
$ clawo goal --cursor
```

Configure the Cursor engine in a session config to enforce sandbox mode and auto-review by default.

```yaml
engine: 'cursor'
sandboxMode: 'workspace-write'
permissionMode: 'auto'
```

— Every flag and config key for the engine is named. v7.35.0

02 Sandboxed checks no longer inherit server environment

Not checked yet 95

BREAKING

Checks run where the agent is sandboxed (Codex, `codex-app` by default, any run with explicit `read-only` / `workspace-write` `sandboxMode`, or a `launcher`) no longer inherit the server's environment — variables whose names contain `KEY`, `SECRET`, `TOKEN`, `PASSWORD`, `AUTH` and similar are stripped. Checks needing those vars must declare them via `passEnv` or `--pass-env`, or set `checkEnv: "inherit"` / `--check-env inherit`.

Allow a sandboxed check to use a specific secret it needs while keeping all other credentials out of the check environment.

```
$ clawo goal --claude-account auto --pass-env OPENAI_API_KEY --
pass-env AWS_* 'Benchmark model outputs'
```

— Exact stripped patterns and migration flags given. v7.33.0

---

03 **Per-run ledger query by engine, model or workflow** Not checked yet

NEW 90

Query per-run ledger data — verdict, cost, duration, and spend by engine and model — with `clawo runs --by model` (or `engine`, `workflow`, `run`) or via `GET /runs/outcomes`; data is written to `~/.claw-orchestrator/runs/outcomes/`.

— Command, endpoint and storage path all named. v7.32.0

---

04 **Session control: model switching and interruption** Not checked yet

NEW 88

Session control adds live model switching mid-conversation via `POST /session/model` with `"now": true` (`switchModel`) — omitting `"now"` stores the model for the next restart only — with session status exposing the engine-reported model via `reportedModel` when it differs from the configured model and recorded per goal run in the ledger; it also adds the ability to stop a running session turn mid-flight without ending the session using `clawo session-interrupt <name>` or `POST /session/interrupt`, with the session and its conversation continuing on the next send and unattended turns (goals, inbox deliveries) also stopped.

Swap the active model immediately in a running session without losing conversation history.

```
$ curl -X POST https://<host>/session/model \
  -H 'Content-Type: application/json' \
  -d '{"model": "claude-sonnet-4-5", "now": true}'
```

Interrupt a runaway or long-stalled session turn from the CLI without killing the session itself

```
$ clawo session-interrupt my-session
```

Interrupt a session turn programmatically via the REST API, for example from a monitoring script that detects a stalled turn

```
$ curl -X POST http://localhost:<port>/session/interrupt -H
'Content-Type: application/json' -d '{"name": "my-session"}'
```

— Endpoint, field names and behavior difference all given. Exact command and endpoint with behavior described.

v7.35.0 v7.34.0

---

05 **Agy session compaction fixed and configurable** Not checked yet

IMPROVED 88

Agy sessions can now be compacted through agy's own `/compact`, run as a turn with slash commands enabled — `compactSession` was previously a no-op on agy. The compaction threshold is configurable via `autoCompactWindow` in an Antigravity session config, exposed as the `AGY_CLI_AUTO_COMPACT_OVERRIDE` env var (requires agy 1.5.15+), accepting an integer of at least 100000 or `auto`.

— Config key, env var and version requirement named. v7.35.0 v7.32.0

---

06 **Per-run spend caps** Not checked yet NEW 86

Cap total run spend with `maxBudgetUsd` on a spec or `--max-budget` on `solve`, `goal`, and `fanout`, failing the run before any node once the limit is reached.

Cap total LLM spend for a solve run so a runaway agent cannot exceed a budget before you review results.

```
$ clawo solve --max-budget 2.00 --wait
```

— Config key and flag with fail behavior named. v7.32.0

---

07 **Git worktree isolation for runs** Not checked yet NEW 86

Run workflows in an isolated git worktree with `isolation: "worktree"` on a spec or `--isolate` on `clawo solve` and `clawo goal`, committing results to a local branch `clawo/<runId>` without touching the caller's checkout.

— Config key, flag and branch naming scheme given. v7.32.0

08 **Contract system enhancements** Not checked yet NEW 86

Contracts gain a new `metric` check type to hold a numeric output from a command to `min` / `max` bounds, with `minDelta` and `maxRegression` comparing against the base commit in a temporary worktree; a `protectPaths` setting protects additional files from agent modification, such as a judge's config or fixtures, alongside linter configs (`tsconfig*.json` , ESLint, Prettier, Biome, Ruff, mypy, flake8, pylint, golangci, stylelint); and Autoloop now accepts `contract` on `autoloop_start` and `POST /autoloop/new` to configure the acceptance gate with a base commit and test snapshot.

Enforce that a benchmark stays within a regression budget — fail the run if the measured value is more than 5% worse than on the base commit.

```yaml
metric:
  command: "python bench/run.py --json"
  path: "$.results.latency_ms"
  maxRegression: 5
```

— Check type and bound fields named with example. Config key named with example protected file types. Field and endpoint named but mechanism thin.

v7.32.0

09 **Automatic Claude account rotation on usage limit** Not checked yet

NEW 84

A goal's Claude session rotates to the next available account automatically when a usage limit is hit, keeping goals running without waiting for the reset. Enable it by setting `claudeAccount` on a goal node, the `goal` template of `workflow_start` , or passing `--claude-account auto` to `clawo goal` .

Keep a long-running goal alive across multiple Claude accounts without manual intervention when a usage limit is reached.

```
$ clawo goal --claude-account auto 'Refactor the authentication module and write tests'
```

— Config key, template and flag for rotation named. v7.33.0

10 **Universal plugin tool command** Not checked yet NEW 83

Call any plugin tool from the shell with `clawo tool <name> '<json>'` , which routes through `POST /tool/<name>` on the server; list available tools with `GET /tool` .

Call the ultrareview plugin tool directly from the shell without a dedicated subcommand.

```
$ clawo tool ultrareview '{"focus": "security", "paths":
["src/"]}'
```

— Command and both endpoints named with usage. v7.32.0

11 **Fan-out outputs saved as files** Not checked yet NEW 81

Fan-out agent outputs are now saved as files (`outputs/<agent>.md` , `outputs/synthesis.md`) in the run directory; `fanout_status` results expose `outputFile` and `synthesisFile` fields, and `clawo fanout --wait` prints each file's location.

— File paths and result fields named precisely. v7.32.0

12 **Configurable fan-out per-agent timeouts** Not checked yet IMPROVED 78

Fan-out per-agent timeouts are configurable via `clawo fanout --agent-timeout <minutes>` ; a timed-out agent now reports `timedOut: true` , an error naming the timeout, and the text it had streamed before the turn ended.

— Flag and reported fields named, no example given. v7.32.0

└──▷ **BREAKING ON UPGRADE**

! Checks in runs where the agent was sandboxed (Codex, `codex-app` by default, or any run with an explicit `read-only` / `workspace-write` `sandboxMode` , or a `launcher`) no longer inherit the server's environment — vars whose names contain `KEY` , `SECRET` , `TOKEN` , `PASSWORD` , `AUTH` and similar are stripped. Checks that relied on those vars being present must now declare them via `passEnv` or `--pass-env` , or set `checkEnv: "inherit"` / `--check-env inherit` .

WAS THIS USEFUL? 👍 👎

atomic-agent

Feedback ⓘ v0.9.0 NOTES ↗ RANK

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

> Atomic Agent's v0.9.0 release adds an OS-level command sandbox, policy-based governance, a Privacy Shield for cloud model calls, named API tokens, agent-to-agent serving, editor integration, and new task channels, while desktop v0.0.10 adds a mini chat overlay, dictation, local file indexing, reusable recipes, and incognito chat.

└──▷ **WHAT SHIPPED · 20 FEATURES** most completely described first ? what's the number?

01 Command sandbox and policy controls

Not checked yet NEW

95

`ATOMIC_AGENT_SANDBOX=workspace` sandboxes shell commands and local MCP servers to the workspace with network disabled unless `ATOMIC_AGENT_SANDBOX_NETWORK=on` is also set. Outbound data-carrying links are now controlled via `http.urlDataGuard` set to `off`, `data`, or `strict`, and prompt for confirmation by default. An admin `policy.json` restricts or allows shell, edits, web, skills, and MCP servers, inspectable with `atag policy show`, and agent web traffic can be routed through a proxy or VPN via `network.proxy.url`.

Sandbox all shell commands to the working directory with no network, to prevent accidental data exfiltration during automated agent tasks.

```
$ ATOMIC_AGENT_SANDBOX=workspace atag serve
```

Enforce an org-wide policy denying shell and MCP server access before distributing a shared agent install.

```
$ atag policy show
```

— Names every config key, flag, and default change v0.9.0

02 Agent serving and API key enforcement

Not checked yet BREAKING

95

`atag serve` now supports agent-to-agent task submission via `atag serve --a2a`, exposes read-only memory and skills to MCP clients at `/mcp`, and can print its auto-generated key via `atag serve --print-api-key` (saved to `~/.atomic-`

`agent/serve-api.key`). As a breaking change, `atag serve` now always requires an API key — auto-generated on first run if not supplied via `--api-key` or `ATOMIC_AGENT_API_KEY` — and all scripts calling serve must send it as `Authorization: Bearer <key>`.

— Endpoints, flags, and migration requirement all named v0.9.0

03 **Chat and memory export/import formats** Not checked yet NEW

90

Export chat history as Markdown, Word, PDF, or HTML using `/export` or `atag export --format pdf`; back up notes and profile facts with `atag memory backup`; export setup without keys using `atag setup export`; and import personal facts from a ChatGPT or Claude export via `/import`.

— Every command and supported format named verbatim v0.9.0

04 **WhatsApp channel and mail task controls** Not checked yet BREAKING

75

Adds WhatsApp as an experimental task channel via `channels.whatsapp` in settings (available in Docker or source installs), ingests tasks from a personal mailbox over IMAP (such as Gmail with an app password), and filters which senders auto-start mail tasks with `email.autoStartSenders` in mail task settings. As a breaking change, mail tasks from senders other than the owner's own address no longer auto-start and wait for manual start unless listed in `email.autoStartSenders`.

— Config keys and default behaviour change both named v0.9.0

05 **Privacy Shield for cloud models** Not checked yet NEW 70

Privacy Shield redacts emails, phone numbers, card numbers, IBANs, and keys before sending data to cloud models. Toggle it in `/privacy` with `m`, or control it via `skills.selfLearning.enabled`.

— Mechanism and toggle named, scope of redaction unclear v0.9.0

06 **Named API tokens with roles** Not checked yet NEW 70

Issue named API tokens with roles using `atag http tokens add <name> --role admin|member|viewer`.

Issue a viewer-scoped API token for a CI pipeline that reads agent memory without admin access.

```
$ atag http tokens add ci-reader --role viewer
```

— Exact runnable command with role options given v0.9.0

07 **OpenAI API key scoping** Not checked yet BREAKING 70

`OPENAI_API_KEY` is now sent only to OpenAI or the local computer; an OpenAI-compatible server now requires its own key in its entry or via `OPENAI_COMPAT_API_KEY`.

— Before/after credential scoping named, migration implied v0.9.0

08 **Mini chat overlay** Not checked yet NEW 70

Open a mini chat overlay with ⌘⇧Space (Ctrl+Shift+Space on Windows and Linux) that stays in front of other apps and can rewrite, translate, explain, or summarise selected text.

— Shortcut and capabilities both named desktop-v0.0.10

09 **Side-question and context-compaction commands** Not checked yet NEW 65

Run `/btw` to ask a side question that is not saved to the chat, or `/compact` to summarise the context, specifying what to keep.

Ask a one-off side question during a session without polluting the saved chat history.

```
$ /btw what port does the target service default to?
```

— Two exact commands with behaviour explained desktop-v0.0.10

10 **Editor integration via atag acp** Not checked yet NEW 60

Connect Zed, JetBrains IDEs, and Neovim to Atomic Agent as their agent with `atag acp`.

— Command and supported editors named, no setup detail v0.9.0

11 **Per-chat reasoning effort control** Not checked yet NEW 60

Set the reasoning effort per chat with `/think auto|low|medium|high`.

— Exact command and options given, mechanism unexplained v0.9.0

12 **Screen capture and button interaction** Not checked yet NEW 60

Enable screen capture and button interaction on Windows and Linux (and macOS button-press by number) with `tools.screen.enabled`.

— Config key named, no usage walkthrough v0.9.0

13 **System-wide dictation** Not checked yet NEW 60

Enable dictation into any app via Settings › General — hold Fn on Mac or Ctrl+Win on Windows to speak and type in place.

— Exact shortcut and settings path given desktop-v0.0.10

THINNER COVERAGE BELOW

14 **Self-hosting via Docker** Not checked yet NEW 55

Self-host the agent on a VPS or second computer using the provided Dockerfile and `compose.yaml`.

— Files named, deployment steps not detailed v0.9.0

15 **Reusable recipes** Not checked yet NEW 55

Save any chat as a reusable recipe via Save as recipe..., then run it from Recipes in the sidebar or share it as a link.

— UI path given, sharing mechanism unexplained desktop-v0.0.10

16 **File references with @ in chat** Not checked yet NEW 50

Add a file to any chat by typing `@` in the message box to reference files from the chat's folder.

— Syntax named, limited mechanism described desktop-v0.0.10

17 **Local file indexing for Q&A** Not checked yet NEW 50

Index local folders in Settings › My files so the agent can answer questions about your documents and open PDFs at the cited page.

— Settings path named, indexing mechanism unexplained desktop-v0.0.10

18 **Incognito chat mode** Not checked yet NEW 40

Start incognito chats that are not saved to history or memory and are deleted on exit.

— Behaviour described, no activation path named desktop-v0.0.10

19 **Canvas Expand view** Not checked yet NEW 40

Use Canvas's Expand to open a page or chart from a reply beside the chat.

— Feature named, interaction detail thin desktop-v0.0.10

20 **Office file preview in side panel** Not checked yet NEW 40

Preview Word, PowerPoint, and Excel files in the side panel.

— Formats named but no further mechanism given desktop-v0.0.10

└──▷ BREAKING ON UPGRADE

! `atag serve` now always requires an API key; without `--api-key` or `ATOMIC_AGENT_API_KEY`, a key is auto-generated on first run to `~/.atomic-agent/serve-api.key`, and all scripts calling serve must send it as `Authorization: Bearer <key>`.

! `OPENAI_API_KEY` is now sent only to OpenAI or the local computer; an OpenAI-compatible server now requires its own key in its entry or via `OPENAI_COMPAT_API_KEY`.

! Mail tasks from senders other than the owner's own address no longer auto-start; they wait for manual start unless the sender is listed in `email.autoStartSenders`.

! Links that carry data out now prompt for confirmation by default; controlled by `http.urlDataGuard`.

WAS THIS USEFUL?  

cotal

Feedback ⓘ v0.181.0 NOTES ↗ RANK

Cotal is the open standard for agent coordination

> Cotal shipped a full `Mesh` TypeScript interface for seat lifecycle, introspection and manager discovery, alongside major ACP protocol advances — streamed agent events, live model switching without despawn, thread titling, and harness command discovery — plus spawn-time permission gating, reasoning-event streaming and auto-naming that bumped the Manager cluster to revision 56, a wave of new SDK primitives (`SpawnRequest` , `SeatFrame` , `FrameMapper` , `HiddenSessions` , prompt-desk types), and CLI polish across `cotal send` , `cotal models` , `cotal setup` , and backup/deploy commands.

└──▷ **WHAT SHIPPED · 45 FEATURES** most completely described first ? what's the number?

01 Mesh interface for seat lifecycle

Not checked yet NEW

95

The `Mesh` interface provides seat lifecycle methods — spawn(request: SpawnRequest) to spawn a seat and wait for manager confirmation (returns `SeatRef`), input(seat, text) to send text as the operator, interrupt(seat) to interrupt a live turn via the manager's connector interrupt key, rename(seat, name) via the agent endpoint's `set-name` command, and spawnNameError(name) to validate a proposed name before creation; introspection methods frameHistory(seat) (oldest first), presence(seat) (`idle` , `working` , or `waiting`), and seats() returning a `SeatList` ; manager discovery methods managers() returning `ManagerChoice[]` , manager(instanceId), catalogs(managerInstanceId?), and personas(managerInstanceId?) (the latter two reject with `LaunchListsRefused` without permission); channels() and roster() listings; close() to close the connection; and `space` / `defaultAgent` properties for the current space and preferred default agent.

Spawn a new agent seat, send it a task, then tear it down when done — a full automated agent workflow.

typescript

```
const seatRef = await mesh.spawn(spawnRequest);
await mesh.input(seatRef, 'Scan the target host and report open
ports');
// ... wait for frames ...
await mesh.despawn(seatRef);
```

Validate a proposed seat name before attempting to spawn, to surface rejection reasons early in a UI or CLI flow.

```typescript
const error = mesh.spawnNameError(proposedName);
if (error) {
  console.error('Invalid seat name:', error);
} else {
  const seatRef = await mesh.spawn({ name: proposedName, ...rest
});
}
```

Monitor all seat activity in real time and cancel the subscription when your session ends.

```typescript
const unfollow = mesh.followFrames((frame) => {
  console.log('Frame from seat', frame.seat, ':', frame);
});
// Later, to stop receiving frames:
unfollow();
```

— Enumerates every method with three runnable examples product docs

02 **CotalEndpoint messaging and presence utilities** Not checked yet

NEW 90

`ensureOpenSpaceStreams` creates an open space's message streams before first use, for senders that consume nothing and may be the space's first participant. `nextPromptsSeq` takes the next prompt sequence number for the current runtime (0 on first call, then one above the highest sequence used), required before a prompt desk can track open prompts. `retryNow` triggers an immediate reconnect attempt during a reestablish-loop backoff, bypassing up to a minute of wait after broker return when sends are queued. `setPromptsPresence` sets the prompts presence field and writes presence in one call, supplying a `PresencePrompts` value with `open` / `sent` prompt counts plus a state-change counter; it throws when the agent endpoint is not currently served. `setCardName` changes a card's display name and republishes presence atomically, leaving the principal unchanged (no-op if identical, throws for names `assertValidName` refuses), and `setCardMeta` sets or removes card meta keys in one presence record, with `setCardName` now also accepting meta changes for the record that carries the new name. Passing `expectedLastMsgId` on publish calls turns a publish into a probe — a cached ID returns the stored sequence, any other ID is refused with error `10070` before storage — and `legacyDiscoveryMessage` supplies a fallback message when membership discovery predates lifecycle support.

Force an immediate reconnect attempt when your agent has queued sends and can't wait out the full reestablish-loop backoff after a broker outage.

```typescript
endpoint.retryNow();
```

Ensure an open space's streams exist before a sender-only participant tries to publish into the space.

```typescript
await endpoint.ensureOpenSpaceStreams();
```

— Names every method and error code with two runnable examples v0.180.0

03 **Per-harness model listing improvements** Not checked yet IMPROVED

90

Launch pi agents at a chosen thinking level with `--variant <level>`, which maps to pi's `--thinking` flag; `cotal models --harness pi` lists each model's thinking levels as variants (requires pi 0.81.0 or later). `cotal models --harness jcode` lists the models available to a Jcode install, including on a default install with no `config.toml`, by running `jcode model list --json` instead of reading `model_catalog` or `reasoning_efforts` from `config.toml`. On OpenCode 2.x, `cotal models --harness opencode` now lists every model instead of an empty list, checking a pinned model against the server's complete list by starting a private `opencode serve --stdio` and reading models only after `GET /api/plugin` confirms a plugin has loaded.

Inspect all available Claude models and their effort-level variants, including alias resolution, before choosing one for a session.

```
$ cotal models --harness claude
```

List the models your Jcode install actually supports on a fresh setup before launching an agent.

```
$ cotal models --harness jcode
```

— Exact flags, endpoint, and version requirement per harness v0.178.0

04 **Harness command discovery and injection** Not checked yet NEW

90

The Manager's new `harness-commands` read (`{ agent, cwd }`) lists the commands and skills a harness offers before any agent starts, and the `spawn` field

`firstCommand: { target, name, args? }` starts an agent with one of them as its first turn; acceptance names `firstCommand.id` and `command-state` reports any refusal. The Manager cluster advances to revision 54.

Discover which commands a harness exposes in a directory before launching an agent, so you can pick a first turn without starting a seat.

```
$ { "read": "harness-commands", "agent": "<agent-id>", "cwd":
"/path/to/project" }
```

Start an agent and immediately queue a specific harness command as its first turn, so the agent acts before any other instruction arrives.

```
$ { "spawn": { ..., "firstCommand": { "target": "<target>",
"name": "<command-name>", "args": ["--flag"] } } }
```

— Exact read shape and spawn field with two runnable examples v0.177.0

---

05 **SpawnRequest interface for agent spawning** Not checked yet NEW

85

The `SpawnRequest` interface configures an agent seat at launch: required fields `agent`, `cwd`, and `prompt`; optional `subscribe` (readonly string array of channels to join), `model` to override the agent's model, `persona` to assign a persona, `managerInstanceId` to link to a managing instance, and `variant` to select an agent variant.

Spawn an agent with a specific model and subscribe it to notification channels from the moment it starts.

typescript

```typescript
const req: SpawnRequest = {
  agent: "my-agent",
  cwd: "/workspace/project",
  prompt: "Analyze the codebase for security issues.",
  model: "gpt-4o",
  persona: "security-analyst",
  subscribe: ["alerts", "findings"]
};
```

— Every field named with a runnable spawn example product docs

---

06 **FrameMapper class for frame ordering** Not checked yet NEW

85

The `FrameMapper` class maps ACP session frames in order, tracking message roles and open tool calls across frames within a session. closeOpen(status) closes any still-open tool calls at turn end (`completed` or `failed`), returning `SessionUpdate[]` so no call stays open once a turn ends. map(frame: SeatFrame) maps individual frames, returning `Mapped` and deduplicating already-seen frames by epoch and sequence number so a history read and a live tap can safely overlap.

Process live ACP seat frames while safely overlapping with a history read — duplicate frames (same epoch and seq) are silently dropped.

typescript

```
import { FrameMapper } from '@cotal-ai/acp';

const mapper = new FrameMapper();

for (const frame of incomingFrames) {
  const mapped = mapper.map(frame);  // frame: SeatFrame, returns
Mapped
  // handle mapped ...
}

// At turn end, close any still-open tool calls
const updates: SessionUpdate[] = mapper.closeOpen('completed');
```

— Both methods named with dedup mechanism and runnable example `product docs`

07 **Spawn permission gating and auto-naming** Not checked yet BREAKING

85

`askPermissions` and `reasoningEvents` spawn options, plus `autoName` auto-naming, are available to agent spawning via `cotal spawn` or any ACP client: `askPermissions` requires approval before tool use, `reasoningEvents` publishes the model's reasoning, and passing `autoName: true` to `spawn` auto-names a new agent from its first message, reaching the seat as `COTAL_AUTO_NAME=1`, storing the label in `meta["ai.cotal.label"]` (up to 40 code points), with presence carrying `meta["ai.cotal.naming"] = "pending"` until naming settles; every `status.connectors` row reports whether its connector accepts the `askPermissions` and `reasoningEvents` options. This bumps the Manager cluster to revision 56, so Managers and their callers must be updated together.

— Full mechanism and keys named; coordinated upgrade required `v0.180.0`

08 **List signed-in user's Manager instances** Not checked yet NEW

80

`AuthProvider.listManagerInstances` retrieves a `ManagerInstanceListing` of the signed-in person's own Manager instances from the space's auth service via `POST`

`/manager-instances`; a failed or partial scan must throw, and an empty array signals an owner with no Manager.

Enumerate all Manager instances for the currently signed-in user to verify they have at least one Manager associated with their account.

```typescript
const listing: ManagerInstanceListing = await
authProvider.listManagerInstances();
if (listing.instances.length === 0) {
  throw new Error('User has no Manager instances');
}
```

— Named endpoint, types, and a runnable example `product docs`

09 **Principal route calling and serving** `Not checked yet` NEW

80

`CotalEndpoint.callPrincipal` invokes a single command on a principal's own endpoint using an `authority` (`owner` or `grant`), a `contract` and `output` (`CompiledContract`), returning an `EpPrincipal` / `EndpointReply` and throwing `unavailable`, `permission-denied`, or `deadline-exceeded` on failure. `servePrincipal` serves this process's agent endpoint on the principal route, registering the `AGENT_CLUSTER` command set with handlers typed as `EpPrincipalCommandDef` and returning a `PrincipalServe` handle, rejecting on open mesh, static credential, missing lifecycle UID, or non-cluster handler sets. Agents on static spaces and open meshes can now answer owner prompts, commands, turns, interrupts, usage reads and renames: `openAgentCaller` in `@cotal-ai/workspace` opens a new static `agent-caller` profile, and `servePrincipal` / `MeshAgent` now serve on every mesh mode.

— Full mechanism and error cases named, no example `v0.180.0`

10 **Idle-exit auto-stop for managers** `Not checked yet` NEW

80

Start an auto-stopping manager with `cotal manager start --idle-exit <seconds>` or `cotal service start manager --idle-exit <seconds>`; the manager deregisters once it has held no agent, call, workflow run, preservation or resume for that many seconds.

Start a manager that shuts itself down and deregisters after 15 minutes with no active agents, useful for temporary CI environments.

```
$ cotal manager start --idle-exit 900
```

— Exact flags with a runnable, concrete example v0.180.0

11 **Automatic T3 Code and Zed setup** Not checked yet NEW 80

`cotal setup` auto-registers Cotal in T3 Code and Zed, installing `@cotal-ai/acp` and adding one `cotal acp --space <space>` entry per space without manual config edits; `cotal update` points T3 Code and Zed ACP entries at the latest installed Cotal binary; `cotal status` checks T3 Code and Zed ACP entries whose command path is gone.

Register Cotal automatically in T3 Code and Zed so agent threads can start without manual config edits.

```
$ cotal setup
```

— Three named commands with a runnable setup example v0.176.0

12 **MeshAgent command dispatch and state** Not checked yet NEW 75

MeshAgent.dispatchNextCommand() hands the oldest waiting command to the connector's dispatcher, callable only while no turn runs, returning a `CommandDispatch` and denying instance-route or manager-call exchanges the caller cannot obtain. The new `queuedCommands` property inspects commands waiting for dispatch. The new `prompts` (`PromptDesk`) property accesses the prompt desk of a process's endpoint, available once the broker accepts the serve, undefined on an open or static mesh or when the endpoint is refused, and refusing with `closed` once the endpoint closes. The new `displayName` property reads the display name visible to peers, reflecting the launch name until `set-name` changes it.

Poll for and dispatch the next queued command from a MeshAgent while no turn is running, then inspect remaining queued commands.

```typescript
const dispatch = await agent.dispatchNextCommand();
console.log('Remaining queued:', agent.queuedCommands);
```

— Four named members with runnable example and edge cases product docs

13 **Connector-core prompt desk primitives** Not checked yet NEW 75

New `AgentEndpoint` and `openAgentEndpoint` primitives open and manage an agent-facing endpoint. `CommandDispatcher` and `CommandQueue` types, plus `COMMAND_LIST_MS` , `ENDED_COMMANDS_KEPT` , and `MAX_QUEUED_COMMANDS` constants, enable structured command dispatch and queue-depth control.

`PromptDesk`, `PromptHold`, `PromptHooks`, `PromptSpec`, and `PromptVerdict` types, with `HookContext` and `HookRelayOptions`, support a full prompt-review workflow, timed by `PROMPT_ANSWER_REPLY_MS`, `PROMPT_OPEN_MS`, `PROMPT_SETTLE_MS`, `PROMPT_SUMMARY_MAX_BYTES`, and `PROMPT_DESK_MIN_MAX_PAYLOAD` constants. `CommandRefused` and `PromptRefused` types surface explicit refusal states, and `AgentCommand`, `CommandDispatch`, `Delivery`, `DeliveryNotice`, and `Evidence` types complete the command-and-delivery data model. Utility functions `isConfigPermissionError`, `isConfigProtected`, `nameEnv`, and `connectionNotReadyWarning` support configuration validation and environment naming, alongside `FROM_MIN_CODE_POINTS` and `MESH_ACTIONS` constants for message-origin filtering and mesh action enumeration.

— Every type and constant named, no runnable example `product docs`

---

14 **Launch dry-run validation mode** `Not checked yet` NEW 70

`LaunchOpts.validateOnly` lets you validate launch options without creating launch files or changing session state, available when a connector declares `Connector.supportsLaunchValidation`; the returned spec is not executable.

Dry-run a launch configuration to catch option errors before any files are written or session state changes — useful in CI pre-flight checks.

```typescript
const result = await launch({ ...opts, validateOnly: true });
// result.spec is returned but not executable
```

— Clear mechanism and runnable example, no limits given `product docs`

---

15 **ACP session slash commands** `Not checked yet` NEW 70

The `name` slash command renames a running agent session. The `open` slash command takes control of an already-running agent from the current session, accepting an agent name, and now also updates the client's model list via a `config_option_update` when it switches the thread's model. The `roster` slash command lists every agent on the mesh and its current activity, `channels` lists all channels in a space, and a `despawn` command was also added. The ACP bridge supports `/name` to name or rename an agent in a thread and `/open` to drive an already-running agent from a new thread.

— Names all five commands and behaviour, no sample invocation `v0.181.0` `v0.176.0`

---

16 **Channel reply support in cotal send** `Not checked yet` NEW 70

Reply to a channel post with `cotal send -c <channel> --reply-to <message-id>`, storing the post with `replyTo` as a direct send already does.

Reply to a specific post in a channel by message ID, now that channel sends support `--reply-to`.

```
$ cotal send -c general --reply-to msg-abc123 'Acknowledged, investigating now'
```

— Exact flag with a runnable example v0.181.0

17 **ACP thread model switching and titling** Not checked yet IMPROVED 70

Choosing another model or reasoning level on a running ACP thread now switches its agent from the next turn with no `/despawn` required; the `agent` endpoint advances to revision 6 with the owner-only `set-model` command. ACP threads are now titled with the name their agent shows; in T3 Code a title you give the thread becomes the agent's label via the new owner-only `set-label` command (revision 7 of the `agent` endpoint, max 40 characters). ACP threads also start agents on a model named `default` when the application lists one.

— Named endpoint revisions and commands, no sample call v0.181.0

18 **Authenticated control server for agents** Not checked yet NEW 65

`startControlServer` runs an authenticated control server with a per-connection control exchange — clients that cannot prove they hold the key are dropped before any handler runs, preventing local processes from driving presence, injecting peer messages, or shutting the agent down. On non-Windows systems it enforces socket directory privacy before binding, treating a non-private directory as a bind error.

— Mechanism described but no usage example product docs

19 **Connector interrupt key and prompts** Not checked yet NEW 65

Declaring `interruptKey` on a Connector specifies the byte sequence the manager's `interrupt` command writes into a seat's terminal to end a live turn (e.g. the Escape byte `\x1b`) — without it, `interrupt` refuses seats on that connector. Declaring `prompts` on a Connector makes sessions it launches serve the harness's tool permissions on the agent's `agent` endpoint (SPEC 13.22) when the mesh is user-auth.

— Two connector fields with concrete mechanism, no example product docs

20 **RunSignal run lifecycle events** Not checked yet NEW 65

The `RunSignal` type in `@cotal-ai/acp` represents run lifecycle events — a `start` signal carrying `runId` and `ts`, and an `end` signal carrying `runId`, `ts`, `outcome`

(`finished` , `error` , or `interrupt`), and an optional `message` — used to notify the bridge when a run completes for prompt completion.

— Fully specified type fields, no usage example `product docs`

21 **ACP client event streaming** `Not checked yet` NEW 65

ACP clients such as T3 Code now receive streamed replies, tool calls with arguments, results and diffs, plans, approvals, questions, agent commands, cancel, image prompts, context usage, reasoning, and model and effort pickers from every Cotal agent; the `agent` endpoint advances to revision 5, adding `turn` and `usage` fields, and presence gains `agentEndpoint: { interrupt, turn, images, usage }` .

— Endpoint revision and fields named, no example `v0.180.0`

22 **ACP bridge bootstrap without install** `Not checked yet` NEW 65

Bootstrap the Cotal ACP bridge without a local install using `npx @cotal-ai/acp` , which offers a terminal sign-in if Cotal or the ACP extension is missing.

Bootstrap the ACP bridge on a machine where Cotal is not yet installed, triggering a terminal sign-in that installs any missing components.

```
$ npx @cotal-ai/acp
```

— Exact command with runnable example `v0.176.0`

23 **Seat display naming at launch** `Not checked yet` NEW 60

`LaunchOpts.displayName` sets a human-readable seat name at launch time, forwarded as the `COTAL_DISPLAY_NAME` environment variable when a seat was assigned a name via `set-name` (SPEC 13.22).

— Names the field and env var, no example given `product docs`

24 **Custom connect approval instructions** `Not checked yet` NEW 60

Setting `connectDescription` in `GatewayOptions` overrides the default approval instructions shown in the `cotal_connect` tool.

Provide custom approval instructions to users when they are prompted by the `cotal_connect` tool at gateway startup.

```typescript
const gateway = new Gateway({
  connectDescription: 'Contact your admin at security@example.com
to approve this connection.',
  // ...other options
});
```

— Single config field with runnable example `product docs`

---

25 **SeatFrame AG-UI event interface** `Not checked yet` NEW 60

The `SeatFrame` interface represents one AG-UI frame from a seat's event channel, exposing fields `actor`, `epoch`, `events` (typed `readonly AguiEventLike[]`), `owner`, `runId`, `seq`, `threadId`, and `ts` (epoch milliseconds).

— All fields named, no usage example `product docs`

---

26 **HiddenSessions interface for hiding runs** `Not checked yet` NEW 60

The `HiddenSessions` interface hides a session run without stopping it, exposing add(lifecycleUid: string) to mark a run as hidden and read() to return the current set of hidden lifecycle UIDs as a `ReadonlySet<string>`.

— Both methods named, no usage example `product docs`

---

27 **Space quiescence observation in auth SDK** `Not checked yet` NEW 60

`observeSpaceQuiescence` in `@cotal-ai/auth` observes when a space becomes quiescent, configured via `SpaceQuiescenceOptions` (including `knownMachineAccounts` to filter known service accounts) and inspected via the `SpaceQuiescenceObservation` return type.

— Function and types named, no usage example `product docs`

---

28 **Per-kind effect budget ceilings** `Not checked yet` NEW 60

Workflow runs can pin separate effect budget ceilings for spawns, turns, and other effect kinds, enforced across restarts; a per-kind refusal names its limit with error `L4029`.

— Names error code but no config surface shown `v0.181.0`

---

29 **Hermes seat initial prompt support** `Not checked yet` NEW 60

Hermes seats now publish an AG-UI event plane and accept an initial prompt via `cotal spawn --prompt`, a manifest `prompt` field, or an ACP client's first message.

— Names flag and manifest field, no runnable example v0.180.0

THINNER COVERAGE BELOW

30 **AgentConfig initial display name** Not checked yet NEW

55

The `displayName` field on `AgentConfig` sets a card's initial display name independently of the launch name.

Override a card's visible label at startup without affecting the persistent launch name stored in `COTAL_DISPLAY_NAME`.

```typescript
const agent: AgentConfig = {
  name: 'my-agent',
  displayName: 'My Agent (Prod)'
};
```

— Single field named with example, thin mechanism product docs

31 **Claude Code session PID binding** Not checked yet NEW 55

Setting `claudePid` in `ClaudeSessionOptions` binds an MCP session to a specific Claude Code process's command socket; leave it unset when one host serves multiple sessions.

— Field named and purpose explained, no example product docs

32 **Endpoint check directory guidance** Not checked yet IMPROVED

55

`cotal endpoint check` now names a directory passed by mistake and shows how to pass its `contract.json` explicitly when it is a regular file.

Pass a contract file path directly to endpoint check when you accidentally point it at a directory.

```
$ cotal endpoint check ./my-endpoint/contract.json
```

— Names command and file, has runnable example v0.176.5

33 **Self-update reinstalls missing extensions** Not checked yet

IMPROVED 55

`cotal update --self` now reinstalls first-party extensions with missing entry files, even if the CLI is already current.

Re-run a self-update to restore missing extension entry files without waiting for a newer CLI version to be released.

```
$ cotal update --self
```

— Exact command with runnable example v0.176.5

34 **CLI error message clarity fixes** Not checked yet IMPROVED

50

`cotal channel set` now points to `--desc` when you type `--description` and no longer suggests passing unknown options as extra channel names. `cotal down` now names the path when a space's recorded project root is no longer a folder, instead of telling you to run it from inside a regular file.

— Two named commands and flags, no example invocation v0.181.0

35 **Backup flag conflict validation** Not checked yet IMPROVED

50

`backup recover` and `backup cleanup-fallback` now reject combining `--dry-run` and `--yes` before looking up the attempt.

— Names flags and commands, no example v0.178.0

36 **Claude Code exposes own commands** Not checked yet NEW

50

Claude Code agents on a signed-in space can now list, run, and interrupt their own `/` commands for their owner, as OpenCode, pi, and Codex agents already could; the `agent` endpoint's `commands` field now lists Claude Code's own `/` commands as `command` entries.

— Names endpoint field, no example v0.176.0

37 **Launch-aware model catalog env allow** Not checked yet NEW

45

`ModelCatalogOpts.envAllow` gives a model catalog read the `spawn.env` names a launch receives, so model lists respect the launch environment.

— Field named, mechanism thin v0.180.0

38 **COTAL_HOME file edge-case handling** Not checked yet IMPROVED 45

When `COTAL_HOME` points to a file, `watch`, `history --follow`, and `channel members --follow` now write an error record and an end record in JSON mode.

— Names three commands affected, no example v0.177.0

39 **Deploy preview broker.host warning** Not checked yet IMPROVED 45

The team deploy preview now warns when a manifest declares `broker.host`, which apply refuses, and lists no channel creation candidates for that manifest.

— Names config key, no command shown v0.176.5

40 **ACP manager service reuse** Not checked yet IMPROVED 40

`cotal acp` now recognises a manager service started by an older release, so new threads use it instead of failing to start a second one.

— Behaviour described, no command or config named v0.181.0

41 **Completion uninstall preserves symlinks** Not checked yet IMPROVED 40

`cotal completion uninstall` now leaves symlinked profiles alone when no stub or ownership record exists.

— Narrow fix, no example v0.176.0

42 **Programmatic stack manager remedy start** Not checked yet NEW 35

`startStackManagerRemedy` in the `@cotal-ai/workspace` control module starts a stack manager remedy programmatically.

— Thin one-line mention, no mechanism product docs

43 **Terminal session API in workspace** Not checked yet NEW 35

`openTerminalSession` opens a live terminal session from the workspace module, returning a `TerminalSession` object.

— Function and return type named only product docs

44 **Foreground and dry-run flag compatibility** Not checked yet IMPROVED 35

`--foreground` and `--dry-run` can now be used together on service and manager commands, as help text now documents.

— Flags named, no example or command context v0.176.0

45 **Manifest teardown retry by run ID** Not checked yet IMPROVED

30

Retry a partial manifest teardown by run ID without passing the manifest path.

— Behaviour stated, no command shown v0.176.5

└──▷ **ALSO FROM THESE RELEASES**

Preview which history entries would be cleared for a stopped space, without restarting it first.

```
$ cotal history clear --dry-run
```

└──▷ **BREAKING ON UPGRADE**

! Manager cluster advances to revision 56: `spawn` gains `askPermissions` and `reasoningEvents`; Managers and their callers must be updated together.

WAS THIS USEFUL? 👍 👎

CopilotKit OpenBot

Feedback ⓘ v1.2.0 NOTES ↗ RANK

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

> OpenBot v1.2.0 reworks several default behaviors around memory, analytics, organizations and app access — all reversible via config keys or env vars — while adding personal AI connections, database sharding, team model allowlists, and new audit trails for restores, runs and refused turns.

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

01 **Memory sources no longer auto-created from connected apps**

Not checked yet BREAKING 95

OpenBot no longer auto-creates memory sources for connected apps; the upgrade deletes all unreviewed imports from those sources and switches them off, with migration `0182_withdraw_default_source_imports` running at startup (unskippable) to withdraw imports made by auto-created memory sources across every organization. Set `config.memory.sourceDefaults: true` or `OPENBOT_MEMORY_SOURCE_DEFAULTS=true` to restore the old behaviour, under which app records reach memory only from sources a person explicitly adds on a Bot's Memory page.

Restore the pre-1.2.0 behaviour where connecting an app automatically makes it a memory source, without changing the migration outcome.

yaml

```yaml
config:
  memory:
    sourceDefaults: true
```

— Exact config key, env var and migration name given v1.2.0

02 **App access defaults overhaul** Not checked yet BREAKING 80

At each organization's first start of this release, apps are opened to all Bot owners and prior per-Bot or per-app access limits are not preserved; use Admin > Apps > Open apps to lock apps back down. Removing a member now deletes all logins they saved in that organization (the first start of this release also deletes saved logins for anyone already removed, and restoring a member does not restore their logins). New migration

`0174_app_access_request_actions` records which actions a Bot-access request names, with earlier requests now reading as naming none.

— UI path and migration named, concrete before/after behavior v1.2.0

---

03 **Per-user organizations no longer default on bundle chart**

Not checked yet

BREAKING

78

Self-hosted installs on the bundle chart now put all users into the shared `default` organization instead of giving each user their own; set `config.organizations.ownOrganizations: true` or `OPENBOT_OWN_ORGANIZATIONS=true` to restore the previous per-user organization behaviour.

Give every user who signs in their own organization rather than landing them all in `default`, preserving pre-1.2.0 self-hosted multi-org behaviour.

yaml

```yaml
config:
  organizations:
    ownOrganizations: true
```

— Exact config key and env var with before/after stated v1.2.0

---

04 **Personal AI connections** Not checked yet NEW

78

Each person can now connect their own AI, stored via migration `0153_personal_ai_connections`, sealed with the vault key and never shared with an organization; it follows the person across organizations and takes precedence over the organization's AI on their runs. Migration `0179_additional_ai_connections` further stores an organization's and a person's additional AI connections beyond the default, one per provider.

— Clear mechanism and precedence, but no config/command shown v1.2.0

---

05 **Database sharding for organizations** Not checked yet NEW

75

Organizations can be spread across multiple databases via `database.shards`, tracked by migration `0167_shard_directory`; an organization with no shard row defaults to the first database.

— Config key and migration named, no runnable example v1.2.0

---

06 **Product analytics on by default** Not checked yet

BREAKING

65

Product analytics (daily count rollups sent to CopilotKit management) are now on by default; disable with `config.productAnalytics: false` or, on the bundle chart, `openbot.config.productAnalytics: false`.

Disable product analytics on a self-hosted bundle-chart install to prevent daily count rollups from being sent to CopilotKit management.

```yaml
openbot:
  config:
    productAnalytics: false
```

— Config keys given but mechanism is thin v1.2.0

---

07 **Team model allowlists** Not checked yet

BREAKING

65

New migration `0165_team_models` stores the model allowlist for each team, enabling per-team restrictions on which models members may run. An empty allowlist no longer blocks all runs — organizations that relied on an empty Only allowed models list to prevent all runs must now explicitly configure allowed models.

— Behavior change explained, but no exact config surface shown v1.2.0

---

08 **Audit trail for restores, runs and refused turns** Not checked yet

NEW

60

New migration `0162_computer_restores` records each restore of a computer from a saved copy, including who requested it, which copy was used, and how it ended. New migration `0152_bot_run_client_and_outcome` tracks, per Bot run, which client initiated it (web, Desktop, API, or other) and how the run ended. New migration `0166_refused_turns` stores turns refused before they ran so the conversation history still shows them.

— Three migrations named but no interaction surface given v1.2.0

BREAKING ON UPGRADE

! OpenBot no longer auto-creates memory sources from connected apps; the upgrade deletes all unreviewed imports from those sources and switches them off. Set `config.memory.sourceDefaults: true` or `OPENBOT_MEMORY_SOURCE_DEFAULTS=true` to restore the old behaviour for new connections.

! Product analytics are now on by default; set `config.productAnalytics: false` or `openbot.config.productAnalytics: false` to disable.

! An empty model allowlist no longer blocks all runs — organizations that relied on Only allowed models on, nothing allowed to prevent all runs must now explicitly configure allowed models.

! Self-hosted installs on the bundle chart now put all users into the shared `default` organization instead of giving each user their own; set `config.organizations.ownOrganizations: true` or `OPENBOT_OWN_ORGANIZATIONS=true` to restore the previous behaviour.

! Apps are opened to all Bot owners at each organization's first start of this release; prior per-Bot or per-app access limits are not preserved. Use Admin > Apps > Open apps to lock apps down again.

! Removing a member now deletes all logins they saved in that organization; the first start of this release also deletes saved logins for anyone already removed. Restoring a member does not restore their logins.

! Migration `0182_withdraw_default_source_imports` runs at startup and cannot be skipped; it withdraws imports made by auto-created memory sources across every organization.

WAS THIS USEFUL? 👍 👎

AWS Strands Agents

Feedback ⓘ harness-cli/v0.5.0 NOTES ↗

RANK

Strands Agents is AWS's open-source SDK for production agents in Python and TypeScript, with a model-driven loop, tool integration, and deployment targets.

> Strands Agents shipped a new `strands dev` CLI workflow for isolated agent development and auto-optimization, an `HttpDecisionModel` integration for System One HTTP servers, per-model cost streaming, and several breaking changes around background-task defaults, context-manager naming, and subagent tooling.

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

01 **`strands dev` command for isolated workspaces and auto-optimization**

Not checked yet NEW 85

The `strands` CLI gains a `strands dev` command that provides isolated dev workspaces, experiments, evaluation, and optimization. It accepts `--max-rollouts` and other optimize-run flags to control auto-optimization runs, e.g. `strands dev --max-rollouts 10`.

Run auto-optimization in an isolated dev workspace, capping the number of rollouts to control cost.

```
$ strands dev --max-rollouts 10
```

— Runnable command shown with concrete flag harness-cli/v0.5.0

02 **Background-task management via agent.cancel()** Not checked yet

85

BREAKING

Background-task management is now available as an opt-in via `cancel_background_tasks` or `cancelBackgroundTasks` on agent.cancel(), e.g. await agent.cancel(cancel_background_tasks=True). The harness now defaults to `manage_background_task` for background-task lifecycle management, so harnesses that depended on the previous default behaviour will behave differently on upgrade.

Let an agent cancel background tasks it launched when the user cancels the main run.

```python
await agent.cancel(cancel_background_tasks=True)
```

— Runnable code plus explicit before/after of default `harness-cli/v0.5.0`

THINNER COVERAGE BELOW

03 **`stash` renamed to `archive` on context manager** Not checked yet

55

BREAKING

The `stash` method on `context-manager` / `ContextManager` is renamed to `archive` in both Python and TypeScript; any code calling .stash() will break.

— Names exact rename, no migration helper shown `python/v1.67.0` `harness-cli/v0.5.0`

04 **HttpDecisionModel for System One HTTP servers** Not checked yet

NEW

50

Adds `HttpDecisionModel` to connect any System One HTTP server as a decision model.

— Names the integration but no usage shown `harness-cli/v0.5.0`

05 **Context summarization uses agent's own model** Not checked yet

IMPROVED

50

Context summarization now uses the agent's own model instead of `aux_model` (Python).

— Clear before/after but no example or config shown `python/v1.67.0`

06 **Per-model cost streaming and usage breakdowns** Not checked yet

NEW

45

A new cost plugin (Python) streams per-model cost estimates and exposes per-model usage breakdowns.

— Described in prose, no command or config key `harness-cli/v0.5.0`

07 **Harness switches to SDK's vended subagent tool** Not checked yet

45

BREAKING

The harness now uses the SDK's vended subagent tool instead of its own; custom harnesses that depended on the previous subagent tool wiring will need to update.

— States impact but not the new tool's interface `python/v1.67.0` `harness-cli/v0.5.0`

08 **Token usage on model spans via telemetry** Not checked yet NEW

40

Token usage on model spans can now be opted into via telemetry, available in both Python and TypeScript.

— Named surface but no mechanism or example `harness-cli/v0.5.0`

└──▷ **BREAKING ON UPGRADE**

! The harness now uses `manage_background_task` by default — harnesses that depended on the previous default background-task behaviour will behave differently on upgrade.

! The `stash` method on `context-manager` is renamed to `archive` in both Python and TypeScript — any code calling .stash() will break.

! The harness now uses the SDK's vended subagent tool instead of its own — custom harnesses that depended on the previous subagent tool wiring will need to update.

! The `stash` method on `ContextManager` is renamed to `archive` in both Python and TypeScript.

! The harness now uses the SDK's vended subagent tool (`refactor(harness)!: use the SDK's vended subagent tool`).

WAS THIS USEFUL? 👍 👎

LangChain

Feedback ⓘ `langchain-anthropic==1.7.13` NOTES ↗ RANK

LangChain is an open-source framework that orchestrates applications powered by language models.

> LangChain's Anthropic integration now supports passing custom HTTP or SDK clients for the chat model.

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

01 **Custom client support for Anthropic chat model** Not checked yet

IMPROVED

50

The Anthropic chat model integration now accepts a custom HTTP or SDK client, allowing control over transport-level behaviour such as proxies, timeouts, or auth.

— Describes capability but no code example or exact parameter name

`langchain-anthropic==1.7.13`

WAS THIS USEFUL? 👍 👎

LangChain LangGraph

Feedback ⓘ cli==0.4.44 NOTES ↗

RANK

LangGraph is a low-level orchestration framework for building, managing, and deploying long-running, stateful AI agents.

> LangGraph CLI's deploy tooling gained machine-readable deployment listings and simplified BYOC authentication.

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

01 Deployment URLs in `deploy list --json`

Not checked yet IMPROVED

65

Running `langgraph deploy list --json` now includes deployment URLs in the machine-readable output, making it usable for scripting or CI pipelines.

List all deployments and get their URLs in machine-readable JSON for scripting or CI pipelines.

```
$ langgraph deploy list --json
```

— Names exact command and output change, runnable today cli==0.4.44

THINNER COVERAGE BELOW

02 BYOC deploy with API key only

Not checked yet IMPROVED

55

`langgraph deploy` can now deploy to BYOC data planes using only an API key, removing the need for additional credentials.

— Names command and benefit but no credential mechanism detail cli==0.4.44

WAS THIS USEFUL? 👍 👎

Cloudflare Agents

Feedback ⓘ agents@0.34.1 NOTES ↗ RANK

The Cloudflare Agents SDK builds stateful agents on Workers and Durable Objects, with persistent state, scheduling, WebSockets, and MCP server support.

> Cloudflare Agents added Pi 1.6.1 support, which introduces breaking changes to the tool parking and signal state shapes in `agents/models/pi-ai`, alongside improved stream error handling and faster SSE decoding.

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

01 Pi 1.6.1 tool parking and signal state shape changes

Not checked yet 93

BREAKING

Parking tools in `agents/models/pi-ai` must now return `{ kind: "sleep", signals, until }` instead of the old `{ sleep: { signals, until } }` shape — the old shape still compiles but yields an empty result at runtime under Pi 1.6.1. Signal state also changes from a flat object to a discriminated union: `{ kind: "value", value }` or `{ kind: "expired" }`.

Update a Pi parking tool to use the new 1.6.1 shape so it no longer returns an empty result after upgrading.

typescript

```
// Before (Pi < 1.6.1 — still compiles but yields empty result)
return { sleep: { signals: [signalId], until: deadline } };

// After (Pi 1.6.1)
return { kind: "sleep", signals: [signalId], until: deadline };
```

Read a signal's state using the new discriminated-union shape in `agents/models/pi-ai`.

typescript

```
// Before
if (signal.state === "expired") { ... }

// After (Pi 1.6.1)
if (signal.state.kind === "expired") { ... }
if (signal.state.kind === "value") { const v =
signal.state.value; }
```

— Names exact old/new shapes with before/after code examples agents@0.34.1

THINNER COVERAGE BELOW

02 **Retryable Pi stream errors and full onResponse coverage**

Not checked yet · IMPROVED · 59

The `agents/models/pi-ai` chat-completions stream now treats `finish_reason: "error"` as a retryable error under the Pi harness retry settings, and the `onResponse` callback now fires for every response, including failures.

— Names the exact field and callback but no usage example `agents@0.34.1`

---

03 **Linear-time SSE decoding for large streams** Not checked yet

IMPROVED · 32

The `agents/models` SSE decoder now reads large streamed events in linear time.

— States the change but no mechanism or benchmark detail `agents@0.34.1`

└──▷ **BREAKING ON UPGRADE**

! Parking tools in `agents/models/pi-ai` must be updated to return `{ kind: "sleep", signals, until }` instead of `{ sleep: { ... } }`; the old shape still compiles but produces an empty result at runtime.

! Signal state shape in `agents/models/pi-ai` changes to `{ kind: "value", value }` or `{ kind: "expired" }` — code reading the old flat signal shape will break silently.

WAS THIS USEFUL? 👍 👎

PydanticAI

Feedback ⓘ v2.75.0 NOTES ↗ RANK

PydanticAI is a framework that builds type-safe Python agents with dependency injection, model integrations, tools, and structured outputs.

> PydanticAI shipped Pax, a standalone coding-agent terminal UI built on the PydanticAI Harness, alongside major computer-use agent support across eleven sandbox runtimes with native tool routing for OpenAI, Anthropic, and Google, plus a long list of breaking changes touching fallback handling, workspace file authority, Coder instructions, Gemini usage reporting, and Playwright's default egress policy.

└──▷ **WHAT SHIPPED · 18 FEATURES** most completely described first ? what's the number?

01 Pax coding-agent terminal UI

Not checked yet NEW 88

Pax is a coding-agent terminal UI built on the PydanticAI Harness, installed with `uv tool install pydantic-ai-pax` and launched with `pax`; point it at a custom agent with `pax --agent mymodule:agent`, check the installed version with `pax --version`, and the transcript now shows user prompts on a grey background under a `/theme` palette while a new `/trust` command (with a status-row indicator) asks whether to trust a new folder.

Install Pax and start the coding-agent TUI immediately in your terminal.

```
$ uv tool install pydantic-ai-pax
pax
```

Run Pax against your own PydanticAI agent instead of the built-in default.

```
$ pax --agent mymodule:agent
```

Check the installed Pax version at the terminal before filing a bug report or upgrading.

```
$ pax --version
```

— Names every flag and command, with runnable install and version commands. v2.75.0

02 Automatic fallback on malformed model responses

Not checked yet 88

BREAKING

New `ModelEmptyResponseError` and `ModelInvalidResponseError` make `FallbackModel` and `ModelFallback` automatically fall back on an empty stream, an OpenAI Chat completion without `choices`, or a response body that is not JSON or doesn't match the schema, instead of raising `UnexpectedModelBehavior`. An `except ModelAPIError` placed before `except UnexpectedModelBehavior` now also catches these new errors; pass a `fallback_on` such as lambda e: isinstance(e, ModelAPIError) and not isinstance(e, UnexpectedModelBehavior) to keep raising instead of falling back.

— Exact exception types and fallback_on lambda given. v2.75.0

---

03 **Coder SubAgents delegation instruction gated by profile**

Not checked yet 88

BREAKING

Coder with `SubAgents` no longer sends the no-unrequested-sub-agents line to models whose profile does not set `delegates_unprompted`, sending a `UserWarning` instead behind a `FallbackModel`. On GPT models the line moves into a rewrite-mode block (`capability:coder_instructions:delegation`), so the first request after upgrading misses the prompt cache; set `profile={'delegates_unprompted': True}` or add the rule to Coder(instructions=...) to keep the line on another model.

— Exact profile flag and cache-miss note given. v2.75.0

---

04 **Computer-use agents across eleven sandbox runtimes**

Not checked yet NEW 85

`ComputerTool` from `pydantic_ai.computer_use` and the ComputerUse() harness capability, backed by the Display protocol, enable computer-use agents; DockerContainer(desktop=True), DaytonaSandbox(desktop=True), E2BSandbox(desktop=True), SpritesSandbox(desktop=True), BoxLiteSandbox(desktop=True), BoxdSandbox(desktop=True), SmolSandbox(desktop=True), RenderSandbox(desktop=True), ModalSandbox(desktop=True), KubernetesSandbox(desktop=True), and OpenSandbox(desktop=True) run them in a desktop environment. PydanticAI routes `ComputerTool` to OpenAI's native computer tool on GPT-5.4 and later, to each Claude model's computer-use tool via Anthropic, and to Google's computer-use tool on Gemini.

— Names every sandbox and model routing, no usage example. v2.74.0

---

05 **Gemini usage and thinking-token accounting changes (breaking)**

Not checked yet 83

BREAKING

A `gemini*` model accessed through `OpenAIProvider` now reports thinking tokens in `output_tokens` and `output_reasoning_tokens`, raising output token counts and cost on thinking models and tripping `UsageLimits` token limits sooner — on a server

whose `total_tokens` counts more than input, output, and thinking tokens, pass profile=OpenAIModelProfile(openai_chat_completion_tokens_omit_reasoning=False); `thinking=False` on `gemini-2.5-flash` through the Gemini API now sends `thinking_budget=128` instead of `0`, and after the 2026-12-01 reroute to `gemini-3.5-flash-lite` a budget of `0` will fail with a `400` (pass profile=GoogleModelProfile(thinking_always_enabled=False) or `google_thinking_config={'thinking_budget': 0}` to keep `thinking_budget=0` until the reroute); and streaming a `gemini*` model through `OpenAIProvider` now reports the endpoint's final usage totals instead of their sum, so cost figures drop and `UsageLimits` token limits trip later.

— Exact profile flag workaround provided. Exact values and profile override given. States the effect but offers no fix or example.

v2.75.0 v2.74.0

06 **Pax approval handler requires resume token** Not checked yet

80

BREAKING

`pydantic_ai_pax.runtime.permissions.approval_handler`'s handler now takes a resume token; in a durable run on any engine but AWS Lambda it raises `UserError` instead of asking. A custom handler run after it must also take a token; under a durable engine other than Lambda, bind ApprovalHandler(lambda ctx, requests: ask_permissions(requests, approve)) instead.

— Exact handler signature and workaround code given. v2.75.0

07 **ClearToolResults targeting and retention controls** Not checked yet

IMPROVED 80

`ClearToolResults` now accepts a `tools` argument to target specific tools, a `keep_turns` argument to retain whole user turns, and a callable `placeholder` argument, letting a noisy tool's results be dropped while keeping conversation structure intact.

Let `ClearToolResults` drop only the results of one noisy tool while keeping every user turn intact, to trim context without losing the conversation structure.

```python
from pydantic_ai_harness import ClearToolResults

capability = ClearToolResults(
    tools=['search_web'],
    keep_turns=True,
    placeholder='[search result cleared]',
)
```

— Exact arguments and a runnable example provided. v2.75.0

08 **GPT-6.2 response continuation inherits static instructions**

Not checked yet 80

BREAKING

Continuing a GPT-6.2 response through an explicit `openai_previous_response_id='resp_…'` without its `message_history` now inherits that response's static instructions; pass `message_history`, or set profile=OpenAIModelProfile(openai_responses_stable_instructions_as_input=False) to restore the previous behaviour.

— Exact flag and remedy given. v2.74.0

09 **lowest_authority_under restricts workspace files from system prompt**

Not checked yet 78

BREAKING

`lowest_authority_under` on `RepoContext`, Skills, and Memory controls which workspace files are kept out of the system-prompt instructions. Workspace files below operator authority no longer reach the system prompt: `RepoContext` names the file for the model to read with its file tools and sends an ancestor's file as framed data, Skills offers such skills through `list_skills` and `load_skill` tools instead of `load_capability`, and Memory attributes the snapshot to its `MEMORY.md`. Durable runs now record extra `realpath` workspace calls at run start; let in-flight runs finish, or use Temporal worker versioning or a new DBOS app version, before upgrading.

– Detailed before/after per module, but no direct example. v2.75.0

10 **submit_output() ends a run from a tool** Not checked yet NEW

75

Tools can call RunContext.submit_output() to submit the run's final output and end the run directly from inside a tool function, without waiting for the model to call a dedicated output tool.

Have a tool decide the run is done and hand back a typed result directly, without waiting for the model to call a final output tool.

python

```
@agent.tool
async def finish_early(ctx: RunContext[MyDeps], result: str) ->
str:
    """Submit the final answer and end the run."""
    await ctx.submit_output(result)
    return 'done'
```

– Runnable code example shows exact usage. v2.75.0

11 **Lowered-authority capability instructions sent as framed data**

Not checked yet BREAKING

62

A capability lowered with `authority` or with_authority() — including an `MCP` capability — now sends its instructions as framed data on the user channel instead of in the system prompt. Runs in flight across the upgrade that lower an `MCP` capability with `with_authority` may hit a nondeterminism error on Temporal or DBOS if another MCP server follows it; under Prefect, a flow retried across the upgrade re-runs a finished model request once.

– Describes mechanism and engine risks but no fix command. v2.75.0

12 **ModelMessagesTypeAdapter instructions equality change**

Not checked yet BREAKING

60

`ModelMessagesTypeAdapter` is now built with an `AfterValidator`; in a loaded history, requests with equal `instructions` now hold the same string object, so instructions should be compared with `==`, not `is`.

– Clear migration note on comparison operator. v2.75.0

THINNER COVERAGE BELOW

13 **AnnounceWorkingDirectory capability** Not checked yet NEW

45

The new `AnnounceWorkingDirectory` harness capability makes `Shell`, `FileSystem`, and custom tools announce the working directory once per run instead of repeating it in every tool description.

— Names capability and tools but no usage example. v2.75.0

14 **PlaywrightBrowser disables WebRTC and FedCM under egress policy**

Not checked yet BREAKING 45

Under an enforced egress policy (the default), `PlaywrightBrowser` pages no longer have `RTCPeerConnection`, `WebTransport`, `WebSocketStream`, or FedCM, and browsers it launches start with those Blink features off.

— Lists disabled features but no config to restore them. v2.74.0

15 **Temporal ExternalEventWaiter for deferred tool calls**

Not checked yet NEW 35

A new Temporal `ExternalEventWaiter` lets `HandleDeferredToolCalls` token handlers and `AskUser` wait durably inside a Temporal workflow.

— Thin description, no example or config shown. v2.75.0

16 **JSON-string tool output for Qwen3 Next** Not checked yet NEW

35

Tool text output can be sent to Qwen3 Next as a JSON string field with `needs_json_string_tool_returns`.

— Names the flag but no example or mechanism. v2.74.0

17 **Coder sends OpenAI reasoning guidance via profile flags**

Not checked yet IMPROVED 30

Coder now sends OpenAI-specific `'max'` reasoning guidance through per-model profile flags instead of static instructions.

— One-line description, no mechanism detail. v2.75.0

18 **CDPBrowser harness capability** Not checked yet NEW 25

A new CDPBrowser harness capability drives Chrome directly with raw DevTools Protocol commands.

— Bare name, no mechanism or example given. v2.74.0

└──▷ BREAKING ON UPGRADE

! Workspace files below `'operator'` authority no longer reach the system-prompt instructions in `RepoContext`, `Skills`, and `Memory`: `RepoContext` names the file for the model to read with its file tools and sends an ancestor's file as framed data; `Skills` offers such skills through `list_skills` and `load_skill` tools instead of `load_capability`; `Memory` attributes the snapshot to its `MEMORY.md`. Durable runs now record extra `realpath` workspace calls at run start — let in-flight runs finish, or use Temporal worker versioning or a new DBOS app version, before upgrading.

! `FallbackModel` and `ModelFallback` now fall back on an empty stream, an OpenAI Chat completion without `choices`, and an OpenAI Chat body that is not JSON or does not match the schema. An `except ModelAPIError` placed before `except UnexpectedModelBehavior` will now catch `ModelEmptyResponseError` and `ModelInvalidResponseError`. To keep raising instead of falling back, pass a `fallback_on` such as lambda e: isinstance(e, ModelAPIError) and not isinstance(e, UnexpectedModelBehavior).

! `pydantic_ai_pax.runtime.permissions.approval_handler`'s handler now takes a resume token; in a durable run on any engine but AWS Lambda it raises `UserError` instead of asking. To run a custom handler after it, that handler must also take a token; under a durable engine other than Lambda, bind ApprovalHandler(lambda ctx, requests: ask_permissions(requests, approve)) instead.

! `ModelMessagesTypeAdapter` is now built with an `AfterValidator`; in a loaded history, requests with equal `instructions` hold the same string object. Compare `instructions` with `==`, not `is`.

! `Coder` with `SubAgents` no longer sends the no-unrequested-sub-agents line to models whose profile does not set `delegates_unprompted`, and sends a `UserWarning` instead of the line behind a `FallbackModel`. On GPT models the line moves into a rewrite-mode block (`capability:coder_instructions:delegation`), so the first request after upgrading misses the prompt cache. To keep the line on another model, set `profile={'delegates_unprompted': True}` or add the rule to Coder(instructions=...).

! A `gemini*` model through `OpenAIProvider` now reports thinking tokens in `output_tokens` and `output_reasoning_tokens`, so output token counts and cost rise on thinking models and `UsageLimits` token limits trip sooner. Review any `UsageLimits` tuned against the old counts. On a server whose `total_tokens` counts more than input, output, and thinking tokens, pass `profile=OpenAIModelProfile(openai_chat_completion_tokens_omit_reasoning=False)`.

! `thinking=False` on `gemini-2.5-flash` through the Gemini API now sends `thinking_budget=128` instead of `0`; after the 2026-12-01 reroute to `gemini-3.5-flash-lite`, a budget of `0` fails with a `400`. To keep `thinking_budget=0` until the reroute, pass `profile=GoogleModelProfile(thinking_always_enabled=False)` or `google_thinking_config={'thinking_budget': 0}`.

! A capability lowered with `authority` or with_authority() — including an `MCP` capability — now sends its instructions as framed data on the user channel instead of in the system prompt. Runs in flight across the upgrade that lower an `MCP` capability with `with_authority` may hit a nondeterminism error on Temporal or DBOS if another MCP server follows it; under Prefect, a flow retried across the upgrade re-runs a finished model request once.

! Continuing a GPT-6.2 response through an explicit `openai_previous_response_id='resp_…'` without its `message_history` now inherits that response's static instructions; pass `message_history`, or set `profile=OpenAIModelProfile(openai_responses_stable_instructions_as_input=False)` to restore the previous behaviour.

! Streaming a `gemini*` model through `OpenAIProvider` now reports the endpoint's final usage totals instead of their sum, so cost figures drop and `UsageLimits` token limits trip later.

! Under an enforced egress policy (the default), `PlaywrightBrowser` pages no longer have `RTCPeerConnection`, `WebTransport`, `WebSocketStream`, or FedCM, and browsers it launches start with those Blink features off.

WAS THIS USEFUL? 👍 👎

Mastra

Feedback ⓘ @mastra/core@1.86.0 NOTES ↗ RANK

Mastra is a TypeScript framework for AI applications and agents, providing workflows, agent memory, RAG, tool calling, and evals.

> Mastra's 1.86.0 release adds durable dataset experiments, an experimental Knowledge importer contract, per-agent durable execution engine selection, channel-aware (and Slack-specific) chat tools for agents, and new safety hooks across Knowledge storage and memory, alongside several breaking API changes in process output, connect events, MCP boundary fields, and playground-ui traces.

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

01 **Knowledge importer contract** Not checked yet NEW 95

Knowledge gains an experimental importer contract: pass `importers: [...]` to new Knowledge() and call knowledge.importers.run(bindingId) to sync external sources with batching, leases, and replay protection. Create the importer-bookkeeping tables in in-memory, PostgreSQL, and LibSQL stores with await knowledge.importers.init(), or generate the DDL via KnowledgePG.getImportersExportDDL() or KnowledgeLibSQL.getImportersExportDDL().

Bootstrap Knowledge importers and initialize their bookkeeping tables, then trigger a sync for a specific binding.

typescript

```typescript
import { Knowledge } from '@mastra/core/knowledge';

const knowledge = new Knowledge({
  storage,
  importers: [mySourceImporter],
});

await knowledge.importers.init();
await knowledge.importers.run('my-binding-id');
```

— names every method and backing store, runnable example @mastra/core@1.86.0

02 **Channel-aware and Slack chat tools for agents** Not checked yet NEW 90

Agents gain channel-aware chat tools — `fetchMessages`, `fetchChannelMessages`, `fetchThread`, `listThreads`, `getThreadParticipants`, `getChannelInfo`, `postMessage`, `postChannelMessage`, and optionally `sendDirectMessage` — via

createChannelTools() from `@mastra/core/channels`, plus Slack-specific capabilities including search, attachment download, and user and channel lookup via createSlackTools() from `@mastra/slack`.

Equip an agent with channel-read and post tools so it can fetch and reply to messages in the active chat thread without being tied to a specific platform.

```typescript
import { createChannelTools } from '@mastra/core/channels';
import { Agent } from '@mastra/core';

const agent = new Agent({
  name: 'support-agent',
  model: openai('gpt-4o'),
  tools: {
    ...createChannelTools({ adapter: channelAdapter }),
  },
});
```

— all tool names listed, runnable setup example `@mastra/core@1.86.0`

03 **Optimistic concurrency and transaction deadlines in Knowledge storage**

Not checked yet IMPROVED 90

Knowledge.upsertPage() and Knowledge.removePage() accept `expected: { id, version }` for optimistic-concurrency protection, throwing `KnowledgeConflictError` instead of silently overwriting a concurrent edit. A `deadlineAt` can be set on knowledge.storage.transaction() to cap how long a transaction waits; on PostgreSQL (`enforcesTransactionDeadline: true`) the deadline covers connection waits, statements, and lock waits, throwing `KnowledgeTransactionDeadlineError` on expiry.

— exact fields and error types named, clear usage path `@mastra/core@1.86.0`

04 **Durable dataset experiments** Not checked yet NEW 85

Experiments can now run durably via mastra.experiments.startExperimentAsync({ experimentId, durable: true }), so a redeploy or worker crash does not leave items stuck at `running` — another worker picks up unacknowledged items and continues.

Run a dataset experiment durably so that a redeploy or worker crash does not leave items stuck at `running` — another worker picks up unacknowledged items and continues.

```typescript
await mastra.experiments.startExperimentAsync({
  experimentId: 'my-experiment',
  durable: true,
});
```

— runnable command shown, mechanism explained `@mastra/core@1.86.0`

05 **Memory hooks for working and observational memory** Not checked yet

NEW

78

`MastraMemory` implementations can now override getWorkingMemoryInstruction() to supply or suppress the working-memory instruction an agent receives (return `undefined` to keep the built-in, or an empty string to send none), and expose getObservationalMemoryConfig() and getObservationalMemoryRecords() so hosted memory implementations can surface their Observational Memory settings and records via status and config routes.

— methods named precisely, no code example given `@mastra/core@1.86.0`

06 **Per-agent durable execution engine selection** Not checked yet

70

BREAKING

Agents can select the durable execution engine individually via `durable.executionEngine`, or a default can be set for all registered agents via `Mastra.durable.executionEngine`. The earlier prerelease value `executionEngine: 'durable'` is now `executionEngine: 'default'`.

— names config keys but no code example `@mastra/core@1.86.0`

07 **get_process_output wait flag removed** Not checked yet BREAKING

60

`get_process_output` no longer supports `wait: true`; use `waitSeconds` (1–300) instead — the old flag is ignored.

— exact flag and replacement range named `@mastra/core@1.86.0`

THINNER COVERAGE BELOW

08 **knowledge_activity tool for agents** Not checked yet NEW

55

Agents can use a new `knowledge_activity` tool to query what Knowledge pages and notes changed between `since` and `until`, optionally narrowed by `from`.

— tool and params named, no usage example `@mastra/core@1.86.0`

09 **@mastra/connect events API replaced** Not checked yet BREAKING

55

In the `@mastra/connect` prerelease, `EventsClient`, connectEvents(), and related events exports are replaced by the new `ConnectEvents` iterator API.

— named old and new APIs, no migration example `@mastra/core@1.86.0`

10 **Playground-ui traces components API changed** Not checked yet

50

BREAKING

`@mastra/playground-ui` traces components have breaking API changes: `ThreadTrace` is now slot-based, `TracesExplorer` is data-in with no internal fetch, and navigation callbacks replace `href` / Link props.

— components named, no migration code shown `@mastra/core@1.86.0`

11 **Optional lookup for getWorkspaceById** Not checked yet IMPROVED

45

mastra.getWorkspaceById() now accepts `{ optional: true }` to return `undefined` for an unknown workspace ID instead of throwing an error, enabling existence checks before creation.

— exact option named, clear usage `@mastra/core@1.86.0`

12 **Snapshot pruning for workflows** Not checked yet NEW 45

Enabling `pruneSnapshot` together with the experimental `storageVersion: 2` option on a workflow trims stored step results when a run finishes.

— option names given, no code example `@mastra/core@1.86.0`

13 **Recommended flag on ask_user options** Not checked yet NEW

40

Individual `ask_user` options can now be marked with the optional `recommended` flag so a host UI can highlight the agent's suggested choice in the suspend payload.

— small named flag with no example `@mastra/core@1.86.0`

14 **Memory-failure abort reason on agent_end** Not checked yet NEW

40

Runs stopped by a memory failure are now detectable via the new `abortReason: 'memory-failure'` field on `agent_end` events.

— field named but no usage shown `@mastra/core@1.86.0`

15 **MCP boundary field renamed** Not checked yet 35

BREAKING

In `@mastra/mcp-apps-components`, the MCP boundary field `projectId` is renamed to `project`.

— single field rename, minimal context `@mastra/core@1.86.0`

└──▷ BREAKING ON UPGRADE

! `get_process_output` no longer supports `wait: true`; use `waitSeconds` (1–300) instead — the old flag is ignored.

! `@mastra/connect` PRE-RELEASE: `EventsClient`, connectEvents(), and related events exports are replaced by the new `ConnectEvents` iterator API.

! `@mastra/mcp-apps-components`: the MCP boundary field `projectId` is renamed to `project`.

! `@mastra/playground-ui` traces components have breaking API changes: `ThreadTrace` is now slot-based, `TracesExplorer` is data-in with no internal fetch, and navigation callbacks replace `href` / `Link` props.

! The earlier prerelease `executionEngine: 'durable'` value is now `executionEngine: 'default'`.

WAS THIS USEFUL? 👍 👎

EXXETA exxperts

Feedback ⓘ v0.16.1 NOTES ↗ RANK

The exxperts framework runs local-first AI agents with approval-gated memory, model-provider integrations, MCP tools, and web search.

> exxperts v0.16.1 adds a `--container` flag for running as a headless server, granular per-connector tool permissions with new status pages, and skill auto-updates, while also removing several built-in connectors and changing default behavior for unprotected Auto rooms.

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

01 Linux container mode via `--container` flag

Not checked yet NEW

75

Run exxperts as a headless server inside a Linux container with the `--container` flag, which prints a sign-in link at every start; the server exposes only on localhost so it can be reached from a browser without installing Node or Git.

Run exxperts as a headless server inside a Linux container, exposing it only on localhost, so you can access it from your browser without installing Node or Git.

```
$ exxperts --container
```

— Exact flag and behavior given, runnable command provided v0.16.1

02 Per-connector tool permissions and status pages

Not checked yet

NEW

60

Each connector tool can be set to Always allow, Ask first, or Blocked to control whether it runs without prompting, including while the user is away. Every connector now has its own page showing its status, its tools, and the rooms that use it.

— Describes the three permission states and new status page but no exact UI path v0.16.1

03 Slash skill picker in message box

Not checked yet NEW

60

Type `/` in the message box to see and invoke a room's skills directly, without leaving the message composer.

Quickly see and pick a skill while composing a message in a room, without leaving the message box.

📍 In any room's message box, type / to open the skill picker and select the skill to invoke.

— Exact trigger key and steps provided v0.16.1

THINNER COVERAGE BELOW

04 **Auto rooms skip permission prompts when protection is off**

Not checked yet

BREAKING

53

Rooms set to Auto now run without asking for permission when protection is off, including using connectors, tools set to Ask first, and sending data to websites. A room already set to Auto without protection runs this way from this update on, including after the user leaves it.

— Explains the behavior change but no mitigation steps given v0.16.1

05 **Skill auto-updates for repository skills** Not checked yet NEW

52

Skills added from a repository are kept up to date: exxperts checks for new versions shortly after start and then daily, with Update automatically available as a switch in Settings.

— Names check cadence and the Settings switch, no exact config key v0.16.1

06 **Cheaper long Claude runs through caching gateways** Not checked yet

IMPROVED

52

Long Claude runs through a gateway such as LiteLLM now cost less where the gateway reports that it caches that model, and exxperts now sends its name and version with each request.

— Names LiteLLM and the request metadata change, no config detail v0.16.1

07 **Several built-in connectors removed from directory**

Not checked yet

BREAKING

50

HubSpot, Asana, Google Drive, Gmail and Microsoft 365 have been removed from the connector directory; connectors already added to a room remain functional.

— Names all five removed connectors but no migration path v0.16.1

08 **Large image handling in conversations** Not checked yet IMPROVED

40

Conversations now support many large pictures: older images stay in context but are not re-sent once total size exceeds a limit, and the assistant can re-read a file when needed.

— Describes mechanism but no concrete size limit given v0.16.1

09 **Windows PowerShell no longer requires Git for Windows**

Not checked yet IMPROVED 32

Windows PowerShell now runs exxperts commands without requiring Git for Windows.

— States the change but gives no further mechanism v0.16.1

└──▷ **BREAKING ON UPGRADE**

! HubSpot, Asana, Google Drive, Gmail and Microsoft 365 have been removed from the connector directory; connectors you had already added remain.

! A room already set to Auto without protection now runs without asking for permission from this update on, including after you leave it.

WAS THIS USEFUL? 👍 👎

openexecutive

Feedback ⓘ v0.7.8 NOTES ↗ RANK

Openexecutive is an ai-powered virtual executive team — a single coherent executive persona backed by 8 specialist agents (FastAPI + Next.js).

> OpenExecutive's v0.7.8 release adds finer orchestrator routing rules, own-inbox monitoring for the Executive, a per-turn web search budget, individual undo actions on alerts, and direct email sending from delegation cards.

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

01 **Executive inbox watching in watch list** Not checked yet NEW 55

Adds the ability to watch your own inbox from the watch list, with Executive inbox watching automatically joining a card on matched threads and sending a ping when no reply arrives.

— Describes behaviour and trigger conditions clearly v0.7.8

02 **Direct email send from delegation cards** Not checked yet NEW 50

Adds the ability to send email directly from every email card in the delegation view, with tables rendered as tables.

— Names exact UI surface and a formatting detail v0.7.8

03 **Topic and Next time orchestrator rules** Not checked yet NEW 40

Adds topic rules and a card's Next time rule to the orchestrator, enabling finer-grained workflow routing for how cards are handled.

— Names the rule types but no mechanism detail v0.7.8

04 **Undo actions on Needs you alerts** Not checked yet NEW 40

Adds Undo actions to each Needs you alert, shown individually with a reason explaining the action.

— Clear UI behaviour but no further mechanism v0.7.8

05 **Per-turn web search budget** Not checked yet NEW 35

Adds a web search budget per chat turn so the assistant can look things up during a conversation.

— Names the feature but no budget size or config given v0.7.8

WAS THIS USEFUL? 👍 👎

nac

Feedback ⓘ v0.1.4 NOTES ↗ RANK

NAC open-source agent harness

> nac v0.1.4 adds the ability to fork a completed chat into a new independent session that preserves the original workspace, model, project and conversation state.

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

01 Conversation forking into new session

Not checked yet NEW

67

Using the Create fork action on a completed model turn, a finished chat can be cloned into a new independent session that keeps the same workspace, model, project, threads, episodes, live thread events, steering, worksets, and workspace revisions.

— Names exact UI action and all preserved state, but no further mechanism detail. v0.1.4

WAS THIS USEFUL? 👍 👎

LlamaIndex

Feedback ⓘ announced · Nov 23, 2026 BLOG ↗ RANK

LlamaIndex is an open-source framework that connects language models to data sources.

> LlamaIndex shipped a new local, CPU-only document parser (OCRia-nano-1) via a standalone LiteParse package, and added quality, confidence, and metadata controls to its Extract API's agentic_plus tier.

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

01 OCRia-nano-1 local document parser via LiteParse

Not checked yet

NEW

90

LlamaIndex introduced OCRia-nano-1, an 8.1M-parameter model for local CPU document parsing covering text recognition, table structure recovery, and layout identification inside the new LiteParse pipeline. Install it with `pip install liteparse` and `lit models pull ocria-nano`, then run `lit parse doc.pdf --format markdown` to output Markdown, JSON, or plain text with reconstructed tables, headings, and bounding boxes. It supports multilingual text recognition across Latin, CJK, right-to-left, and other scripts via a script router that handles per-line overrides and upside-down page detection.

Install LiteParse and download OCRia-nano-1 weights to parse a local PDF to Markdown on a CPU with no GPU required.

```
$ pip install liteparse
lit models pull ocria-nano
lit parse doc.pdf --format markdown
```

— Full command sequence, model specs, and mechanism given launch-20261123-3badd4b0

02 Extract quality, confidence and metadata controls

Not checked yet

NEW

80

LlamaIndex Extract added a `max` setting for the `effort` field on the `agentic_plus` tier, trading variable per-page billing for higher extraction quality on complex documents. The extraction configuration passed to `client.extract.run` now also supports `confidence_scores` and `cite_sources` fields, and job metadata can be retrieved by passing `expand=["extract_metadata"]` to `client.extract.get`.

Run a high-quality extraction job on a complex PDF where missing values are costly, using max effort on the agentic_plus tier.

```python
job = client.extract.run(
    file_input=file.id,
    configuration={
        "tier": "agentic_plus",
        "effort": "max",
        "version": "latest",
        "extraction_target": "per_doc",
        "data_schema": Invoice.model_json_schema(),
        "confidence_scores": True,
        "cite_sources": True,
    },
)
```

— Names exact config fields and API calls, limited on billing mechanics

launch-20261029-a4ab8ffd

WAS THIS USEFUL? 👍 👎

◆ Applied AI Agents

OpenClaw

Feedback ⓘ v2026.10.34 NOTES ↗ CODE ↗ RANK

OpenClaw is a personal AI assistant platform that connects messaging channels, tools, and model providers through a gateway.

> OpenClaw's latest release adds GPT-6.2 Astra and Sol model support with prompt caching, refines Tlon group DM approval handling, and retires MiniMax M2 models in favor of M3 for PDF extraction.

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

01 **GPT-6.2 Astra and Sol model support** Not checked yet NEW

65

Users can now select the GPT-6.2 Astra and Sol models in both OpenAI and ClawRouter, with system-prompt caching enabled on the OpenAI API-key route and the Sign in with ChatGPT Responses route.

— Names models and routes but no setup steps v2026.10.34

02 **MiniMax M2 retirement and M3 PDF extraction** Not checked yet

60

BREAKING

MiniMax now uses MiniMax-M3 for PDF text extraction after MiniMax M2 models were retired. MiniMax M2 models are no longer offered, including during live discovery, so any configuration that explicitly selects an M2 model must be switched to a supported model.

— Clear migration note but no exact config key given v2026.10.34

THINNER COVERAGE BELOW

03 **Tlon group DM approval policy changes** Not checked yet IMPROVED

45

Tlon group DMs now use their own approval policy separate from other channels, and approved clubs remain authorized across Gateway restarts.

— Describes behavior change but no config detail v2026.10.34

└──▷ **BREAKING ON UPGRADE**

! MiniMax M2 models are no longer offered, including during live discovery. If your configuration explicitly selects an M2 model, switch it to a supported model.

WAS THIS USEFUL?  

◆ Agent Sandboxes

cplt

Feedback ⓘ 2026.11.25-191731-6d6c48f NOTES ↗ RANK

cplt is a kernel-level sandbox for AI coding agents, running Copilot CLI, Claude Code, OpenCode, Gemini CLI, and others with git and gh guards.

> cplt's eight releases extend Linux sandboxing to parity with macOS (Landlock/seccomp-BPF for worktree add, devpts access under bwrap), add worktree lifecycle and policy-conflict safeguards, and introduce new config introspection, session-summary, and per-repo network-policy surfaces.

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

01 `allow.domains` in `.cplt.toml` policy

Not checked yet NEW

90

Adds `allow.domains` support at the managed policy layer in `.cplt.toml`, giving teams per-repo control over outbound network filtering without overriding user-level config.

Permit specific outbound domains for a repo's sandboxed agent sessions via committed policy, so all developers sharing the repo get the same network allowlist without touching their personal config.

toml

```toml
[allow]
domains = ["api.example.com", "registry.npmjs.org"]
```

— Names config file, key, and shows runnable config snippet 2026.11.24-163056-57898c7

02 `--json` flag for `config get`

Not checked yet NEW

85

`config get` now accepts a `--json` flag that reports the effective configuration value and which layer it comes from (global default, user config, or repo policy).

Inspect where a config value is being set (global default, user config, or repo policy) in a scriptable format.

```
$ cplt config get --json sandbox.quiet
```

— Exact flag plus runnable example command 2026.11.24-205609-1ddbc16

03 Linux sandboxing improvements

Not checked yet IMPROVED

80

Improves Linux sandboxing: grants `/dev/pts` access inside bubblewrap's private devpts namespace for the `bwrap` sandbox backend, and runs `cplt worktree add` sandboxed under Landlock LSM and seccomp-BPF on Linux, alongside the existing macOS Seatbelt implementation.

— Names mechanism and backend but no usage command Names exact enforcement mechanisms and platforms, lacks example

2026.11.25-093326-bd46900 2026.11.25-090317-831cdfa

---

04 **Session summary via CPLT_SESSION_SUMMARY** Not checked yet NEW

80

Writes a session summary to a file path set via the `CPLT_SESSION_SUMMARY` environment variable, allowing external launchers to read structured results after the agent exits.

Capture a session summary to a temp file so a launcher script can inspect what the agent did after it exits.

```
$ CPLT_SESSION_SUMMARY=/tmp/cplt-summary.json cplt -- -p "fix the tests"
```

— Names env var and shows runnable example 2026.11.24-194709-a9e458a

THINNER COVERAGE BELOW

05 **Worktree removal command** Not checked yet NEW

50

Adds a new `cplt worktree remove` subcommand to remove a worktree from an active sandboxed session.

Remove a worktree from an active sandboxed session after finishing work on a feature branch.

```
$ cplt worktree remove
```

— Names exact subcommand with a runnable example 2026.11.25-191731-6d6c48f

---

06 **Browser opening for private-domain links on macOS** Not checked yet

IMPROVED

50

cplt now opens `allow_private_domains` links in the host browser from every macOS session, not just the one that initiated it.

— Names the config field but gives no command or UI path 2026.11.24-102148-27e7fbf

---

07 **Policy allow/deny overlap warning** Not checked yet NEW

40

cplt now warns when an allow override in policy config overlaps a deny entry, surfacing conflicts before they cause unexpected access.

— Explains behavior but no config key or command named `2026.11.25-191731-6d6c48f`

---

08 **Gradle loopback launch hint** Not checked yet NEW 30

Adds a launch hint when a Gradle project meets a closed loopback condition, surfacing guidance before the agent stalls.

— No command, config key, or condition detail given `2026.11.24-171721-e131eaf`

WAS THIS USEFUL? 👍 👎

NVIDIA OpenShell

Feedback ⓘ v0.1.10 NOTES ↗ RANK

OpenShell runs autonomous AI agents in policy-controlled sandboxes with isolated filesystem, network, and credential access.

> OpenShell v0.1.10 adds offline policy linting and experimental Cedar-based sandbox policies, richer audit records, OTLP metrics export, workspace secrets, and SPIFFE-based sandbox identities, alongside several breaking changes to Helm chart distribution, gateway-mediated app URLs, and the TypeScript SDK package.

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

01 Richer audit records with resource-level capture

Not checked yet

IMPROVED

90

Gateway audit logs can now record setting values or exec arguments via `capture_setting_values = true` or `capture_exec_command = true` under `[openshell.gateway.audit]`. Audit records now name the specific resources each change touched — workspaces, sandboxes, providers, credential keys, settings, draft policy chunks, exec channels, port forwards, and SSH sessions — and flag no-op and replayed changes.

— Config keys and full resource list named; mechanism clear. v0.1.10

02 Offline policy linting and experimental Cedar policies

Not checked yet NEW

80

`openshell policy lint policy.yaml` lints policy files offline before deploying them to a gateway. In development builds, OpenShell also adds experimental support for creating and enforcing sandbox policies with Cedar via `--features experimental`, with existing YAML policies continuing to work.

Catch policy errors locally before pushing to a gateway — useful in CI or during policy authoring.

```
$ openshell policy lint policy.yaml
```

— Runnable command and flag named; Cedar policy scope left unspecified. v0.1.10

03 Workspace secrets shared across providers

Not checked yet NEW

80

Store a shared secret once with `openshell secret create` and reference it across providers using `--material-secret`, enabled per-gateway with

`enable_workspace_secrets`.

— Command and flags given; no detail on secret storage backend. v0.1.10

04 **Helm supervisor capacity sizing and IPv6 listening**

Not checked yet NEW 80

New Helm values size shared supervisor capacity on Kubernetes with `supervisor.pool.maxInstancesPerMember` and `supervisor.pool.maxMembersPerGateway`, and enable running the gateway and shared supervisor on IPv6 clusters by setting `gateway.listenAddress` and `supervisor.listenAddress` to `::`.

— All four Helm keys named; no guidance on sizing values. v0.1.10

05 **Application URLs now sourced from supervisor origin**

Not checked yet 80

BREAKING

Application URLs now come from the supervisor application origin rather than the gateway connection. If you use the app listener, replace `domain` with `base_url`, deploy matching gateway, driver, and supervisor images, recreate existing dedicated sandbox pods with the new supervisor image, and upgrade the CLI.

— Full migration steps and field rename given. v0.1.10

06 **Gateway-mediated service routing disabled by default**

Not checked yet 75

BREAKING

Kubernetes Helm deployments no longer serve gateway-mediated sandbox service URLs or HTTP/WebSocket forwarding by default; set `gatewayConfig.'openshell.gateway'.enable_gateway_service_routing: true` before upgrading if you rely on that path.

— Exact Helm key and migration step given. v0.1.10

07 **Container images and Helm chart renamed and republished**

Not checked yet 75

BREAKING

OpenShell container images now publish to `quay.io/ai-openshell`; if `global.image.registry` is set to a mirror, the chart now pulls `<mirror>/ai-openshell/<name>` instead of `<mirror>/openshell/<name>`. The Helm chart itself is renamed from `helm-chart` to `openshell` and now publishes to `oci://quay.io/ai-openshell/charts/openshell` (previously `oci://ghcr.io/nvidia/openshell/helm-chart`).

— Exact old and new registry and chart paths named. v0.1.10

08 **OTLP metrics export from the gateway** Not checked yet NEW

70

Gateways can export metrics via OTLP by setting `enabled = true` under `[openshell.gateway.otlp.metrics]`, alongside existing Prometheus scraping.

Send gateway metrics to an OTLP collector (such as the OpenTelemetry Collector) while keeping Prometheus scraping active.

```toml
[openshell.gateway.otlp.metrics]
enabled = true
```

— Exact config key given; no detail on supported collectors. v0.1.10

09 **SPIFFE identities for Kubernetes sandboxes** Not checked yet NEW

70

Kubernetes sandboxes can obtain provider access tokens via the gateway's SPIRE Broker by configuring `[openshell.gateway.spiffe_broker]`, giving sandboxes SPIFFE-based identity for provider authentication.

— Config section named; no example values or broker setup steps. v0.1.10

10 **Wildcard credential path templates** Not checked yet IMPROVED

65

Credential `path_template` values can now end in `/**` to let one token cover every method under an API path prefix, such as `/bot{credential}/**`.

— Exact syntax and example given; scope of use left implicit. v0.1.10

11 **TypeScript SDK package moved** Not checked yet BREAKING

65

The TypeScript SDK has moved to `@ai-openshell/sdk` on npmjs.org; `@nvidia/openshell-sdk@0.1.10` is the last release of the old package.

— Exact old and new package names given. v0.1.10

THINNER COVERAGE BELOW

12 **OAuth refresh provider parameters** Not checked yet IMPROVED

55

OAuth refresh providers now accept `token_endpoint_auth_method=client_secret_basic` and `resource=<uri>` parameters.

— Parameters named; no example of full provider config. v0.1.10

13 **Scoped GitHub App installation tokens** Not checked yet NEW

50

Providers can mint and rotate GitHub App installation tokens scoped to a defined repository and permission set, so long-running workloads keep access without seeing the app's private key.

— Mechanism described but no flag or config key named. v0.1.10

14 **Docker gateway persistent storage and image updates**

Not checked yet NEW

50

Docker gateways now support persistent `/sandbox` storage for new workspaces and allow updating a stopped sandbox's image without changing its sandbox ID.

— Behavior described; no config key or command shown. v0.1.10

15 **Kubernetes 1.33 minimum tested version** Not checked yet BREAKING

30

Kubernetes 1.33 is now the minimum tested version; older clusters are not tested and may not work.

— Minimum version stated; no migration guidance given. v0.1.10

└──▷ **BREAKING ON UPGRADE**

! Kubernetes Helm deployments no longer serve gateway-mediated sandbox service URLs or HTTP/WebSocket forwarding by default; set `gatewayConfig.'openshell.gateway'.enable_gateway_service_routing: true` before upgrading if you rely on that path.

! Application URLs now come from the supervisor application origin rather than the gateway connection; if you use the app listener, replace `domain` with `base_url`, deploy matching gateway, driver, and supervisor images, recreate existing dedicated sandbox pods with the new supervisor image, and upgrade the CLI.

! OpenShell images now publish to `quay.io/ai-openshell`; if you set `global.image.registry` to a mirror, the chart now pulls `<mirror>/ai-openshell/<name>` instead of `<mirror>/openshell/<name>`.

! The Helm chart is renamed from `helm-chart` to `openshell` and now publishes to `oci://quay.io/ai-openshell/charts/openshell` (previously `oci://ghcr.io/nvidia/openshell/helm-chart`).

! The TypeScript SDK has moved to `@ai-openshell/sdk` on npmjs.org; `@nvidia/openshell-sdk@0.1.10` is the last release of the old package.

! Kubernetes 1.33 is now the minimum tested version; older clusters are not tested and may not work.

WAS THIS USEFUL? 👍 👎

DEPLOY

◆ ML Platforms & SDKs

Hugging Face OpenEnv

Feedback ⓘ v0.15.0 NOTES ↗ RANK

OpenEnv is an interface library for RL post training with environments.

> OpenEnv's v0.15.0 release adds a new sandboxed data analysis environment, FastMCP 4 session support, and a series of sandbox isolation hardening measures spanning environment server launches and session isolation, alongside a new public documentation site.

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

01 Environment server launch hardening

Not checked yet IMPROVED

70

UVProvider now starts environment servers without the caller's credentials, env servers run as PID 1 by using `exec` in shell-form CMDs for better signal handling in containers, and env values are kept out of the Docker command line in container launches.

– Names components and mechanism, no exact command shown. v0.15.0

02 Session isolation hardening across environments

Not checked yet IMPROVED

70

`ortools_env` now assigns one UID per session, kills all processes when a call ends, and blocks network access for sandboxed code, while sessions are now isolated in `julia_env` and `tbench2_env` local mode.

– Names specific environments and isolation mechanisms applied. v0.15.0

THINNER COVERAGE BELOW

03 openenv init runtime replay plan declaration

Not checked yet NEW

50

Adds `openenv init` runtime replay plan declaration so runtime validation runs.

– Names exact CLI command, limited explanation of mechanism. v0.15.0

04 New data_analysis_env environment

Not checked yet NEW

45

Adds `data_analysis_env`, a new environment that answers questions about real data files from a sandboxed Python session.

– Names environment but gives no usage example or config. v0.15.0

05 Published documentation site

Not checked yet NEW

35

Adds a published documentation site on GitHub Pages with versioning and PR previews.

— Describes a doc site location but no further specifics. v0.15.0

06 **FastMCP 4 session support and sampling guards** Not checked yet

NEW

30

Adds FastMCP 4 session support and harness sampling guards.

— Names FastMCP 4 but no mechanism or usage detail. v0.15.0

07 **Harbor sandbox start supervision** Not checked yet IMPROVED

30

Harbor now supervises sandbox starts and cleans up only its own sandboxes on shutdown.

— Brief behavioral note, no command or config named. v0.15.0

WAS THIS USEFUL? 👍 👎

OpenAI Python SDK

Feedback ⓘ v3.60.0 NOTES ↗ RANK

The OpenAI Python SDK provides typed Python clients for OpenAI model APIs, streaming, and asynchronous requests.

> OpenAI Python SDK added spend alert and response compaction fields, session vault updates via the API, and typed healthcare model IDs for the Responses API.

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

01 **Healthcare model IDs in Responses API** Not checked yet NEW

60

The SDK now exposes identifiers for healthcare-specific models as typed options on `client.responses.create`, letting callers target healthcare model IDs in the Responses API.

– Names exact client method and model targeting v3.59.0

THINNER COVERAGE BELOW

02 **Spend alert and compaction fields on responses** Not checked yet

NEW

33

API response objects now expose spend alert and response compaction fields for reading.

– Names fields but no exact names or usage v3.60.0

03 **Session vault updates via API** Not checked yet NEW

31

The SDK now supports updating session vaults via the API, a capability it did not previously have.

– No method name or endpoint given v3.60.0

WAS THIS USEFUL? 👍 👎

OpenAI TypeScript SDK

Feedback ⓘ v7.65.0 NOTES ↗

RANK

The OpenAI TypeScript SDK provides typed JavaScript and TypeScript clients for OpenAI APIs and streaming responses.

> OpenAI's TypeScript SDK added support for updating session vaults, new spend alert and response compaction fields in the API types, and healthcare model IDs for the Responses API.

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

01 **Healthcare model IDs in Responses API** Not checked yet NEW

45

New healthcare model IDs can now be used in client.responses.create() calls via the Responses API.

— Names exact method but not the specific model IDs v7.64.0

02 **Session vault updates via SDK** Not checked yet NEW

26

The SDK now supports updating session vaults through the API.

— No endpoint or method name given, only prose description v7.65.0

03 **Spend alert and response compaction fields** Not checked yet NEW

25

New spend alert and response compaction fields were added to the API types.

— Names the fields but no usage or structure detail v7.65.0

WAS THIS USEFUL? 👍 👎

Tracer Cloud OpenSRE

Feedback ⓘ v0.1.2026.11.26 NOTES ↗

RANK

OpenSRE is an open-source toolkit for building AI SRE agents.

> OpenSRE gave its agent native file editing and browsing tools plus batched tool calls with persistent prompt caching, cutting reliance on shell access and reducing multi-turn latency and cost.

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

01 **Native file tools for the agent** Not checked yet NEW INTEGRATION

NEW 62

The agent gains `apply_patch` as a tool to apply patches directly to files, matching the tool name already referenced in the system prompt, alongside native file read, search, and list tools that let it browse a codebase without spawning a shell, returning plain-text Python results.

– Names the apply_patch tool and read/search/list but no limits or syntax v0.1.2026.11.26

THINNER COVERAGE BELOW

02 **Batched tool calls with persistent prompt cache** Not checked yet

IMPROVED 35

Agents now batch tool calls and retain the prompt cache across reveals and sessions, reducing latency and cost for multi-turn investigations.

– Describes benefit but no mechanism or config detail v0.1.2026.11.26

NEW INTEGRATION Official MCP server presets for Grafana, Datadog, PagerDuty, and AWS

> v0.1.2026.11.25 NOTES ↗ STABLE
>
> └──▷ **GET THIS VERSION** ▾

```
$ git clone --branch v0.1.2026.11.25 https://github.com/Tracer-Cloud/opensre.git
# already have the repo? check out this version:
$ git checkout v0.1.2026.11.25
```

WAS THIS USEFUL?  

MotleyAI SLayer

Feedback ⓘ v2.0.0 NOTES ↗ RANK

SLayer is an embeddable semantic layer that lets AI agents and people query warehouse data through shared metric definitions.

> SLayer's v2.0.0 release adds full Ossie semantic-model interoperability (import, export, conformance mode, portability grading) alongside a new storage-upgrade CLI and a SQL-like expression syntax, while shipping a string of breaking changes to its MCP integration, module layout, and relationship/granularity handling.

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

01 Ossie semantic-model import, export and conformance mode

Not checked yet NEW 90

SLayer can bring in an Ossie semantic-model document verbatim with `slayer import-osi`, which rejects invalid documents with full error details, and write it back with `slayer export-osi`. Queries can be run in strict Ossie conformance mode by setting `conformance="ossie-l3"` to get Ossie-exact wide-table rows, output names, and error codes, and a model's portability across all Ossie ladder rungs can be graded with `slayer ossie-roundtrip <in> <out>`.

Bring an existing Ossie semantic-model file into SLayer, then export it back to verify round-trip fidelity.

```
$ slayer import-osi my_model.osi.yaml
slayer export-osi my_model.osi.yaml --out exported.osi.yaml
```

Run a query in strict Ossie conformance mode to get Ossie-exact rows and error codes when validating cross-tool portability.

```python
result = engine.query(
    population="orders",
    metrics=["sum(orders.order_total)"],
    dimensions=["orders.status"],
    conformance="ossie-l3",
)
```

— Names exact commands and conformance flag with runnable examples v2.0.0

02 Storage upgrade CLI and library function Not checked yet NEW

85

Pending storage-format migrations can be previewed with `slayer storage upgrade --dry-run`, then applied with `slayer storage upgrade` to write everything in the 2.0 format. Dataset files SLayer does not manage can be upgraded in-memory by calling slayer.storage.upgrade.upgrade_datasets(), which returns them in the 2.0 format without touching disk.

Check what stored documents would be rewritten before committing to the 2.0 format, then apply the migration.

```
$ slayer storage upgrade --dry-run
slayer storage upgrade
```

– Exact CLI flags and function path with a runnable dry-run example v2.0.0

---

03 **Core module paths reorganized** Not checked yet BREAKING

72

`slayer.engine.syntax` and `slayer.core.formula` moved to `slayer.core.syntax`; `slayer.osi.expression` moved to `slayer.importers.aggregates`. Old paths raise an error, except `parse_formula`, `parse_filter_expr` and their 1.x node classes, which keep working with a deprecation warning until 3.0.

– Exact old and new module paths given, with deprecation window v2.0.0

---

04 **Granularity name collision auto-rename** Not checked yet BREAKING

62

A 1.x custom granularity whose name collides with a new reserved function name (e.g. `left`) is renamed to `left_granularity` on load; saving that datasource is refused until `slayer storage upgrade` has stored the rename.

– Mechanism and example collision named, with upgrade path pointer v2.0.0

---

05 **SQL-like expression syntax** Not checked yet BREAKING

60

Expressions are now SQL-like across all databases: names match in any case, double quotes name a field, single quotes make a string, and `--` starts a comment — replacing the old `amount:sum` colon spellings.

– Rules are concrete but no migration command is given v2.0.0

---

06 **MCP integration requires mcp 2.x** Not checked yet BREAKING

60

SLayer now requires `mcp>=2.3`; create_mcp_server() now returns an mcp 2.x `MCPServer` and its call_tool() returns a `CallToolResult` instead of the previous types.

— Names exact version and return types but no code fix example v2.0.0

THINNER COVERAGE BELOW

07 **to_many_handling options deprecated** Not checked yet DEPRECATED

48

`to_many_handling: "broadcast"`, `strict`, and references relative to the population are deprecated and will be removed in 3.0.

— Names the deprecated values but no replacement syntax shown v2.0.0

08 **Relationship values require trust flag** Not checked yet BREAKING

43

A declared many-to-one relationship no longer changes a value unless it is marked `trusted`.

— States the behavior change but not how to mark trusted v2.0.0

└──▷ **BREAKING ON UPGRADE**

! SLayer now requires `mcp>=2.3`.

! create_mcp_server() now returns an mcp 2.x `MCPServer` and its call_tool() returns a `CallToolResult` instead of the previous types.

! `slayer.engine.syntax` and `slayer.core.formula` moved to `slayer.core.syntax`; `slayer.osi.expression` moved to `slayer.importers.aggregates` (old paths raise an error, except `parse_formula`, `parse_filter_expr` and their 1.x node classes, which keep working with a deprecation warning until 3.0).

! `to_many_handling: "broadcast"`, `strict`, and references relative to the population are deprecated and will be removed in 3.0.

! A 1.x custom granularity whose name collides with a new reserved function name (e.g. `left`) is renamed to `left_granularity` on load; saving that datasource is refused until `slayer storage upgrade` has stored the rename.

! A declared many-to-one relationship no longer changes a value unless it is marked `trusted`.

WAS THIS USEFUL?  

Anthropic Python SDK

Feedback ⓘ v2.14.0 NOTES ↗ RANK

The Anthropic Python SDK provides typed Python clients for Claude APIs, streaming responses, and tool use.

> The Anthropic Python SDK added API support for compliance review cases in v2.14.0.

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

01 **Compliance review case support** Not checked yet NEW 25

The SDK adds support for compliance review cases, notices, and items via the API.

— Names a capability but no endpoints, fields, or methods given v2.14.0

WAS THIS USEFUL? 👍 👎

◆ AI / LLM Gateways

dario

Feedback ⓘ v6.24.2 NOTES ↗ RANK

dario is a local proxy that routes AI coding tools through your pool of Anthropic API keys with seat tracking and family-level failover.

> dario added Claude Code 2.1.332 compatibility and reworked how it merges client tool declarations with bundled definitions, including smarter required-field handling and independent routing for case-only tool name aliases.

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

01 **Client tool declaration merging improvements** Not checked yet

IMPROVED 80

The bundled `ToolSearch` definition now makes `max_results` optional with a default of 5, matching Claude Code 2.1.332 behaviour. The proxy also preserves a client tool's top-level required fields when merging with the bundled definition — for example, a client declaring `ToolSearch` with `max_results` required gets `max_results` marked required upstream. Case-only aliases of a bundled tool name (such as `toolsearch` next to `ToolSearch`) now go out as independent tools carrying the client's own `input_schema`, routed under names like `mcp__client__read` or a numbered variant such as `mcp__client__read_2` when the base name is already taken.

– Mechanism and named surfaces detailed but no user-facing command v6.24.2

THINNER COVERAGE BELOW

02 **Claude Code 2.1.332 support** Not checked yet IMPROVED 45

Claude Code 2.1.332 is now in dario's supported range, so `dario doctor` and proxy startup no longer warn that it is untested.

– Names command and version but no further mechanism v6.24.2

WAS THIS USEFUL? 👍 👎

Mozilla AI Otari

Feedback ⓘ v0.57.0 NOTES ↗ RANK

Mozilla AI Otari is an open-source AI gateway you run yourself. One endpoint for 40+ providers, with virtual keys, budgets, and usage tracking.

> Otari's biggest addition this window is cost-saving visibility and auditability for its subagent router, alongside per-key account/server isolation for organization keys, a new Keenable search provider, tenant-endpoint authentication, operator-controlled SSO disable, and stable API error codes.

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

01 Subagent router cost tracking and decision auditability

Not checked yet NEW 75

Otari now records each subagent routing decision for auditability, accepts a subagent's outcome to price its cost baseline for comparison, reports per-routed-subagent token usage via `otari-router`, and tracks in its inference layer whether a deployment paid for a given completion, while a new router savings panel in the sessions view shows cost saved per session and the home screen now shows the total cost saved by the subagent router across all sessions.

— UI path given but no config or API detail Names `otari-router` surface but no exact command

v0.57.0

02 Stable error codes for two API error conditions Not checked yet

IMPROVED 60

The user-mismatch 403 and the request-MCP-servers 400 error responses now have stable machine-readable codes.

— Names exact status codes but not the code values v0.56.0

THINNER COVERAGE BELOW

03 Keenable search provider in any-search Not checked yet NEW

50

Adds Keenable as a search provider in `any-search`, with or without an API key.

— Names config surface `any-search` but no exact key v0.56.1

04 Mail status field now reports delivery progress Not checked yet

50

BREAKING

The mail status field now publishes how far an email got in delivery, not whether sending is possible — consumers reading the previous boolean capability signal must update to

the new delivery-progress value.

– Migration note given but no exact field name v0.57.0

05 **Invitations response replaces booleans with next_step field**

Not checked yet BREAKING 50

The invitations response replaces correlated booleans with a single `next_step` field.

– Names exact field but not the endpoint path v0.56.0

06 **Per-key account and server binding for organization keys**

Not checked yet IMPROVED 45

Organization keys now carry their own account or their own server, isolating provider context per key.

– Explains isolation but no config key or endpoint named v0.56.0

07 **Operator-controlled SSO disable in identity settings**

Not checked yet NEW 40

Operators can now turn organization SSO off for a deployment via the identity settings.

– UI path given, no API or flag named v0.57.0

08 **Tenant endpoint authentication behind edge proxy** Not checked yet

NEW 40

Adds support for authenticating to a tenant endpoint behind an edge proxy in `providers`.

– Names `providers` module, no exact flag given v0.56.1

09 **Native pi harness in otari hook** Not checked yet NEW 40

Adds a native pi harness to `otari hook`.

– Bare addition with no further mechanism described v0.56.1

10 **Workspace-scoped sessions view links** Not checked yet IMPROVED

35

Otari's sessions view now lets a link specify which workspace it displays and carries the team home workspace filter into sessions views.

– No concrete parameter name or URL format given v0.57.0

└──▷ BREAKING ON UPGRADE

! The mail status field now publishes how far an email got in delivery, not whether sending is possible — consumers reading the previous boolean capability signal must update to the

new delivery-progress value.

! The invitations response replaces correlated booleans with a single `next_step` field.

WAS THIS USEFUL? 👍 👎

opencode-models-discovery

Feedback ⓘ v2.3.0 NOTES ↗

RANK

opencode-models-discovery is an OpenCode plugin that discovers available models from local LLM servers like Ollama, vLLM, and LM Studio.

> opencode-models-discovery v2.3.0 adds preservation of reasoning effort tiers from models.dev and introduces structured surfacing of authentication errors across the TUI and agent tools.

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

01 Reasoning effort variant preservation from models.dev

Not checked yet NEW 84

When `modelInfoFormat` is set to `"models.dev"`, the plugin preserves declared reasoning effort tiers from `models.dev` — including custom tiers — with provider-specific efforts taking precedence over defaults, and variants are cached across discovery runs.

Enable reasoning effort variant preservation from models.dev for a provider so discovered models carry declared effort tiers automatically.

json

```
{
  "providers": {
    "gateway": {
      "settings": {
        "modelsDiscovery": {
          "enabled": true,
          "modelInfoFormat": "models.dev"
        }
      }
    }
  }
}
```

— Names exact config key and behavior with a runnable config example. v2.3.0

02 Authentication error surfacing in discovery UI and tools

Not checked yet NEW 74

Authentication errors (HTTP 401/403) and provider discovery failures are now surfaced as `discoveryFailed` RPC events, as TUI toasts with `/connect` guidance, and as `[Auth Error]` badges and alerts in `/models-discovery-console`. The

`models_discovery_refresh` and `models_discovery_status` agent tools also return structured per-provider failure details and `/connect` hints when authentication fails.

— Names RPC event, UI path, badge, and two agent tools, but no runnable command. v2.3.0

WAS THIS USEFUL? 👍 👎

BerriAI LiteLLM

Feedback ⓘ v1.111.0 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 latest releases add OpenSearch pass-through and attribute-based access groups to the proxy, finer-grained per-model rate limits and team-level spend attribution, a new Rust-based cost-map crate, support for MiniMax-H3-Image and Claude Fable 5.2 models, and two breaking changes to transcription billing and RAG ingest file IDs.

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

01 **MiniMax-H3-Image on image generation and edits** Not checked yet

NEW

75

Call MiniMax-H3-Image on `/v1/images/generations` and `/v1/images/edits` with cost tracking now included.

— Exact endpoints named, directly runnable v1.108.17

02 **RAG ingest returns managed file IDs with team credentials**

Not checked yet

BREAKING

65

RAG ingest now returns managed file IDs and uses team credentials on vector store file routes — callers that expected raw upstream file IDs will receive different values.

— Names the affected routes and ID change, migration note implicit v1.111.0

03 **Per-model TPM/RPM limit maps for budgets and customers**

Not checked yet IMPROVED

60

Adds per-model `tpm` and `rpm` limit maps to budgets and customers, so rate limits can be set at the individual model level rather than globally.

— Names the limit fields but not the exact config path v1.111.0

04 **litellm-s3 Rust crate for cost-map handling** Not checked yet NEW

60

Adds the `litellm-s3` Rust crate with `AwsCredentialsProvider` and cost-map fetch, moving cost-map inputs and registration into Rust.

— Names the crate and class but no usage entry point v1.111.0

05 **New Claude Fable and Mythos model support** Not checked yet NEW

60

Adds `claude-fable-5-2` and its AWS Bedrock mirror to the Anthropic provider, plus `azure_ai/claude-fable-5-2` and `claude-fable-5-2-fast` from Azure AI Foundry, along with GovCloud regional pricing rows for Bedrock Claude Fable 5.2 and Mythos 5.2.

— Names every model and region added, no usage example given v1.111.0

06 **Transcription billing now uses provider-reported seconds**

Not checked yet BREAKING 60

Transcription cost billing now uses the seconds the provider reports rather than a locally measured duration — existing integrations that relied on client-side duration measurement will see different spend figures.

— States before/after clearly but no endpoint or flag named v1.111.0

THINNER COVERAGE BELOW

07 **OpenSearch index pass-through with per-key grants** Not checked yet

NEW 55

Adds `opensearch` index pass-through to the proxy, letting individual API keys carry scoped read/write grants to OpenSearch indexes.

— Names the surface but no exact route or config key v1.111.0

08 **Gemini page and sheet range scoping for file inputs**

Not checked yet NEW 55

Scope Gemini file reads to a page or sheet range on chat completions so only that portion is read and billed.

— Behaviour described, no exact parameter name given v1.108.17

09 **Logging credential decoupled from key** Not checked yet IMPROVED

50

Lets a key's logging entry reference a proxy credential, decoupling the logging destination auth from the key itself.

— Explains the mechanism but no named field or endpoint v1.111.0

10 **Customer spend split by team** Not checked yet IMPROVED

40

Splits customer spend by the key's current team, giving per-team attribution inside a single customer's usage report.

— Describes the behaviour change without a specific API or report path v1.111.0

11 **Boot-time warning for deprecated empty-models default**

Not checked yet IMPROVED 40

Logs keys and teams still on the deprecated empty-models default at proxy boot, surfacing misconfigured objects before they cause runtime errors.

— Clear behaviour but no log format or flag named v1.111.0

12 **Attribute-based access group grants** Not checked yet NEW 35

Grants access groups to principals through attribute rules, enabling policy-driven group membership without manual assignment.

— Mechanism named only in prose, no config surface given v1.111.0

└──▷ **BREAKING ON UPGRADE**

! Transcription cost billing now uses the seconds the provider reports rather than a locally measured duration — existing integrations that relied on client-side duration measurement will see different spend figures.

! RAG ingest now returns managed file IDs and uses team credentials on vector store file routes — callers that expected raw upstream file IDs will receive different values.

WAS THIS USEFUL? 👍 👎

◆ Inference Serving & Optimization

Niko1221 Strata

Feedback ⓘ v0.2.22 NOTES ↗ RANK

Strata is a local inference engine with one-click install for Windows and Linux, providing an OpenAI/Anthropic-compatible API on localhost for models like Qwen and DeepSeek.

> Strata's v0.2.22 release adds fine-grained controls over RAM allocation, host access, engine startup timing, and Inkling 2 Mini's reasoning behaviour, alongside a sandboxing fix for SVG/HTML rendering.

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

01 Tunable RAM reserve for 2-card layer splits

Not checked yet NEW

90

`STRATA_SPLIT_LEND_RESERVE` controls how much RAM a 2-card layer split reserves for the prompt path: set it to `0` to fill experts first for up to +9% decode speed, to a value between `1` - `99` to reserve that percentage, or leave it unset to keep the existing default.

Maximise decode throughput on a 2-card NVIDIA setup by filling experts before reserving prompt RAM.

```
$ STRATA_SPLIT_LEND_RESERVE=0 strata
```

— Names exact env var, value range, and measured speed gain v0.2.22

02 Inkling 2 Mini thinking-token budget cap control

Not checked yet

NEW

90

Set `"inkling_budget_cap": false` in the run config or `STRATA_INKLING_BUDGET_CAP=0` to lift the 8,192-token thinking-cap ceiling on Inkling 2 Mini, letting Claude Code's default 31,999-token requests through in full.

Remove the 8,192-token thinking cap so Claude Code's full 31,999-token budget reaches Inkling 2 Mini.

```
$ STRATA_INKLING_BUDGET_CAP=0 strata
```

— Names both config key and env var plus exact token numbers v0.2.22

03 Configurable engine readiness timeout

Not checked yet IMPROVED

75

`engine_ready_s` or `STRATA_ENGINE_READY_S` sets how long Strata waits for the engine to become ready; a slow disk no longer terminates startup after 15 minutes because the timer now resets on each new log line or disk read, up to a 2-hour maximum.

— Names config key, env var, and before/after timeout behaviour v0.2.22

---

04 **Host allowlisting for local API server** Not checked yet BREAKING

70

`STRATA_ALLOWED_HOSTS` lets specific hostnames reach Strata's built-in server when using a hosts-file alias or reverse proxy; requests from other browser tabs are now blocked by default.

Let Strata be reached through a local reverse-proxy hostname without blocking the request as an untrusted origin.

```
$ STRATA_ALLOWED_HOSTS=strata.local strata
```

— Named env var and runnable example, but default-change impact unquantified v0.2.22

---

05 **Inkling 2 Mini image size cap** Not checked yet NEW 70

Images above 0.8 MP handed to Inkling 2 Mini are now downscaled automatically; set `inkling_image_max_mp` to another size, or `0` to disable the cap entirely.

— Names config key and threshold but has no runnable example v0.2.22

---

06 **Per-request reasoning effort for Inkling 2 Mini** Not checked yet NEW 60

Inkling 2 Mini now honours the `reasoning_effort` field in a run config and the shared Chat settings, with priority order: request, Chat settings, config, then falling back to `low`.

— Names field and priority order but no runnable example v0.2.22

THINNER COVERAGE BELOW

---

07 **Sandboxed SVG and HTML rendering** Not checked yet IMPROVED 30

SVG and HTML files handed to the browser are now sandboxed so they cannot run script as part of the app.

— Brief security fix with no mechanism or surface named v0.2.22

WAS THIS USEFUL? 👍 👎

Experiential Labs Experiential

Feedback ⓘ

v0.7.263 NOTES ↗ RANK

Experiential is the open source, zero markup gateway for BYOK, self-hosted and 1000+ marketplace models. It learns from your traffic to cut costs, recommend better models, and train a specialized model you own.

> Experiential added support for serving Google Antigravity plans through Cloud Code Assist.

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

01 **Google Antigravity plan serving via Cloud Code Assist**

Not checked yet NEW 30

Experiential adds Google Antigravity plan serving through Cloud Code Assist, with native version `0.3.207`.

— Names integration and version but no usage detail v0.7.263

WAS THIS USEFUL? 👍 👎

Etalab OpenGateLLM

Feedback ⓘ 0.9.4 NOTES ↗ RANK

OpenGateLLM is an open-source, unified interface for running and managing self-hosted LLMs.

> OpenGateLLM 0.9.4 introduces OpenFGA as a required authorization engine governing all admin routes, alongside a reworked model-health endpoint and several breaking changes to admin API fields and container images.

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

01 OpenFGA authorization engine for admin routes

Not checked yet

95

BREAKING

OpenFGA becomes a required dependency, configured via `dependencies.openfga` (`url` , `store_name` , `api_token`) backed by environment variables `OPENFGA_HOST` , `OPENFGA_PORT` , and `OPENFGA_AUTHN_PRESHARED_KEYS` ; the API refuses to start without a reachable OpenFGA holding a store and a model. Idempotent backfill migrations run via `db-migrate` : `m001_platform_admins` grants `platform:api#admin` to every user whose role holds `ADMIN` , and `m002_organizations_platform` attaches every organization to the platform, with a `--dry-run` option that writes nothing; existing admins and organizations get `403` on the admin users/organizations routes until these migrations run.

Add the required OpenFGA block to your configuration file so the API can start and enforce relation-based access control on admin routes.

yaml

```yaml
dependencies:
  openfga:
    url: "http://openfga:8080"
    store_name: "opengatellm"
    api_token: "${OPENFGA_AUTHN_PRESHARED_KEYS}"
```

— Full config keys, env vars, and migration commands given. 0.9.4

02 Router load_balancing_strategy field removed

Not checked yet

85

BREAKING

The `load_balancing_strategy` field on routers is ignored when sent and no longer returned by admin router endpoints; remove it from router configuration and run Alembic migration `2b8e23234402` .

— Named field, endpoint behaviour, and exact migration id. 0.9.4

03 curl removed from api and playground images

Not checked yet

85

BREAKING

The `api` and `playground` images no longer ship `curl`; replace any healthcheck that uses it with the recommended `python -c "import urllib.request; urllib.request.urlopen('http://localhost:<port>/<path>', timeout=5)"`, e.g. `8000/health` for the api and `8501/ping` for the playground.

— Exact replacement command and ports given. 0.9.4

---

04 Model health recomputed from liveness and concurrency

Not checked yet

BREAKING

80

`GET /health/models` status is now computed from provider liveness and gateway concurrency load against `qos_limit`, replacing the previous provider `/metrics` source; a vLLM provider with waiting requests is no longer `yellow`, a closed provider that answers is `red`, and a model with one healthy provider out of several is `green`.

— Endpoint and exact status semantics given, no runnable command. 0.9.4

---

05 Role admin-permission changes blocked via PATCH

Not checked yet

65

BREAKING

`PATCH /v1/admin/roles/{role_id}` now returns `409` when the request adds or removes the `admin` permission of an existing role; change the user's role instead.

— Exact endpoint and status code, limited mechanism detail. 0.9.4

---

06 Provider basic_auth field removed

Not checked yet

BREAKING

60

The `basic_auth` field on providers is removed from `POST /v1/admin/providers` and from the configuration file; it is ignored and no longer present in the OpenAPI schema.

— Named field and endpoint, no migration or further scope. 0.9.4

└──▷ BREAKING ON UPGRADE

! `dependencies.openfga` (`url`, `store_name`, `api_token`) is now required in the configuration file; the API refuses to start without a reachable OpenFGA. Set `OPENFGA_HOST`, `OPENFGA_PORT`, and `OPENFGA_AUTHN_PRESHARED_KEYS` and create the `openfga` PostgreSQL database before upgrading.

! Existing admins and organizations get 403 on `POST` / `DELETE /v1/admin/organizations` and all `/v1/admin/users` routes until `db-migrate` runs `m001_platform_admins` and `m002_organizations_platform` to backfill OpenFGA relations.

! `PATCH /v1/admin/roles/{role_id}` now returns 409 when the request adds or removes the `admin` permission of an existing role; change the user's role instead.

! The `load_balancing_strategy` field on routers is ignored when sent and no longer returned by admin router endpoints; remove it from router configuration and run Alembic migration `2b8e23234402`.

! The `basic_auth` field on providers is removed from `POST /v1/admin/providers` and from the configuration file; it is ignored and no longer in the OpenAPI schema.

! The `api` and `playground` images no longer ship `curl`; replace any healthcheck that uses it. The recommended replacement is `python -c "import urllib.request; urllib.request.urlopen('http://localhost:<port>/<path>', timeout=5)"` (e.g. `8000/health` for the api, `8501/ping` for the playground).

! `GET /health/models` status values now reflect provider liveness and gateway concurrency load against `qos_limit` rather than provider `/metrics`; a vLLM provider with waiting requests is no longer `yellow`, a closed provider that answers is `red`, and a model with one healthy provider out of several is `green`.

WAS THIS USEFUL? 👍 👎

vLLM

Feedback ⓘ announced · Nov 24, 2026 BLOG ↗ RANK

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

> vLLM introduces vLLM-GR, a new open-source serving framework extending its engine to generative recommendation models with catalog-constrained beam search across NVIDIA and Ascend hardware.

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

01 **vLLM-GR framework for generative recommendation models**

Not checked yet NEW 78

vLLM now includes vLLM-GR, a serving framework for generative recommendation models such as `OneRec-1.7B`, `HSTU`, and `OneTrans` that runs beam search over candidate semantic IDs within a single request, accessible via an offline Python interface, OpenAI-compatible HTTP, or gRPC. Catalog-constrained beam search guides generation toward semantic IDs that map to real items in the item catalog, with per-stage beam widths and per-parent candidate limits configurable per request. It runs on NVIDIA GPUs (H20, A100, L20, RTX PRO 6000) and Ascend NPUs (A2, A5), reusing vLLM's paged KV cache and scheduler.

— Names models, interfaces and hardware but no exact commands shown.

launch-20261124-907b71ab

NEW INTEGRATION ERNIE 5.5 Flash Preview model support via vLLM-ERNIE plugin: baidu/ERNIE-5.5-Flash-Preview

> announced · Nov 24, 2026 BLOG ↗
>
> vLLM adds day-0 ERNIE 5.5 Flash Preview support via a standalone plugin, with DSpark speculative decoding, FP4/FP8 quantization, and FlashIndexer acceleration.
>
> Serve ERNIE 5.5 Flash Preview (280B MoE, 1M-token context) via the `vllm-ernie` plugin installed alongside `vllm==0.32.0`, discovered automatically through the `vllm.general_plugins` entrypoint with no fork or custom image required.
> Not checked yet
>
> Enable DSpark speculative decoding for ERNIE 5.5 Flash by passing `--speculative-config '{"method": "dspark", "num_speculative_tokens": 7, "draft_sample_method": "probabilistic"}'` to `vllm serve`, drafting up to 7 tokens per step. Not checked yet
>
> Control ERNIE 5.5 Flash reasoning and tool use via `--reasoning-parser ernie5`, `--tool-call-parser ernie5`, and `--enable-auto-tool-choice` flags on `vllm`

`serve` . Not checked yet

Set the DSA indexer cache to FP4 with `--attention-config '{"indexer_kv_dtype": "mxfp4"}'` on `vllm serve` , cutting long-context indexing cost by 40% at 1M context via FlashIndexer. Not checked yet

Disable chain-of-thought thinking per request by passing `"chat_template_kwargs": {"enable_thinking": false}` to the `/v1/chat/completions` endpoint. Not checked yet

---

WAS THIS USEFUL? 👍 👎

Google LiteRT

Feedback ⓘ v2.4.0 NOTES ↗

LiteRT is Google's on-device runtime, formerly TensorFlow Lite, for running ML and generative models on phones and edge hardware.

LiteRT v2.4.0 adds MoE model support on GPU/CPU, fused GPU kernels for hybrid LLMs, versioned NPU C ABIs, selective-op builds, and new per-inference timing APIs across C, C++, and Kotlin/Java.

└──▷ GET THIS VERSION ▾

```
$ git clone --branch v2.4.0 https://github.com/google-ai-edge/LiteRT.git
# already have the repo? check out this version:
$ git checkout v2.4.0
```

LiteRT now runs Mixture of Experts models such as Gemma 4 26B-A4B on GPU and CPU, executing only active experts per token; weight caching on disk cuts repeat startup time from ~30-45 s to ~2 s and saves over 11 GB of setup memory. Not checked yet

UPDATED SUPPORTED PLATFORM LiteRT adds fused GPU kernels across Metal, WebGPU, and OpenCL for hybrid models, cutting GPU launches per token by 58% and speeding up generation by up to 35% on Metal, 23% on WebGPU, and 18% on Android. Not checked yet

Override dynamic input shapes at AOT and on-device JIT compile time via new `CompilerOptions` APIs in C, C++, and Kotlin/Java. Not checked yet

Reduce binary size by selecting only the ops your model uses with the new `--//litert/build_common:cpu_backend={xnnpack,builtin,selective}` build flag; pass `models = ["my_model.tflite"]` to `litert_cc_binary`, `litert_cc_test`, or `litert_cc_library` to auto-register required ops, or use `litert::tflite_support::SetOpResolver` for explicit control. Not checked yet

Retrieve per-inference timing via `LiteRtGetCompiledModelLastInferenceDuration` (C), CompiledModel::LastInferenceDurationNanoseconds() (C++), and `CompiledModel.lastNativeInferenceDurationNanoseconds` or getLastNativeInferenceDurationNanoseconds() (Kotlin/Java). Not checked yet

+4 more

NEW SUPPORTED PLATFORM Windows ARM64 platform support: ai-edge-litert, litert_prebuilts_windows_arm64.zip

NEW SUPPORTED PLATFORM Linux aarch64 Qualcomm NPU support: ai-edge-litert, ai-edge-litert-sdk-qualcomm

NEW SUPPORTED PLATFORM iOS and macOS Swift Package Manager support: LiteRT-Swift

NEW SUPPORTED PLATFORM Arm Mali G2-Ultra NX experimental GPU backend via Arm Vulkan ML API

UPDATED INTEGRATION QAIRT SDK dependency upgrade to 2.51.1.261002

└──▷ **BREAKING ON UPGRADE**

- ! The `use_local_tf` bazelrc config and the `USE_LOCAL_TF` / `TF_LOCAL_SOURCE_PATH` CMake variables were removed.
- ! The LiteRT Dispatch C ABI and Compiler Plugin C ABI are now versioned interface tables (`LiteRtDispatchQueryInterface` , `LiteRtCompilerPluginQueryInterface` , `LiteRtAbiHeader`) with narrowed shared library exports, breaking out-of-tree backends built against the previous unversioned ABIs.
- ! Python 3.10 is the last supported release; `ai_edge_litert` on Python 3.10 now emits a `FutureWarning` , and v2.5.0 will require Python >= 3.11 and will not publish `cp310` wheels.

WAS THIS USEFUL? 👍 👎

◆ AI Chat Apps

Open WebUI

Feedback ⓘ v0.13.5 NOTES ↗ RANK

Open WebUI is a self-hosted chat interface for local and remote models, with RAG over uploaded documents, tool calling, and multi-user access control.

> Open WebUI's v0.13.5 release adds folder workspaces with terminal and browser previews, automatic model routing, ComfyUI-based video generation, SSH connections to remote machines, a Status panel for agent threads, richer task automation with webhooks and sharing, and notes embedded in sidebar folders.

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

01 **Project instructions from AGENTS.md or CLAUDE.md** Not checked yet

NEW 86

Project instructions and skills are read from an `AGENTS.md` or `CLAUDE.md` file placed in the project folder and applied automatically to every new chat in that folder, when connected via Open Terminal 0.18.0 or later.

Set standing instructions for every new chat in a project folder so the assistant knows the codebase conventions without being told each time.

📍 In the project folder, create AGENTS.md with contents such as: # Project conventions - Always use pnpm, not npm - Tests live in tests/ - Run `pnpm test` before committing

— Names exact filenames, version requirement, and worked example. v0.13.5

02 **Tool output size cap via `TOOL_OUTPUT_MAX_CHARS`** Not checked yet

NEW 83

Setting `TOOL_OUTPUT_MAX_CHARS` in the environment caps how much tool output is sent to the model's context window; when a machine is connected, the model can still open the full output as a file.

Prevent a long shell command output from flooding the model's context window by capping tool results at 4,000 characters.

```env
TOOL_OUTPUT_MAX_CHARS=4000
```

— Names exact env var and describes fallback behaviour. v0.13.5

03 **Video generation via ComfyUI workflows** Not checked yet NEW

83

Admins import a ComfyUI workflow in API format under Admin Settings > Images & Videos and configure inputs for prompt, dimensions, frame count, and frame rate, enabling video generation directly in chat.

Let an admin import a ComfyUI video-generation workflow so users can create videos from chat without opening ComfyUI.

> 📍 In Open WebUI, go to Admin Settings › Images & Videos, import your ComfyUI workflow in API format, then map the prompt, dimensions, frame count, and frame rate inputs.

– Exact UI path, inputs, and step-by-step example given. v0.13.5

---

04 **Remote machine connections over SSH** Not checked yet NEW

78

Connect a remote workstation over SSH to use its files and command-line tools in chat, requiring Open Terminal 0.17.21 or later on Linux x86-64, ARM64, or Apple silicon; on single-instance deployments, admins can also run commands on the Open WebUI server itself as a machine, via Open Terminal 0.17.21 or later.

– Names exact tool version and supported platforms. v0.13.5

---

05 **Webhook triggers and task sharing for automations** Not checked yet

NEW

63

Webhook triggering lets an external form submission pass its answers into an Open WebUI task prompt via a secret webhook URL, and task sharing lets owners grant teammates permission to view past runs or run and reschedule the task, which then executes with the owner's permissions.

– Names webhook mechanism and permission model clearly. v0.13.5

THINNER COVERAGE BELOW

06 **Automatic model routing by request type** Not checked yet NEW

58

A router can be placed in the model picker and configured to route coding questions to one model and writing requests to another, using text patterns or natural-language descriptions.

– Explains mechanism but no config key or UI path named. v0.13.5

---

07 **Notes inside sidebar folders with docked view** Not checked yet

NEW

55

Notes can live inside sidebar folders, inheriting the folder's read or edit access, with docked note chats that keep a note and a chat visible side by side and an option to collapse the chat to restore full writing width.

— Describes feature and UI behaviour, no config key named. v0.13.5

08 **Status panel for delegated agent threads** Not checked yet NEW

52

A Status panel in chat lets you follow delegated agent threads, approve tool calls, stop jobs, and see model context usage and queue depth.

— Describes capability without a config or command surface. v0.13.5

09 **Folder workspaces with terminal and browser preview**

Not checked yet NEW

50

Folder workspaces let you run terminal tests and a browser preview alongside a chat, with per-user tab and layout memory preserved in shared folders.

— Describes behaviour but no config surface given. v0.13.5

10 **Recurring tasks continue conversation thread** Not checked yet

IMPROVED

40

Recurring tasks can now continue a conversation thread, reading replies posted since the last run so each automation run accounts for recent discussion.

— Behaviour change described without any named surface. v0.13.5

WAS THIS USEFUL?  

◆ Inference Providers & GPU Clouds

Eigen Labs Darkbloom

Feedback ⓘ v0.10.19 NOTES ↗

CODE ↗ RANK

Darkbloom is a private Inference Network on Idle Macs

> Darkbloom added native Inkling 2 Mini support — a 524,288-token context window, native `tool_choice` enforcement, and thinking-off response cleanup — alongside a rollback command and a long list of per-feature environment variable kill switches.

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

01 **Per-feature rollback via environment variables** Not checked yet

NEW 90

Roll back any individual v0.10.18 behaviour change before `darkbloom start` by setting the corresponding environment variable to `0` — available variables include `DARKBLOOM_GLM_REFERENCE_TOOL_HISTORY`, `DARKBLOOM_GLM_SHARED_GRANT_WAIT`, `DARKBLOOM_MEDIA_ON_EVERY_ROLE`, `DARKBLOOM_MTP_STOP_STRING_ROWS`, `DARKBLOOM_MTP_PENALTY_ROWS`, `DARKBLOOM_GEMMA_MTP_M5_ULTRA_LONE_ROW`, `DARKBLOOM_GEMMA_NATIVE_ROWS_BEFORE_ASSISTANT`, `DARKBLOOM_CLUSTER_LEADERSHIP_WINDOW_WAIT`, `DARKBLOOM_CLUSTER_STOP_ALONE_UNCONFIRMED`, `DARKBLOOM_TTFT_SHAPE_WARMUP`, `DARKBLOOM_INKLING_UNDECLARED_CALL_TEXT`, `DARKBLOOM_DECODE_RATE_IDLE_MAX_AGE_SECONDS`, `DARKBLOOM_PREFIX_CACHE_PARALLEL_STAGE`, `DARKBLOOM_TTFT_SEGMENT_HOST_ZEROS`, `DARKBLOOM_GEMMA_MTP_MAX_PROMPT_TOKENS`, and `DARKBLOOM_TTFT_MEDIA_PREP_STREAM`. Setting `DARKBLOOM_TTFT_SHAPE_WARMUP=0` specifically disables Inkling 2 Mini shape warm-up at load time.

Pin a Mac Studio running Inkling 2 Mini at the previous thermal-width behaviour while testing the new fair-state rule, then start the provider.

```
$ DARKBLOOM_TTFT_SHAPE_WARMUP=0 darkbloom start
```

Roll back GLM-5.3-Flash tool-history handling and Gemma 4 26B speculative decoding with stop strings to 0.10.17 behaviour before starting the provider.

```
$ DARKBLOOM_GLM_REFERENCE_TOOL_HISTORY=0
DARKBLOOM_MTP_STOP_STRING_ROWS=0 darkbloom start
```

— Enumerates all named env vars with runnable examples v0.10.18

02 **Native tool_choice enforcement for Inkling 2 Mini** Not checked yet

NEW 85

Inkling 2 Mini now natively enforces `tool_choice` `required` or a named function at sampling time, returning a tool call instead of an error. Operators can disable this and restore the prompt-forced-only behaviour from v0.10.18 by setting `DARKBLOOM_INKLING_FORCED_TOOLS=0` before `darkbloom start`.

Opt out of native tool-call enforcement on Inkling 2 Mini and keep the prompt-forced behaviour from v0.10.18.

```
$ DARKBLOOM_INKLING_FORCED_TOOLS=0 darkbloom start
```

— Names exact env var and start command v0.10.19

03 **Hidden thought tokens excluded from completion metadata**

Not checked yet IMPROVED 80

Inkling 2 Mini now reports `completion_tokens_details.reasoning_tokens` as 0 when thinking is off (`reasoning.enabled` false, `enable_thinking` false, or effort `none` / `off` / `0`), keeping hidden thought tokens out of the reasoning count. It also omits the hidden thought from `logprobs` entries when thinking is off and `logprobs: true`, returning log-probability entries only for answer tokens and tool-call message tokens.

— Names exact fields and conditions, no command v0.10.19

04 **Inkling 2 Mini native serving with 524,288-token context**

Not checked yet NEW 75

Darkbloom now serves Inkling 2 Mini natively with a 524,288-token context window (prompt plus output); prompts of 524,288 tokens or more are refused with HTTP 400, and output is capped at whatever tokens the prompt leaves.

— Exact limits and error code, no direct command v0.10.18

05 **darkbloom update --rollback for paired Macs** Not checked yet NEW

60

Roll back to v0.10.18 on a paired Mac setup with `darkbloom update --rollback`.

Roll a paired Mac back to v0.10.18 after a failed upgrade.

```
$ darkbloom update --rollback
```

— Runnable rollback command, thin on mechanism v0.10.19

06 **Thermal-aware serving width extended to 'fair' state**

Not checked yet IMPROVED 60

Automatic serving width on M5 hardware now holds at thermal state `fair` in addition to `nominal`, increasing request concurrency for Macs under sustained load without manual profile changes.

— Names thermal states, no direct config action v0.10.18

THINNER COVERAGE BELOW

07 **darkbloom start resilience during model download** Not checked yet

IMPROVED 50

`darkbloom start` now stays up when an Autopilot model is still downloading, naming the held model and serving with the remaining verified models.

— Describes behaviour but no flag or command v0.10.19

└──▷ **ALSO FROM THESE RELEASES**

Audit an update that ran overnight without keeping a terminal open — the new disk-level log lines remain visible hours later.

```
$ darkbloom logs --last 24h
```

WAS THIS USEFUL? 👍 👎

Fireworks AI

Feedback ⓘ

Flex tier now available for GLM 5.3 and GLM 5.3 Flash at 50% off Standard

NOTES ↗ RANK

Fireworks AI provides hosted inference, model fine-tuning, and deployment APIs for open-weight models.

> Fireworks AI shipped an ultra-fast preview variant of Kimi K3.1 Pro, brought ultra-fast DeepSeek V4.1 Flash 1017 to general availability, added discounted Flex tier pricing for GLM 5.3 models, and tripled adaptive rate limits for medium and large serverless tiers.

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

01 Ultra-fast preview variant of Kimi K3.1 Pro

Not checked yet NEW

95

A preview ultra-fast variant of Kimi K3.1 Pro targets 3–6x generation speed (up to 8x for agentic coding tasks), accessed by setting `model` to `accounts/fireworks/routers/kimi-k3p1-pro-ultrafast`. It is priced at $5.00 input / $0.50 cached input / $25.00 output per 1M tokens, with preview ceilings of 21.6M Total Prompt TPM, 5.4M Uncached Prompt TPM, and 216k Generated TPM.

Call the ultra-fast Kimi K3.1 Pro preview variant to get 3–6x faster completions for agentic coding workflows.

```
$ curl https://api.fireworks.ai/inference/v1/chat/completions \
  -H 'Authorization: Bearer <API_KEY>' \
  -H 'Content-Type: application/json' \
  -d '{"model": "accounts/fireworks/routers/kimi-k3p1-pro-
ultrafast", "messages": [{"role": "user", "content": "Refactor
this function to use async/await."}]}'
```

Route an agentic coding workload to the ultra-fast Kimi K3.1 Pro variant to get up to 8x faster generation.

```
$ curl https://api.fireworks.ai/inference/v1/chat/completions \
  -H 'Authorization: Bearer <API_KEY>' \
  -H 'Content-Type: application/json' \
  -d '{"model": "accounts/fireworks/routers/kimi-k3p1-pro-
ultrafast", "messages": [{"role": "user", "content": "Refactor
this Python function for performance."}]}'
```

Call Kimi K3.1 Pro in preview ultra-fast mode to evaluate speed gains before the mode reaches general availability.

```
$ curl https://api.fireworks.ai/inference/v1/chat/completions \
  -H 'Authorization: Bearer <API_KEY>' \
  -H 'Content-Type: application/json' \
  -d '{"model": "accounts/fireworks/routers/kimi-k3p1-pro-
ultrafast", "messages": [{"role": "user", "content": "Summarise
this threat report."}]}'
```

— Full pricing, TPM limits, and runnable model ID given

Flex tier now available for GLM 5.3 and… | Tripled serverless rate limits for medi… | Ultra-fast is now generally available

---

02 **DeepSeek V4.1 Flash 1017 ultra-fast reaches GA** Not checked yet

IMPROVED 80

Ultra-fast DeepSeek V4.1 Flash 1017 is now generally available, delivering 3–6x faster generation (up to 8x for agentic coding), accessed via `model` set to `accounts/fireworks/routers/deepseek-v4p1-flash-1017-ultrafast`.

Call DeepSeek V4.1 Flash 1017 at generally available ultra-fast speed for high-throughput agentic coding pipelines.

```
$ curl https://api.fireworks.ai/inference/v1/chat/completions \
  -H 'Authorization: Bearer <API_KEY>' \
  -H 'Content-Type: application/json' \
  -d '{"model": "accounts/fireworks/routers/deepseek-v4p1-flash-
1017-ultrafast", "messages": [{"role": "user", "content":
"Refactor this function for performance."}]}'
```

— Runnable model ID and GA status clearly stated | Ultra-fast is now generally available

---

03 **Tripled serverless rate limits for Medium and Large tiers**

Not checked yet IMPROVED 65

Adaptive rate-limit ceilings were tripled for serverless models: Medium-tier now reaches 129.6M Total Prompt TPM, 32.4M Uncached Prompt TPM, and 1.296M Generated TPM; Large-tier now reaches 64.8M Total Prompt TPM, 16.2M Uncached Prompt TPM, and 648k Generated TPM.

— Exact numeric limits but no config action needed

Tripled serverless rate limits for medi…

THINNER COVERAGE BELOW

04 **Flex tier pricing for GLM 5.3 models** Not checked yet NEW

35

Flex tier pricing - 50% off Standard - is now available for GLM 5.3 and GLM 5.3 Flash via the Serverless modes model IDs.

— No exact model ID or command provided Flex tier now available for GLM 5.3 and…

WAS THIS USEFUL?  

Together AI

Feedback ⓘ announced · Nov 25, 2026 BLOG ↗ RANK

Run, train, and serve open-weight models on Together AI.

> Together AI launched Hurdle, an open-source npm library for running automated agent evals against APIs, SDKs, and docs, with suite definitions, per-run checks, guardrails, a results viewer, and sandboxed execution through Together Sandbox.

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

01 Guardrails to block dangerous agent actions

Not checked yet NEW

88

Block dangerous agent actions mid-run with `defineGuard` in a `guard.mjs` file, using guard.block.request() to intercept and fail specific HTTP methods and paths before they reach your API.

Prevent an agent from accidentally creating a paid resource during an eval by blocking the relevant API endpoint with a guard.

javascript

```
// guard.mjs
import { defineGuard } from 'hurdle';

export default defineGuard((guard) => {
  guard.block.request({
    id: 'no-reservation',
    method: 'POST',
    path: '/v1/compute/clusters',
  });
});
```

— Exact function names plus a runnable blocking example launch-20261125-13f1c697

02 Suite definition and API config in .suite.mjs

Not checked yet

NEW

85

Define reusable eval suites in a `.suite.mjs` file using `defineSuite`, specifying `api`, `tasks`, `agents`, `conditions`, `repeats`, `verdict`, and `judge` fields to control every eval dimension. Configure the API under test by setting `authorizationCredential` and `hosts` in the `api` block of a suite file, keeping keys out of source via `.env.local`.

Run the same task across two models under two conditions (with and without an MCP skill) 15 times each, scoring with both code checks and an LLM judge.

```javascript
// example.suite.mjs
import { defineSuite } from 'hurdle';

export default defineSuite({
  api: {
    authorizationCredential: 'TOGETHER_API_KEY',
    hosts: ['api.together.ai'],
  },
  tasks: ['./tasks/gpu-cluster'],
  agents: [
    { runner: 'claude-code', model: 'claude-opus-4-8' },
    { runner: 'claude-code', model: 'claude-sonnet-4-6' },
  ],
  conditions: {
    baseline: {},
    'with-skill': {
      skills: [{ name: 'together-gpu-clusters', sourceDirectory:
'skills/together-gpu-clusters' }],
    },
  },
  repeats: 15,
  verdict: 'hybrid',
  judge: { model: 'claude-opus-4-8' },
});
```

— Names every suite field and config key with a worked example `launch-20261125-13f1c697`

03 **Agent eval project bootstrap CLI** Not checked yet NEW

75

Bootstrap a new Hurdle eval project with `npx hurdle init`, or target Codex instead of the default Claude Code with `npx hurdle init --agent codex`.

Bootstrap a new Hurdle eval project and immediately run the starter task to verify your setup end-to-end.

```
$ mkdir my-evals && cd my-evals
npx hurdle init
npx hurdle run --view
```

— Exact commands and flag, minimal explanation of behavior `launch-20261125-13f1c697`

04 **Together API example eval project** Not checked yet NEW

70

Initialize a sandboxed Together API example project with `npx hurdle init api-evals --example together-api` and prepare credentials with `npx hurdle setup api-evals`.

— Exact scaffolding commands, thin on what the example covers launch-20261125-13f1c697

05 **Per-run checks and LLM judging** Not checked yet NEW 65

Write per-run assertions in a `checks.mjs` file using `defineChecks` from `hurdle`, with run.expect.request() for HTTP call checks and run.judge() for LLM-based rubric scoring.

— Names file and functions but no worked example launch-20261125-13f1c697

06 **Results viewer for eval runs** Not checked yet NEW 65

Run an eval suite and open the results viewer in one step with `npx hurdle run --view`; reopen saved results at any time with `npx hurdle view`.

— Exact commands given but no detail on viewer contents launch-20261125-13f1c697

07 **Local and sandboxed eval execution** Not checked yet NEW 60

Run agent evals locally with no sandbox or network isolation, or in isolated Together Sandbox environments where Guardian proxies all traffic, enforces guards, and swaps in real API keys without exposing them to the agent.

— Describes mechanism and scope but gives no command launch-20261125-13f1c697

WAS THIS USEFUL?  

DATA

◆ Vector DBs & RAG

liliang-cn CortexDB

Feedback ⓘ v2.171.0 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 shipped a new command-line interface exposing every MCP tool as a subcommand, plus a plugin for the Antigravity (agy) agent.

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

01 **CLI exposing every MCP tool** Not checked yet NEW 75

A new `cortexdb` CLI lets users call any MCP tool from a shell, with each tool exposed as its own subcommand with dedicated help text, e.g. `cortexdb search` and `cortexdb tool list`.

Search shared agent memory from the shell without opening any agent UI.

```
$ cortexdb search what did we decide about the deploy key
```

List all available MCP tools to discover what CortexDB exposes over the command line.

```
$ cortexdb tool list
```

— Names exact commands but little on internal mechanism. v2.171.0

02 **Antigravity (agy) agent plugin** Not checked yet NEW 65

CortexDB adds a plugin for the Antigravity (`agy`) agent, installable with `agy plugin install ./plugins/cortexdb`.

Install the CortexDB plugin into an Antigravity agent from a local clone of the repo.

```
$ agy plugin install ./plugins/cortexdb
```

— Install command given but no detail on plugin behavior. v2.171.0

WAS THIS USEFUL? 👍 👎

ngt

Feedback ⓘ v2.8.3 NOTES ↗ RANK

Ngt is a nearest Neighbor Search with Neighborhood Graph and Tree for High-dimensional Data

> NGT v2.8.3 adds inline filtering to graph and quantized searches and supports node removal in bi-directed graphs via a new incoming edge index.

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

01 **Inline filtering for graph and quantized searches** Not checked yet

NEW

40

NGT now supports inline filtering during graph and quantized searches, allowing result sets to be narrowed at query time without a post-filter pass.

— Explains mechanism but no API names or config given v2.8.3

02 **Node removal in bi-directed graphs** Not checked yet NEW

38

NGT now supports node removal in bi-directed graphs via an incoming edge index, enabling cleaner index maintenance when objects are deleted.

— Describes behaviour but no concrete API or flag named v2.8.3

WAS THIS USEFUL? 👍 👎

dosco GraphJin

Feedback ⓘ v3.22.0 NOTES ↗ RANK

GraphJin is a governed graph for AI agents — GraphQL + MCP over your databases, files, APIs, and code.

> GraphJin's biggest change this window was replacing its three operating modes with two servers whose config files declare sources directly, alongside a string of onboarding upgrades adding setup screens, model and database selection, and authentication to the console and agentic subsystem.

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

01 Guided setup and authentication across console and agentic mode

Not checked yet NEW 55

GraphJin's setup flow now lets users choose an AI model directly from the setup page, connect a database via a setup-page connection form with config boot from the store in agentic mode, and the console gained a setup screen with email-and-password sign-in; the agentic subsystem separately gained local accounts, first-admin setup, and password sign-in.

Select the AI model to use for agentic features without editing config files.

📍 In the GraphJin UI, go to Setup and select your preferred model from the model picker.

Connect a database to GraphJin's agentic mode without editing config files manually.

📍 In the GraphJin UI, go to Setup and use the database connection form to enter your database credentials and connect.

First-time setup of the GraphJin console using the new setup screen and email-and-password sign-in flow.

📍 Open the GraphJin console in your browser, follow the setup screen to configure your instance, then sign in with your email and password.

— UI navigation steps given, no config keys or APIs named v3.21.29 v3.21.30 v3.21.32 v3.21.33

02 Two-server architecture replacing three modes Not checked yet

39

BREAKING

GraphJin now runs as two servers instead of three operating modes, with config files declaring sources directly.

— No config key names or migration steps given v3.22.0

WAS THIS USEFUL? 👍 👎

synaplan

Feedback ⓘ v5.10.2 NOTES ↗ RANK

Synaplan is an own the flow: one open-source AI platform for your knowledge, models and channels – web, WhatsApp, email, Outlook, Nextcloud & ownCloud. Local and cloud models side by side, with the cost on every answer.

> Synaplan v5.10.2 adds optional API key authentication for Qdrant requests and automatic tagging of email-to-me results.

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

01 **Qdrant API key authentication** Not checked yet NEW 35

Synaplan now supports authenticating Qdrant requests by sending an optional Qdrant API key.

– Names the integration but no config key or flag given v5.10.2

02 **Automatic tagging for email-to-me results** Not checked yet

IMPROVED 23

Email-to-me results are now marked as automatic.

– Brief statement with no mechanism or UI detail v5.10.2

NEW INTEGRATION SpecStream web search provider

NEW INTEGRATION LLMTR API and models hosted in Turkey

NEW SUPPORTED PLATFORM Gemini 4 Neon model

v5.9.0 NOTES ↗ STABLE

synaplan adds chat-run routing to a run kernel or loop, source verification before delivery, cost limits, cycle detection, a task result-destination and mail-without-approval flow, and a server load capacity dashboard.

└──▷ **GET THIS VERSION** ▾

Docker

```
$ docker pull ghcr.io/metadist/synaplan:5.9.0
```

git

```
$ git clone --branch v5.9.0 https://github.com/metadist/synaplan.git
# already have the repo? check out this version:
$ git checkout v5.9.0
```

published under the project's GitHub account · checked 26 Nov Wrong command?

Adds a task flow that asks where the result goes and mails it without requiring approval. Not checked yet

Adds chat-run cost limits and loop-cycle detection for chat runs (flag off). Not checked yet

Adds source verification, checking a run's answer against its sources before delivery (flag off). Not checked yet

Adds ability to route a streamed chat turn to a run on the loop (flag off). Not checked yet

Adds ability to run a chat turn as a run on the run kernel (flag off). Not checked yet

+3 more

WAS THIS USEFUL? 👍 👎

synthadoc

Feedback ⓘ v1.5.4 NOTES ↗ RANK

Synthadoc: An open-source LLM knowledge compilation engine that turns raw documents into structured, local-first wikis. A transparent, human-readable alternative to traditional RAG, which can be self-managed and self-improved without the use of any tools.

> Synthadoc v1.5.4 hardens URL ingestion against SSRF and adds broad secret redaction across logs and APIs, alongside several Ollama agent configuration fixes.

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

01 SSRF guard for URL ingest

Not checked yet BREAKING 90

URL ingest and the lint URL check now refuse private, loopback and link-local addresses by default, blocking SSRF-style fetches to internal infrastructure; cloud-metadata addresses and local wiki ports remain blocked regardless. Set `allow_private_urls = true` in `[server]` or `SYNTHADOC_ALLOW_PRIVATE_URLS=1` to restore access to intranet wikis or `http://localhost` documentation.

Allow URL ingest from an intranet wiki while keeping cloud-metadata addresses blocked — set this in your server config after upgrading.

ini

```
[server]
allow_private_urls = true
```

— Full before/after with exact override config and env var. v1.5.4

02 Ollama agent configuration options

Not checked yet IMPROVED 85

Synthadoc now honours `agents.default.base_url` for Ollama so a Docker sidecar pointed at `http://ollama:11434` is actually used, adds optional `max_num_ctx` to cap context size for GPUs with limited VRAM, supports `thinking = "disabled"` to skip reasoning for faster responses, and makes `--timeout` bound the full HTTP call instead of dying at a hard-coded 300s.

Run a local Ollama model with thinking disabled and a generous timeout to keep queries practical on CPU-only hardware.

```ini
[agents]
[agents.default]
base_url = "http://ollama:11434"
max_num_ctx = 8192
thinking = "disabled"
```

— Names all four config keys/flags with a runnable example. v1.5.4

---

03 **Secret redaction across logs and API** Not checked yet NEW

80

Redacts secrets — userinfo in URLs, Authorization and Cookie headers, Bearer tokens and secret-looking query parameters — from console and file logs, job errors, scheduler output, and `/jobs` and `/audit/*` API responses, including jobs stored by older versions.

— Names exact redacted fields and affected API endpoints. v1.5.4

---

04 **Download size cap for URL fetches** Not checked yet NEW

70

URL fetch responses are now capped via `[ingest] max_download_bytes` (default `26214400`), limiting what a URL fetch can download after decoding.

— Names config key and default but no usage example. v1.5.4

└──▷ BREAKING ON UPGRADE

! URL ingest and the lint URL check now refuse private, loopback and link-local addresses by default; set `allow_private_urls = true` in `[server]` or `SYNTHADOC_ALLOW_PRIVATE_URLS=1` to restore access to intranet wikis or `http://localhost` documentation (cloud-metadata addresses and local wiki ports remain blocked regardless).

WAS THIS USEFUL? 👍 👎

stashbase

Feedback ⓘ v2.19.0 NOTES ↗ RANK

stashbase is an open-source, local-first writing workspace where you write with Claude Code and Codex, drawing on your own sources.

> stashbase's only change this window is a small usability tweak: you can now close document tabs with a middle click.

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

01 **Middle-click to close document tabs** Not checked yet NEW

45

In the documents view, you can close a document tab by middle-clicking it, without reaching for a close button.

Quickly dismiss a document tab without moving to a close button — middle-click the tab to close it.

📍 In the documents view, click a tab with the middle mouse button to close it.

— Clear UI action but no further mechanism detail v2.19.0

WAS THIS USEFUL? 👍 👎

infino

Feedback ⓘ v0.12.0 NOTES ↗ RANK

Infino is a search library built on Parquet. Fast, efficient, and scalable.

> Infino's v0.12.0 release focuses on query performance: a new keyword analyzer with term-index-powered cardinality queries, several I/O-reduction optimizations across vector, full-text, and SQL search paths, and new operational controls for hosted request deadlines and cache behavior.

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

01 **Keyword analyzer with term-index aggregation** Not checked yet

NEW 58

Adds a keyword analyzer enabling exact match, grouping, and aggregation on keyword columns. Value counts for a keyword column are read directly from the term index, enabling fast cardinality queries without a full column scan.

— Describes mechanism but no config key or API surface given v0.12.0

02 **Selective SQL wide-column decoding** Not checked yet IMPROVED

50

SQL wide-column decoding now applies only to rows selected by a filtered `LIMIT` or a top-k answer, reducing decode overhead.

— Names `LIMIT` and top-k scope but no benchmark v0.12.0

03 **Vector search file pruning via WHERE clause** Not checked yet

IMPROVED 43

Vector search now skips files ruled out by a `WHERE` clause, reducing unnecessary I/O on filtered vector queries.

— Clear mechanism but no concrete numbers or syntax v0.12.0

04 **Predicate evaluation inside full-text search pass** Not checked yet

IMPROVED 43

Full-text search now evaluates `WHERE` predicates inside the search pass, scoring best-matching rows first before filtering.

— Names mechanism but lacks measurable impact v0.12.0

05 **Superfile pruning for ranked and exact-match search**

Not checked yet IMPROVED 43

Ranked and exact-match full-text search now skips superfiles that cannot contribute results, reducing read amplification on large tables.

— Names mechanism and scope but no measured improvement v0.12.0

06 **Disk cache eviction logging** Not checked yet IMPROVED 42

Disk cache now logs when each evicted file was first cached and last used, giving operators visibility into cache churn.

— Clear operational benefit but no log format or field names v0.12.0

07 **Programmatic decoded-column cache sizing** Not checked yet NEW 40

Embedders can now size the decoded-column cache in code, allowing programmatic control over memory use per embedding workload.

— No API name or default size specified v0.12.0

08 **Deadline bound on remote hosted requests** Not checked yet NEW 38

Remote hosted requests are now bounded by a total deadline, preventing runaway query execution on the hosted path.

— No config key or default value for the deadline given v0.12.0

WAS THIS USEFUL? 👍 👎

redis-vl-python

Feedback ⓘ v0.33.0 NOTES ↗ RANK

Redis-vl-python redis Vector Library. The AI-native Python client for Redis.

> redis-vl-python v0.33.0 adds OpenTelemetry tracing for managed LangCache, fixes Redis Cluster index listing, and tightens validation on SVS-VAMANA vector fields, alongside two breaking changes affecting JSON index queries and Cohere embedding caches.

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

01 **Breaking: JSON index FilterQuery id field change** Not checked yet

88

BREAKING

On a JSON index, `FilterQuery` without `return_fields` now returns the full Redis key in `result['id']` instead of the document's own `id` field. Applications that relied on reading the stored document ID from `result['id']` must add `'$.id'` to `return_fields` and read `result['$.id']`, or use index.fetch().

— Exact migration path with fields and method given v0.33.0

02 **SVS-VAMANA training_threshold validation** Not checked yet IMPROVED

75

SearchIndex.from_existing() now preserves a custom `training_threshold` on SVS-VAMANA vector fields, and raises `ValueError` at schema-build time if `training_threshold` is set without `compression`.

— Named method, field, and exact validation condition v0.33.0

03 **Breaking: Cohere embedding cache corruption fix** Not checked yet

70

BREAKING

If you used `CohereTextVectorizer` with a dtype such as `'INT8'` or a `VectorDataType` enum member together with `EmbeddingsCache`, the cached embeddings may be all-zero vectors. Clear the affected cache entries for that model after upgrading so RedisVL can regenerate them.

— Named classes and dtype with a clear remediation step v0.33.0

04 **Fixed `rvl index listall` on Redis Cluster** Not checked yet

IMPROVED 65

Running `rvl index listall` on Redis Cluster previously failed; this is now fixed so index operations work correctly against clustered Redis.

— Exact command named and runnable, with before/after v0.33.0

05 **Fixed MCP server startup timeout double-counting** Not checked yet

IMPROVED 65

The MCP server now applies `runtime.startup_timeout_seconds` once across both the Redis connection and the initial `INFO` call, so a slow server no longer consumes the full timeout twice.

— Config key named with clear before/after behaviour v0.33.0

THINNER COVERAGE BELOW

06 **OpenTelemetry tracing for LangCache semantic cache**

Not checked yet NEW 50

RedisVL now tracks hit rate and latency for `LangCacheSemanticCache` through its existing OpenTelemetry instrumentation, covering both sync and async lookups, stores, and deletions.

— Names the cache class and coverage but no setup steps v0.33.0

└──▷ **BREAKING ON UPGRADE**

! On a JSON index, `FilterQuery` without `return_fields` now returns the full Redis key in `result['id']` instead of the document's own `id` field. Applications that relied on reading the stored document ID from `result['id']` must add `'$.id'` to `return_fields` and read `result['$.id']`, or use index.fetch().

! If you used `CohereTextVectorizer` with a dtype such as `'INT8'` or a `VectorDataType` enum member and `EmbeddingsCache`, the cached embeddings may be all-zero vectors. Clear the affected cache entries for that model after upgrading so RedisVL can regenerate them.

WAS THIS USEFUL? 👍 👎

rabitq-library

Feedback ⓘ v0.9.0 NOTES ↗ RANK

Official library for RaBitQ vector quantization and approximate nearest-neighbor search, in C++ and Python.

> rabitq-library v0.9.0 introduces three new index types — IVF, HNSW, and SymphonyQG — with matching Python bindings, alongside a new on-disk index format and a set of breaking API cleanups.

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

01 **New IVF, HNSW, and SymphonyQG index classes** Not checked yet NEW

75

Build, search, and save/load indexes from C++ using the new `rabitqlib::IvfIndex`, `rabitqlib::HnswIndex`, and `rabitqlib::SymqgIndex` classes, each with a matching Python binding.

Build and query an IVF-RaBitQ index in Python using the new `IvfIndex` class added in v0.9.0.

python

```python
import numpy as np
from rabitqlib import IvfIndex

rng = np.random.default_rng(42)
data = rng.standard_normal((500, 64)).astype(np.float32)
queries = rng.standard_normal((5, 64)).astype(np.float32)

index = IvfIndex(64, 5, metric="l2")
index.add(data)
ids, distances = index.search(queries, k=10, nprobe=5)
print(ids.shape, distances.shape)  # (5, 10) (5, 10)
```

— Names three classes with a runnable Python example v0.9.0

02 **API cleanup: removed deprecated forms, metric name, tools.hpp**

Not checked yet BREAKING

65

The Python forms deprecated in 0.7 and 0.8 are removed; only the 0.9 names (e.g. `capacity`, `routing`, `ef_search`, k-means `nbits`) are accepted. The inner-product metric name is now accepted only as `"ip"`; other spellings are rejected. The evaluation helpers previously in `tools.hpp` are removed.

— Names exact flags and files removed, clear migration note v0.9.0

THINNER COVERAGE BELOW

03 **Index format 2 on-disk format** Not checked yet NEW 35

Indexes now save and load in index format 2, a new on-disk format applied across all index types (IVF, HNSW, SymphonyQG).

– Named but no mechanism or command shown v0.9.0

└──▷ **BREAKING ON UPGRADE**

! The Python forms deprecated in 0.7 and 0.8 are removed; only the 0.9 names (e.g. `capacity`, `routing`, `ef_search`, k-means `nbits`) are accepted.

! The inner-product metric name is now accepted only as `"ip"`; other spellings are rejected.

! The evaluation helpers previously in `tools.hpp` are removed.

WAS THIS USEFUL? 👍 👎

◆ **Document Parsing**

Docling

Feedback ⓘ v2.153.0 NOTES ↗ RANK

Docling parses PDFs, Office files, and HTML into structured documents with layout, tables, and reading order preserved for downstream RAG pipelines.

> Docling v2.153.0 adds a single new option for controlling PDF character normalization during parsing.

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

01 **Character normalization control in PDF backend** Not checked yet NEW 55

A new `do_char_normalization` option in `PdfBackendOptions` exposes the underlying docling-parse setting, giving callers direct control over PDF character handling.

– Names exact config key and class but no behavior detail v2.153.0

WAS THIS USEFUL? 👍 👎

kordoc

Feedback ⓘ v4.25.19 NOTES ↗ RANK

Kordoc is a 한국 공문서를 AI가 읽고 쓰게 합니다. HWP·HWPX·PDF·오피스·스캔 이미지를 Markdown으로, Markdown을 HWPX 공문서로 바꾸고 양식 채우기·신구대조·서식 보존 편집까지.

> Kordoc's release window centers on ML-based PII redaction for Korean names and addresses, alongside a cluster of OCR output enrichments (per-character confidence, per-line candidates, page image export, per-cell table coordinates, and perspective dewarp) plus smarter PDF word-joining across line breaks.

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

01 Per-character OCR confidence and alternatives

Not checked yet

NEW

95

Get per-character OCR confidence scores, bounding boxes, and up to `n` alternative characters per position (n: 1–5) by passing `--ocr-chars [n]` on the CLI, or `ocrChars` (JS), `ocr_chars` (Python), ocrChars(...) (Java). Each character returns as `{ text, box, confidence }` under `chars`, with `alternatives: [{ text, confidence }]` when `n` is supplied.

Scan a document and emit per-character confidence with the top-3 alternative readings at each position — useful for auditing low-confidence OCR regions in scanned government forms.

```
$ npx kordoc 스캔본.pdf --ocr --ocr-chars 3 --format json -o 스캔본.json
```

— Runnable command plus exact field/flag names across SDKs v4.25.17

02 ML-based PII detection for redaction

Not checked yet NEW

90

Adds model-driven detection of names and addresses using the Hugging Face model `schift-io/schift-ko-pii-ax-v1` (int8 151 MB, q4 81 MB), downloaded on demand with a pinned revision and SHA-256 verification. Exposed via CLI `redact --model [int8|q4]`, the MCP `use_model` tool, and the library's loadPiiModel(); requires the `onnxruntime-node` peer dependency.

— Names model, sizes, flags and dependency; no worked example v4.25.5

03 Per-page dewarp correction for OCR

Not checked yet NEW

88

Adds per-line vertical-shift correction for perspective-distorted OCR pages, exposed as `dewarp: { cx, knots }` on `ocrLines` page output, with `knots` as `[y(px),`

`θ(°)]` pairs and `deskewAngle` set to 0 on affected pages. Accessible via `result.ocr_dewarp` in the Python SDK and result.ocrDewarp() in the Java SDK, both returning empty/None for pages not dewarped.

— Exact field names and SDK accessors but no sample command v4.25.19

---

04 **Page image export from OCR pipeline** Not checked yet NEW

75

Export the page images read by the OCR path as `pageImages` in PNG or WebP format using `--page-images png|webp` on the CLI or the `pageImages` API setting; images are produced even for non-OCR pages and when `ocr: false`, serving as the coordinate reference for `ocrLines`.

— Flag, formats and behaviour named, no runnable sample v4.25.4

---

05 **Per-line OCR candidate scores** Not checked yet NEW 70

Get the top N recognition candidates per OCR line, including confidence scores, via `--ocr-line-candidates <n>` (1–5) on the CLI or `ocrLineCandidates` in the API; each line in `ocrLines` gains a `candidates: [{ text, probability }]` array.

— Flag and output field named, no worked example given v4.25.4

---

06 **Per-cell bounding boxes in table OCR** Not checked yet NEW

65

Adds `IRCell.bbox` to OCR table results, giving per-cell coordinates in the same PDF-pt (bottom-origin) coordinate space as the table block `bbox`, removing the need to manually re-align OCR lines to cells.

— Named field and coordinate space but no example usage v4.25.8

THINNER COVERAGE BELOW

---

07 **Smarter Korean word-joining across line breaks** Not checked yet

IMPROVED 53

PDF parsing now uses bracket-attached and postposition-attached word stems as in-document evidence when joining Korean words split across line breaks, in addition to exact word matches — e.g. joining '국제⏎에너지기구' when only `국제에너지기구(IEA)는` or '휴머노이드를' appear in the document. Not applied when spaced neighbors exist, and not applied to column text.

— Clear mechanism and example but automatic, nothing to configure v4.25.11

WAS THIS USEFUL?  

anymd

Feedback ⓘ v8.5.17 NOTES ↗ RANK

anymd converts any file to clean Markdown for AI agents: PDF, Word, PowerPoint, Excel, EPUB, HTML and web pages.

> anymd introduced a standalone `anymd-core` Rust library for embedding document conversion without the server binary, alongside small usability improvements to the documentation site and Studio results.

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

01 **Standalone anymd-core Rust library** Not checked yet NEW

78

A new `anymd-core` library lets Rust, WebAssembly, and mobile apps share document conversion logic without depending on the server binary. It exports `anymd_core::outline`, and with the optional `formats` feature flag also exposes `anymd_core::document` with `open_bytes` and `scanned_pages`.

— Names exact modules and functions but no usage example v8.5.17

THINNER COVERAGE BELOW

02 **Copy, download and stats in docs and Studio** Not checked yet

IMPROVED

42

Code blocks in `/docs` can now be copied directly, and Studio results display copy and download buttons along with document stats at the top.

— Describes UI additions but lacks further mechanism detail v8.5.17

WAS THIS USEFUL? 👍 👎

Unstructured

Feedback ⓘ 0.27.38 NOTES ↗ RANK

Unstructured is an open-source ETL library that converts PDFs, images, and office documents into chunked, structured elements for LLM ingestion.

> Unstructured sped up hi_res PDF processing by eliminating redundant page rendering across pipeline stages.

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

01 **Single-render hi_res PDF pipeline** Not checked yet IMPROVED

45

The `hi_res` PDF pipeline now renders each page once and reuses the result across all pipeline stages, instead of re-rendering per stage, reducing processing time for multi-stage PDF partitioning.

— Explains mechanism but no benchmark or usage detail 0.27.38

WAS THIS USEFUL? 👍 👎

EVALUATE

◆ AI Observability & Evals

LangChain LangSmith

Feedback ⓘ updated · Nov 26, 2026

NOTES ↗ RANK

LangSmith provides tracing, evaluation, and deployment tools for LLM applications.

> LangSmith's latest snapshot adds gateway routing visibility via a new response header and trace UI, IAM-role authentication for Bedrock in the LLM Gateway, richer thread-based annotation automations, three new GPT-5.6 chat models, sandbox resource history and WebSocket fixes, and deprecates the Turns view in thread detail.

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

01 **Gateway routing visibility in traces** Not checked yet NEW

84

LangSmith adds an `X-LangSmith-Gateway-Metadata` response header describing gateway routing, limits, guard activity, and upstream errors, which clients can attach to their own traces as `ls_gateway_info` metadata. Trace details now show a Gateway section with LLM Gateway outcomes, policy limits, selected models, and fallback attempts, plus a Blocked badge in the trace tree for runs the gateway blocked.

— Names the exact header and metadata key plus UI elements. snapshot-20261126

02 **CSV export for colon-containing dataset keys** Not checked yet

IMPROVED

60

Datasets whose input, output, or metadata keys contain a colon — including those with `source_run_feedback:<key>` metadata keys — can now be exported to CSV.

— Names the exact metadata key pattern fixed. snapshot-20261126

THINNER COVERAGE BELOW

03 **Turns view deprecated in thread detail** Not checked yet DEPRECATED

55

The Turns view in thread detail is deprecated and will be removed from LangSmith Cloud on October 31, 2026, and from self-hosted in v17; use the Messages or Details view instead.

— Gives exact removal dates/version and migration path. snapshot-20261126

04 **WebSocket query strings preserved through sandbox URLs**

Not checked yet IMPROVED

50

WebSocket connections through a sandbox service URL now reach the service with their query string intact, allowing dev servers that authenticate WebSocket connections with a query parameter (such as Vite's HMR token) to connect.

— Gives a concrete example (Vite HMR token) but no config detail. snapshot-20261126

05 **Tag visibility for tag-on-create permission** Not checked yet

IMPROVED

45

Users with the tag-on-create permission for datasets, projects, or prompts can now see and apply tags when creating those resources.

— Names the permission and resource types affected. snapshot-20261126

06 **API Keys page search across full list** Not checked yet IMPROVED

45

The API Keys page now filters keys across the full paginated list by key, description, and workspace.

— Clear UI path and filter fields named, no further mechanism. snapshot-20261126

07 **New GPT-5.6 models in LangSmith Chat** Not checked yet NEW

35

LangSmith Chat now offers GPT-5.6 Sol, GPT-5.6 Terra, and GPT-5.6 Luna alongside the existing OpenAI models.

— Names the three models but no further detail. snapshot-20261126

08 **Sandbox resource usage history** Not checked yet NEW 35

Sandboxes now retain recent CPU, memory, and network usage history, making it easier to understand resource consumption and diagnose performance bottlenecks.

— States the capability but no retention window or UI path. snapshot-20261126

09 **Dual-stack IPv4/IPv6 connectivity** Not checked yet NEW

25

LangSmith services now accept both IPv4 and IPv6 connections.

— Bare statement with no scope or configuration detail. snapshot-20261126

└──▷ **BREAKING ON UPGRADE**

! The Turns view in thread detail is deprecated and will be removed from LangSmith Cloud on October 31, 2026, and from self-hosted in v17; use the Messages or Details view instead.

WAS THIS USEFUL? 👍 👎

Arize Phoenix

Feedback ⓘ arize-phoenix-v22.1.0 NOTES ↗ RANK

Arize Phoenix is an open-source platform that monitors LLM application traces and evaluates their outputs.

> Arize Phoenix's latest release adds a single-span REST endpoint, two new CLI commands for project setup and diagnostics, a configurable playground timeout, and changes dataset example deletion to be permanent.

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

01 **Single span fetch REST endpoint** Not checked yet NEW 80

A new `GET /spans/{span_identifier}` REST API endpoint lets callers fetch a single span by identifier.

– Exact endpoint named but no usage example given arize-phoenix-v22.1.0

02 **Configurable playground timeout** Not checked yet NEW 80

The playground LLM task timeout can now be configured using the `PHOENIX_PLAYGROUND_TIMEOUT_SECONDS` environment variable, useful for replaying long-running LLM calls that would otherwise time out.

Raise the playground timeout when replaying long-running LLM calls that otherwise time out.

```
$ export PHOENIX_PLAYGROUND_TIMEOUT_SECONDS=120
```

– Exact env var named with a runnable example arize-phoenix-v22.1.0

03 **New CLI commands for project setup and diagnostics**

Not checked yet NEW 70

Two new CLI commands were added: `px project create` to create a Phoenix project from the command line, and `px setup doctor` to diagnose connection problems from the command line.

Create a new Phoenix project from the CLI during pipeline setup or onboarding.

```
$ px project create
```

Run the connection doctor before debugging a tracing issue to surface misconfigurations fast.

```
$ px setup doctor
```

— Both commands named with runnable examples `arize-phoenix-v22.1.0`

THINNER COVERAGE BELOW

04 **Permanent deletion of dataset examples** Not checked yet BREAKING

35

Dataset examples are now permanently deleted rather than soft-deleted when removed.

— Behavior change noted but no migration guidance given `arize-phoenix-v22.1.0`

05 **Model provider listing helper** Not checked yet NEW

20

A new helper in the client library lists available model providers.

— No method name or usage detail given `arize-phoenix-v22.1.0`

WAS THIS USEFUL? 👍 👎

NVIDIA SkillEvaluator

Feedback ⓘ v1.2.1 NOTES ↗ RANK

SkillEvaluator checks AI agent skills for quality and overlap, then runs coding agents on the same tasks with and without each skill to measure its effect.

> Across three releases, SkillEvaluator added paired-case lift statistics and a locked CI container for Tier 2/3 evaluation, new CLI surfaces for hook waivers, CI setup, grader auditing, and skill-version comparison, and a run of breaking changes to scanning, dataset grounding defaults, and Tier 3 grading outcomes.

└──▷ **WHAT SHIPPED · 19 FEATURES** most completely described first ? what's the number?

01 **Paired-case lift statistics for Tier 3 runs** Not checked yet NEW

84

Tier 3 skill runs now write a paired-case lift estimate with a 95% bootstrap interval, a precision label, and a zero-interval warning to `<agent>/statistics.json`, the report JSON, Markdown and HTML reports, and the full terminal output of `validate` and `tier3 evaluate --verbose`.

— Names exact output files, formats and commands. v1.2.1

02 **Fixed Bandit config and bounded YAML/JSON parsing** Not checked yet

75

BREAKING

Skill-owned Bandit configuration and `# nosec` comments no longer suppress findings — Bandit now runs with a fixed configuration controlled by SkillEvaluator. YAML lists and mappings are limited to 1,024 items, so skills relying on larger YAML structures are rejected; scanner baselines affected by either change must be regenerated after upgrading.

— Clear before/after and regeneration guidance, no exact command. v1.2.0

03 **OpenCode image build tools removed** Not checked yet

NEW SUPPORTED PLATFORM BREAKING

75

Generated Debian-based OpenCode images no longer include `gcc`, `g++`, or `make`; skills or tasks needing them must declare packages in `scripts/apt-packages.txt` or a custom evaluation Dockerfile.

— Exact packages removed and remediation file named. v0.9.0

04 **Richer evaluation reports with SARIF findings** Not checked yet

IMPROVED

73

Evaluation reports now show met, unmet, unchecked, and inapplicable expectations; dataset coverage and sufficiency; skill activation; failed-arm coverage; and concrete next steps; HTML report charts now work offline; and dependency advisories now appear as structured findings in reports and `SARIF`, with stable finding IDs to track issues across runs.

— Many concrete fields and SARIF named, no single command. v0.9.0

05 **Opt-in GitHub/GitLab CI integration** Not checked yet NEW

72

Adds opt-in CI integration across GitHub Actions and GitLab CI: a locked Tier 2 CI container with credential-free build verification and opt-in digest-pinned jobs (documented in the Repository Integration guide alongside Tier 1), opt-in stable CI channels that automatically follow promoted Tier 1 releases once a release is verified, and a `skillevaluator ci init` command that sets up CI integration including advisory Tier 2 and Tier 3 evaluations alongside required Tier 1 checks.

— Describes container scope but no exact CLI entrypoint given. Mechanism described but no exact config key shown. Exact command and scope of checks given.

v1.2.1 v1.2.0 v0.9.0

06 **Custom headers for OpenAI-compatible providers** Not checked yet

NEW

70

Set custom HTTP headers for OpenAI-compatible LLM providers with the `SKILL_EVAL_LLM_EXTRA_HEADERS` environment variable, such as for a corporate proxy requiring a custom key.

Pass extra authentication headers to an OpenAI-compatible provider endpoint, such as a corporate proxy requiring a custom key.

```
$ export SKILL_EVAL_LLM_EXTRA_HEADERS='{"X-Custom-Auth": "my-token"}'
skillevaluator validate ./my-skill
```

— Env var and runnable example both given. v1.2.0

07 **Dataset grounding mode control** Not checked yet BREAKING

69

Override dataset grounding mode with `--grounding skill` or `--no-llm` for skill-only starter flows that do not need LLM generation. Dataset generation now defaults to `auto` grounding, so skill-only starter flows must add `--grounding skill` (including when using `--no-llm`) or generation will fail.

— Flags named plus the breaking default-change explained. v0.9.0

08 **GitHub Actions annotation reporter** Not checked yet NEW

67

Export GitHub Actions annotations by passing `-r github` to the reporter, so CI surfaces finding locations inline on pull requests.

Run Tier 1 checks in GitHub Actions and emit inline annotations on the pull request diff.

```
$ skillevaluator validate ./my-skill --tiers 1 -r github
```

– Exact flag and runnable example command provided. v1.2.0

09 **Tier 3 grading outcome changes** Not checked yet BREAKING

65

Some grading outcomes change: accuracy evaluates task outcomes without penalizing a no-skill baseline, standard judge scores are computed from verdicts, a scored attempt fails when goal accuracy or security is below 1.0, and a scored dimension below 0.40 fails the Tier 3 verdict. Re-evaluate before comparing scores with older releases.

– Precise thresholds given, no exact flag or command. v0.9.0

10 **Tier 3 grader auditing and feedback subcommands** Not checked yet NEW

62

The `tier3 validate-grader` subcommand qualifies the shipped verifier and audits grading before publishing results, while `tier3 feedback` lets you ask questions about a completed run for post-run interrogation.

– Two subcommands named with purpose, no examples. v0.9.0

11 **Optional installs for Tier 2/3 and dashboard** Not checked yet NEW

62

Adds optional installs via `skillevaluator[all]` for Tier 2, Tier 3, and Python security dependencies, and a separate `dashboard` extra for installing the optional dashboard.

– Exact extras syntax named for each tier. v0.9.0

12 **Reviewed command hook waivers** Not checked yet NEW

60

Approve eligible command hooks in a reviewed package with the `--hook-waivers` flag, leaving other findings and coverage gaps intact.

– Flag named but no example usage shown. v1.2.0

THINNER COVERAGE BELOW

13 **Deprecated Tier 3 required-gate settings removed** Not checked yet

DEPRECATED

58

The deprecated `--tier3-required`, `tier3-required`, and `SKILL_EVAL_TIER3_REQUIRED` settings no longer make Tier 3 a required CI gate.

— Exact deprecated settings named, no replacement path given. v0.9.0

14 **Skill version comparison** Not checked yet NEW

57

Compare retained skill versions against each other using `--compare-to` and the `compare-versions` subcommand.

— Flag and subcommand named, no usage example. v0.9.0

15 **Machine-readable readiness checks** Not checked yet NEW

53

Get machine-readable readiness output from `doctor --json` for scripted environment checks.

— Command named but output shape unspecified. v1.2.0

16 **Tolerate baseline execution timeouts** Not checked yet NEW

53

Tolerate narrowly qualified baseline execution timeouts by opting in with `--tolerate-baseline-timeouts`.

— Flag named but scope of 'narrowly qualified' unspecified. v0.9.0

17 **Python catalog rebuild replaces selection** Not checked yet

47

BREAKING

Python catalog rebuilds replace the selection: callers reusing a registry must supply the complete intended catalog, as successful builds no longer merge into the previous selection.

— Behavior change described, no concrete command shown. v0.9.0

18 **Experimental Hermes agent support** Not checked yet NEW

35

Adds experimental Hermes agent support with Docker and native NVIDIA Build routing.

— Marked experimental with little further detail. v0.9.0

19 **Custom metrics import with provenance** Not checked yet NEW

32

Import externally computed custom metrics with their provenance into evaluation reports.

— No mechanism, format or command named. v0.9.0

NEW INTEGRATION NeMo Gym execution backend

└──▷ **BREAKING ON UPGRADE**

! Skill-owned Bandit configuration and `# nosec` comments no longer suppress findings; Bandit now runs with a fixed configuration controlled by SkillEvaluator.

! YAML lists and mappings are limited to 1,024 items; skills relying on larger YAML structures will now be rejected.

! Scanner baselines affected by the new bounded YAML/JSON parsing or Bandit configuration changes must be regenerated after upgrading.

! Dataset generation now defaults to `auto` grounding; skill-only starter flows must add `--grounding skill` (including when using `--no-llm`) or generation will fail.

! Some grading outcomes change: accuracy evaluates task outcomes without penalizing a no-skill baseline, standard judge scores are computed from verdicts, a scored attempt fails when goal accuracy or security is below 1.0, and a scored dimension below 0.40 fails the Tier 3 verdict. Re-evaluate before comparing scores with older releases.

! Generated Debian-based OpenCode images no longer include `gcc`, `g++`, or `make`; skills or tasks needing them must declare packages in `scripts/apt-packages.txt` or a custom evaluation Dockerfile.

! The deprecated `--tier3-required`, `tier3-required`, and `SKILL_EVAL_TIER3_REQUIRED` settings no longer make Tier 3 a required CI gate.

! Python catalog rebuilds replace the selection: callers reusing a registry must supply the complete intended catalog, as successful builds no longer merge into the previous selection.

v1.0.1 NOTES ↗ STABLE

SkillEvaluator v1.0.1 adds HarnessCheck, NeMo Gym backend, JUnit XML output, pre-commit hooks, reusable PR-comment workflow, and `--quality-config` / `--profile registry` controls.

└──▷ **GET THIS VERSION** ▾

pipx

```
$ pipx install skillevaluator==1.0.1
# already installed? upgrade to this version:
$ pipx install --force skillevaluator==1.0.1
```

```git
$ git clone --branch v1.0.1
https://github.com/NVIDIA/SkillEvaluator.git
# already have the repo? check out this version:
$ git checkout v1.0.1
```

named in the project's README · checked 26 Nov Wrong command?

Run `skillevaluator harness-check ./my-skill` to find compatibility issues — missed discovery paths, frontmatter rejections, description truncation, invocation fields, plugin formats, and conflicts — across Claude Code, Codex, Copilot CLI, Hermes, OpenClaw, and OpenCode, with no model key required. Not checked yet

Run Tier 3 evaluations against NeMo Gym using `--backend nemo-gym`, sharing the existing grading and reporting pipeline. Not checked yet

Export JUnit XML from any evaluation run with `-r junit` for CI test-result ingestion. Not checked yet

Add offline pre-commit checks to a repository via the new pre-commit hook integration. Not checked yet

Use the reusable GitHub workflow to post scored PR comments without hosted reports. Not checked yet

+3 more

└─▷ **BREAKING ON UPGRADE**

! Existing CI workflow and image pins must be updated separately from the Python package when upgrading from v0.9.0; review the author-policy, security-override, and hook-gate changes in the changelog.

WAS THIS USEFUL? 👍 👎

Superlog Labs Superlog

Feedback ⓘ

changes since 2026-10-27 CODE ↗

Superlog monitors AI agent telemetry with OpenTelemetry and groups signals into incidents for observability.

Superlog adds org-level monthly spend limits, MCP request forwarding for migrated projects, and ClickHouse ingest pacing controls

└──▷ **USE IT**

Slow down ClickHouse part creation on a quorum-write cluster by enforcing a minimum interval between inserts, so batches grow instead of flooding the replication queue.

.env

```
INGEST_CLICKHOUSE_MIN_INSERT_INTERVAL_MS=500
```

Spread ClickHouse inserts evenly across replicas behind a TCP load balancer by disabling keep-alive so each insert opens a fresh connection.

.env

```
INGEST_CLICKHOUSE_KEEP_ALIVE=false
```

Control ClickHouse insert pacing by setting `INGEST_CLICKHOUSE_MIN_INSERT_INTERVAL_MS` to enforce a minimum gap between inserts into the same table per worker, growing batch size instead of part count when inserts accelerate. Not checked yet

Disable ClickHouse keep-alive connections by setting `INGEST_CLICKHOUSE_KEEP_ALIVE=false` to open one connection per insert and spread load evenly across replicas instead of accumulating on the slowest one. Not checked yet

Superlog now forwards MCP requests for projects that have cut over to a workspace, sending them to `${workspace_url}/api/mcp` and streaming the response back, with no MCP client reconfiguration needed. Not checked yet

Superlog adds a `project_migrations` table (fields: `project_id`, `workspace_url`, `cut_over_at`, `created_at`) to track project cut-over state, with `isProjectCutOver` and `cutOverWorkspaceUrl` exported from `@superlog/db`. Not checked yet

Superlog adds a GCP Pub/Sub connection cache via `createGcpConnectionCache` that holds connection lookups for 60 seconds and caches unknown or non-ingestible connections for 30 seconds, keeping busy GCP push subscriptions off Postgres. Not checked yet

+1 more

WAS THIS USEFUL? 👍 👎

Comet Opik

Feedback ⓘ 2.3.4 NOTES ↗ RANK

Opik traces, evaluates, and monitors LLM applications and agentic workflows, with datasets, an LLM-as-judge metric library, and production dashboards.

> Opik shipped a public SQL query endpoint, new guardrail model and topic APIs with vector backfill, cursor-streamed dataset retrieval for experiments, and an additional selectable model for Decision model rules.

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

01 **Guardrail models API and training-run tracking** Not checked yet

NEW

65

Opik adds a guardrail models REST API to the backend for managing guardrail model configuration, along with guardrail models and training-run tables that include a `finished_at` field and a (model_id, created_at) index on training runs.

– Names table fields and index but no endpoint paths 2.3.1 2.3.2

02 **Topics API and vector backfill support** Not checked yet NEW

62

Opik adds a Topics read API with count support for querying topic data, Topics vector tick support with long-window and backfill capabilities for historical topic vector data, and a Topics rebuild service with cleanup hooks for maintaining topic vector state.

– Names three Topics capabilities but no concrete endpoints or flags 2.3.2

THINNER COVERAGE BELOW

03 **Public free-form SQL run endpoint** Not checked yet NEW

50

A new public endpoint lets users run free-form SQL queries against Opik data, gated behind its own feature flag.

– Names endpoint and flag but no path or flag name given 2.3.4

04 **GPT-6 Luna model in Decision rules** Not checked yet NEW

35

Opik adds GPT-6 Luna as a selectable model in Decision model rules.

– Bare addition with model name, no further detail 2.3.1

WAS THIS USEFUL? 👍 👎

Langfuse

Feedback ⓘ v4.86.0 NOTES ↗ RANK

Langfuse provides tracing, evaluation, and monitoring for LLM applications.

> Langfuse's latest releases add inline interactive chart rendering inside MCP Apps, a trace-level transcript evaluation UI, and move Topics to public beta for self-hosted instances, alongside organization-level model management and dashboard filtering improvements.

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

01 **Trace evaluation enhancements** Not checked yet NEW 50

Adds a trace evaluation UI that lets users review a complete conversation transcript at the trace level and run or inspect evaluations across the full context, plus server-side trace evaluator testing and trace rule management.

Review a complete conversation transcript at the trace level to run and inspect evaluations across the full context.

> 📍 In Langfuse, go to Evals › select a trace › open the Transcript tab to view trace-level transcript evaluations.

– UI navigation path given via usage example Bare feature name with no mechanism v4.86.0 v4.85.0

02 **Topics public beta on self-hosted instances** Not checked yet NEW 50

Topics and events dispatch can now be enabled on self-hosted instances, moving Topics to public beta, with cost estimation available before setup and a live cost map once enabled.

– Explains cost estimation and live map, no exact config steps v4.86.0 v4.85.0

03 **Inline interactive charts in MCP App** Not checked yet NEW 45

Charts can now be rendered inline and interactively inside an MCP App served by Next.js.

– Names MCP App and Next.js but no exact command v4.86.0

04 **Traces linked from AI Gateway tables** Not checked yet NEW 45

Traces are now shown directly from the Budgets and Gateway keys tables in the AI Gateway, letting a user jump from a budget or gateway key row to its associated filtered

trace list via a Show traces action.

Jump from a budget or gateway key directly to its associated traces to investigate AI gateway usage.

> 📍 In Langfuse, go to AI Gateway › Budgets (or Gateway Keys) › locate a row › select Show traces to open the filtered trace list.

— Exact UI path given, mechanism is thin v4.86.0

---

05 **Numeric metadata filters for dashboard v2 widgets** Not checked yet

IMPROVED

40

Dashboard observation widgets can now be filtered by numeric metadata values, for v2 widgets.

— Names v2 widgets but no config key given v4.85.0

---

06 **Self-serve enterprise trial via coupon** Not checked yet NEW

25

Users can start a self-serve enterprise trial by redeeming a coupon code.

— Thin description, no coupon flow detail v4.85.0

---

07 **Deny support-team access for listed organizations** Not checked yet

NEW

25

Support-team access can now be denied for specific listed organizations.

— Bare description with no configuration detail v4.85.0

WAS THIS USEFUL? 👍 👎

overmind

Feedback ⓘ v0.1.34 NOTES ↗ RANK

overmind is a platform for continuously improving AI agents through observability, evaluation, and fine-tuning.

> overmind added Kev decision model support for training and serving, and extended fine-tuning capabilities to Gemma 4 on Unsloth 2026.11.9.

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

01 **Kev decision models for training and serving** Not checked yet

NEW 60

overmind adds Kev decision models for training and serving, expanding the set of native decision model workers alongside the existing `modal_decision_evaluation.py` and `modal_decision_serving.py` Modal workers.

— Names the worker files but no usage steps v0.1.34

THINNER COVERAGE BELOW

02 **Gemma 4 fine-tuning on Unsloth** Not checked yet NEW 30

overmind adds support for fine-tuning Gemma 4 models using Unsloth 2026.11.9.

— Only named in summary, no detail or usage v0.1.34

WAS THIS USEFUL? 👍 👎

Braintrust

Feedback ⓘ updated · Nov 26, 2026 NOTES ↗ RANK

Braintrust provides evaluation, tracing, and improvement workflows for AI applications.

> Braintrust rolled out project-wide tracing setup for coding agents, in-browser trace translation, and a handful of CLI and Terraform provider breaking changes.

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

01 Project-scoped tracing install for coding agents

Not checked yet

NEW

90

`bt trace install --scope project` turns on tracing for all supported coding agents across a repo so every agent run in that project is traced without per-agent configuration. Tracing defaults — `project`, `tags`, and `metadata` — can be stored in a `.braintrust.json` file at the repo root so any supported coding agent picks them up automatically.

Turn on repository-wide tracing for all supported coding agents in a repo, so every agent run in that project is traced without per-agent configuration.

```
$ bt trace install --scope project
```

Store tracing defaults at the repo root so any supported coding agent picks up the project, tags, and metadata automatically.

```json
{
  "project": "my-agent-project",
  "tags": ["production"],
  "metadata": {"team": "infra"}
}
```

— Names exact command, flag and config file with its keys snapshot-20261126

02 In-browser translation of non-English traces

Not checked yet

NEW

60

An entire non-English trace — including span fields, span names, thread messages, and custom trace views — can be translated to English using the browser's built-in translation API in desktop Chrome or Edge, keeping trace text on-device without consuming Braintrust credits.

— Clear mechanism but no exact UI path given snapshot-20261126

THINNER COVERAGE BELOW

03 **Optional telemetry argument for MCP tool calls** Not checked yet

IMPROVED

50

`bt mcp run` no longer requires passing the telemetry argument when calling a Braintrust MCP server tool.

— Names the command but little mechanism detail snapshot-20261126

└──▷ **BREAKING ON UPGRADE**

! `bt trace disable` now pauses tracing and keeps the integration and its settings instead of removing tracing; use `bt trace uninstall` in scripts that previously used `bt trace disable` to remove tracing.

! `bt trace setup` is now an alias for `bt trace install` (renamed).

! The Terraform provider (v0.0.9+) requires Terraform or OpenTofu 1.11.1 or later.

WAS THIS USEFUL? 👍 👎

GOVERN

◆ AI Runtime Defense & Guardrails

Prismor

Feedback ⓘ v1.80.5 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 releases add a `--hard` flag to force-lift floor rules during a pause, a new antigravity feature for capturing tool output from transcripts, proxy-based screening for agents without a base URL, new policy guards against destructive Git operations and npm publishing, and dashboard upgrades for session attribution and per-agent secret vaulting.

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

01 **Proxy screening for agentless agents** Not checked yet NEW 80

Screen agents that have no base URL via `prismor proxy -- <agent>` (warp mode), catching misconfigured agent definitions before they reach production.

— Exact runnable command with stated purpose v1.80.4

02 **`--hard` flag for pause command** Not checked yet NEW 70

Passing `--hard` to the `pause` command lifts floor rules during a pause, and pause messages now report which rules stay active.

— Names exact flag and command, explains behaviour change v1.80.5

03 **New policy guards for destructive Git ops and npm publish**

Not checked yet NEW 60

New policy rules now guard against forced Git cleanup and stash deletion, forced local Git branch deletion, and flag `npm publish` commands for review, preventing destructive repository operations and surfacing package-publishing actions by agents.

— Names each guarded operation but no config surface v1.80.4

THINNER COVERAGE BELOW

04 **Per-agent key vault in dashboard** Not checked yet NEW 40

The dashboard adds a vault with per-agent key access, scoping secret retrieval to individual agents.

— Names vault concept but lacks setup steps v1.80.4

05 **Dashboard session attribution** Not checked yet NEW 35

The dashboard now shows which desktop session started a session and who sent each relayed message, adding per-message attribution to session views.

— Describes behaviour but no exact UI path v1.80.4

06 **Antigravity tool-output capture** Not checked yet NEW 25

The antigravity feature captures tool output from the transcript.

— Named but no mechanism or usage detail given v1.80.5

WAS THIS USEFUL? 👍 👎

FailproofAI

Feedback ⓘ v1.0.17 NOTES ↗ RANK

FailproofAI provides observability and enforcement for AI agent harnesses, capturing every run and blocking dangerous tool calls before they execute.

> FailproofAI's v1.0.17 release adds a persistent daemon log with configurable upload controls, plus new Cloud health reporting and hook-silence detection to surface enforcement gaps.

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

01 Daemon log file and upload controls

Not checked yet NEW

88

FailproofAI now writes a persistent local daemon log that can be tailed directly (e.g. `~/.failproofai/logs/daemon.log`) for triaging enforcement errors without waiting on Cloud. Uploads of this log can be disabled per machine by setting `FAILPROOFAI_NO_LOG_UPLOAD=1`, and the 'send daemon logs to failproof' setting in Cloud now controls uploads for every machine in an organization at once, with nothing queued or sent while it is off and nothing logged while off ever uploaded after it is turned back on.

Prevent the daemon from uploading any logs to Cloud on a machine where your policy prohibits telemetry.

```
$ export FAILPROOFAI_NO_LOG_UPLOAD=1
```

Inspect the new local daemon log directly to triage enforcement errors without waiting for Cloud to surface them.

```
$ tail -f ~/.failproofai/logs/daemon.log
```

— Names env var, Cloud setting, and log path with exact toggle behavior. v1.0.17

THINNER COVERAGE BELOW

02 Cloud health reporting

Not checked yet NEW

12

FailproofAI adds Cloud health reporting to surface agent enforcement gaps, per the release summary.

— Only named in summary, no mechanism or detail given. v1.0.17

03 Hook-silence detection

Not checked yet NEW

12

FailproofAI adds hook-silence detection to surface agent enforcement gaps, per the release summary.

— Only named in summary, no mechanism or detail given. v1.0.17

WAS THIS USEFUL? 👍 👎

◆ AI Red Teaming & Testing

AI-Infra-Guard

Feedback ⓘ v4.6.12 NOTES ↗

AI-Infra-Guard is an AI red teaming platform that scans AI infrastructure for CVE vulnerabilities, MCP/Agent skill risks, jailbreak weaknesses, and API relay security issues.

Agent-Scan adds sandboxed `--repo` source tools with Stage 1 profiling; Skill-Scan adds baseline-anchored semantic slicing for pure semantic files.

└──▷ GET THIS VERSION ▾

```
$ git clone --branch v4.6.12 https://github.com/Tencent/AI-Infra-Guard.git
# already have the repo? check out this version:
$ git checkout v4.6.12
```

Scan agent repositories in a sandbox with Stage 1 profiling by passing `--repo` as a sandboxed source in Agent-Scan. Not checked yet

Skill-Scan now applies baseline-anchored semantic slicing when analysing pure semantic files. Not checked yet

NEW DETECTION MCP cross-server tool shadowing detection rule

NEW DETECTION RefusalLadderJailbreaking multi-turn attack technique

WAS THIS USEFUL? 👍 👎

◆ Model & Agent Protection

ToolHive

Feedback ⓘ v0.69.1 NOTES ↗ RANK

ToolHive runs MCP servers in isolated containers and enforces identity and access policies for local and Kubernetes deployments.

> ToolHive's latest releases tighten ID-JAG token handling with per-client lifespans and authorization_details enforcement, add EdDSA token support and open-source Microsoft Entra on-behalf-of backend auth, and introduce new observability for vMCP code mode.

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

01 authorization_details claim rejected in token exchange

Not checked yet BREAKING 88

The embedded authorization server now refuses RFC 8693 token-exchange requests whose `subject_token` or `actor_token` carries an `authorization_details` claim, returning a 400 `invalid_request`, and refuses JWT-bearer or ID-JAG assertions carrying the claim with a 400 `invalid_grant` or `invalid_assertion`. Tokens issued by ToolHive itself are unaffected; previously the claim was silently dropped and the request succeeded.

— Exact endpoints, claim names and error codes given v0.69.0

THINNER COVERAGE BELOW

02 Wildcard directory patterns in .thvignore Not checked yet

IMPROVED 55

ToolHive can now match wildcard directory patterns in `.thvignore` to control which files are excluded from MCP server context.

— Names exact config file; pattern syntax not shown v0.69.1

03 Microsoft Entra on-behalf-of backend auth in OSS edition

Not checked yet NEW 54

ToolHive adds Microsoft Entra on-behalf-of (OBO) backend authentication in the open-source edition, enabling delegated identity flows to downstream APIs without an enterprise license.

— Names provider and flow but no config steps v0.69.0

04 authorization_details claim processing for tool-call authorization

Not checked yet NEW 48

ToolHive now processes the `authorization_details` claim from ID-JAG tokens to enforce fine-grained authorization on tool calls.

— Names claim and mechanism, no usage steps v0.69.1

05 **Observability for vMCP code mode** Not checked yet IMPROVED 48

ToolHive's vMCP code mode now emits OTel metrics, spans, and logs, and reports tool calls in anonymous usage metrics, making code-mode activity observable.

— Names telemetry types but no setup detail v0.69.0

06 **Per-client access token lifespan for ID-JAG redemption**

Not checked yet NEW 45

ToolHive now adds a per-client access token lifespan for ID-JAG redemption, letting each client carry an independently scoped token lifetime.

— Describes behavior but no config surface named v0.69.1

07 **EdDSA token support in token validator** Not checked yet NEW 43

ToolHive's token validator now accepts EdDSA-signed tokens, widening the set of identity providers that can authenticate to the gateway.

— Clear capability, no setup detail given v0.69.0

08 **Admission method recorded on Identity object** Not checked yet

NEW 38

ToolHive now records how a token was admitted on the Identity object, surfacing the admission method in audit and identity records.

— Describes behavior with no concrete field names v0.69.1

09 **MCP tool-call details type** Not checked yet NEW 27

ToolHive now adds the MCP tool-call details type, capturing structured detail about individual tool invocations.

— Bare name, no schema or usage given v0.69.1

└──▷ **BREAKING ON UPGRADE**

! If a trusted external identity provider puts `authorization_details` (RFC 9396) into tokens or ID-JAGs it sends to ToolHive, those requests now fail: RFC 8693 token-exchange requests return 400 `invalid_request`, and JWT-bearer or ID-JAG redemptions return 400 `invalid_grant` or `invalid_assertion`. Previously the claim was silently dropped and the request succeeded.

WAS THIS USEFUL?  

Minimal Software Research minimal Feedback ⓘ

release-7b9ec7d4 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 secrets management via `min secret`, tamper-evident box audit trails, Box Egress Proxy controls for scoping and naming outbound network access, GitHub sign-in via `min login`, and opt-in OpenTelemetry tracing backed by TPM/Secure Enclave key storage.

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

01 Box Egress Proxy (BEP) enhancements Not checked yet NEW

88

`min auth status` now names each software proxy key via the new `bep` field (T132), storing these BEP keys in the Secure Enclave or TPM rather than in plaintext; `--max-client-keys` on `bep serve` limits the Box Egress Proxy's client-key registry size (T69); `min task run` and `min box run` gain network and egress flags to scope outbound access per invocation (T55); and `min box show --network <box>` displays a box's effective networking policy (T55).

Inspect every GitHub credential and software proxy key the host holds, including BEP keys stored in the Secure Enclave or TPM.

```
$ min auth status
```

Inspect what network traffic a running box is actually permitted to make before trusting it with credentials.

```
$ min box show --network my-agent-box
```

Review the effective networking policy for a running box before trusting it with outbound credentials.

```
$ min box show --network mybox
```

— names the bep field and storage mechanism but no flag detail three named flags/commands with clear per-invocation scoping behaviour

release-7b9ec7d4 release-bdc02696

02 **Box audit trail** Not checked yet NEW 86

`min box audit` reads a box's audit trail, `min box audit --parent` merges parent chains for a full history, and `min doctor` detects edited or tampered chains (T34).

Review the tamper-evident audit trail of every proxy decision made by a box, merging its parent chain for a full history.

```
$ min box audit --parent my-agent-box
```

Review the full audit trail for a box, merging its parent chain, to verify that no proxy decisions or egress events have been tampered with.

```
$ min box audit --parent <box>
```

Inspect the full audit trail for a box, merging its parent chain, to verify no proxy decisions were tampered with.

```
$ min box audit --parent mybox
```

— named subcommands with tamper-detection mechanism and examples release-7b9ec7d4

---

03 **Secrets management via min secret** Not checked yet NEW 80

The new `min secret` command lets users store, list and remove secrets (T26), so a box can consume credentials like API tokens without exposing them in environment variables or config files.

Store a secret token so a box can consume it without exposing it in environment variables or config files.

```
$ min secret set MY_API_TOKEN
```

Store a secret in a box's secret store, then list all stored secrets — useful for supplying API keys to an agent without pasting them into the session.

```
$ min secret set MY_API_KEY
min secret list
```

— exact subcommands and example given, mechanism stated release-7b9ec7d4

04 **Box command CLI improvements** Not checked yet IMPROVED 78

`min box show` accepts `-o json` or `-o yaml` for machine-readable output (T43); `min box` gained list/show commands for box types (T19); box name completion now works for every `min box` verb that takes a box argument (T54); and exec attach sessions can be detached using detach keys (T44).

— four named CLI surfaces each with a stated effect release-7b9ec7d4

05 **Sealed box loadout constraints** Not checked yet NEW 70

A sealed box refuses a named loadout, a loadout that sets `loadout_mode`, or a contradicted constraint, and reads no default loadouts (T126).

— clear rule set but no command shown to invoke it release-75bea6c1

06 **Removal of bare min stop and task run --path** Not checked yet 70

BREAKING

The bare `min stop` and `min task run --path` commands are removed; use `min host stop` and `-C <dir>` instead.

— exact migration path given across all three releases release-7b9ec7d4 release-75bea6c1 release-bdc02696

07 **GitHub sign-in via min login** Not checked yet NEW 65

`min login` signs in to GitHub through the published app, minting sealed sign-in references and registering the box with the Box Egress Proxy.

— mechanism described but no flags or follow-up command shown release-bdc02696

08 **OpenTelemetry tracing and export** Not checked yet NEW 60

Adds opt-in OpenTelemetry export with a local spool across the CLI, daemon, and VM host daemon, with spans carried into guests over the VM channel.

— mechanism and components named but no enabling flag given release-7b9ec7d4 release-75bea6c1

THINNER COVERAGE BELOW

09 **Vulnerability report via min diag --security** Not checked yet NEW 55

Send a suspected vulnerability report with `min diag --security` (T28 of diag).

— exact flag named but no further mechanism explained release-75bea6c1

10 **QUIC block helper in finalize-install** Not checked yet NEW

35

Names the QUIC block helper in `min finalize-install`.

— bare naming with no explanation of behaviour release-bdc02696

11 **Box log control-code stripping** Not checked yet IMPROVED

30

Box logs strip terminal control codes when stdout is not a terminal (T45).

— one-line behavioural note with no further detail release-75bea6c1

12 **Daily update check** Not checked yet NEW

25

Checks once a day whether the channel has a newer release and reports it to the user.

— thin description with no named flag or config release-75bea6c1

└──▷ **BREAKING ON UPGRADE**

! The bare `min stop` and `min task run --path` commands are removed. Use `min host stop` and `-C <dir>` instead.

! The bare `min stop` and `min task run --path` commands are removed; use `min host stop` and `-C <dir>` instead.

WAS THIS USEFUL?  

AI MODELS

◆ Open-Weight Models

Alibaba Qwen Models

Feedback ⓘ updated · Nov 26, 2026

NOTES ↗ RANK

Qwen is Alibaba's language and multimodal model family, including coding, vision, image, speech, and reasoning models.

> Qwen added `decision-model`, a new structured classification and scoring model aimed at high-frequency business decisions.

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

01 **New decision-model for classification and scoring** Not checked yet

NEW

55

`decision-model` performs classification, yes-or-no decisions, and scoring in parallel from text, images, or business context, returning probability distributions and confidence scores. Intended use cases include ticket routing, content moderation, agent routing, and result validation.

— Names model and behavior but no API or invocation details snapshot-20261126

WAS THIS USEFUL? 👍 👎

DeepSeek V4 Flash Vision Exp

Feedback ⓘ

price increase · Nov 25, 2026 MODEL ↗

DeepSeek V4 Flash Vision Exp is a multimodal model from DeepSeek.

Price increase on OpenRouter (deepseek/deepseek-v4-flash-vision-exp): $0.2156 → $0.44 in, $0.6468 → $1.32 out per million tokens (+104%), first seen 25 Nov 2026.

Not checked yet

WAS THIS USEFUL? 👍 👎

Z.ai GLM 4.7 Flash

Feedback ⓘ context change · Nov 25, 2026

MODEL ↗

GLM 4.7 Flash is a chat model from Z.ai.

Context window shrank from 200K to 131K tokens on OpenRouter (z-ai/glm-4.7-flash), first seen 25 Nov 2026. Not checked yet

WAS THIS USEFUL? 👍 👎

MiniMax M2.5

Feedback ⓘ context change · Nov 25, 2026 MODEL ↗

STABLE

MiniMax M2.5 is a reasoning model from MiniMax.

Context window shrank from 198K to 66K tokens on OpenRouter (minimax/minimax-m2.5), first seen 25 Nov 2026. Not checked yet

▾ 1 more release in this issue · 2026-11-25

> price increase · Nov 25, 2026 MODEL ↗ STABLE
>
> Price increase on OpenRouter (minimax/minimax-m2.5): $0.27 → $0.30 in, $0.95 → $1.20 out per million tokens (+26%), first seen 25 Nov 2026. Not checked yet

WAS THIS USEFUL? 👍 👎

PrismML

Feedback ⓘ announced · Nov 24, 2026 BLOG ↗ RANK

PrismML builds 1-bit and ternary Bonsai models that pack advanced reasoning, tool use, and multimodal AI into phones, laptops, and efficient data centers.

> PrismML launched Ternary Bonsai 2.5 27B, a 1-bit/ternary model released under Apache 2.0 with cross-platform GPU and Apple support and published generation benchmarks.

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

01 **Ternary Bonsai 2.5 27B model release** Not checked yet NEW

75

Ternary Bonsai 2.5 27B runs on NVIDIA and Intel GPUs via `llama.cpp` and on Apple devices via `MLX`, with weights released under the Apache 2.0 license. On an NVIDIA H100 it achieves 291 tokens/s generation with MTP and 3,200 tokens/s prompt processing; on an RTX 4090 it delivers 204 tokens/s with MTP and approximately 3,060 tokens/s prompt processing.

— Names runtimes, license, and concrete benchmarks but no install command

launch-20261124-13933122

WAS THIS USEFUL? 👍 👎

◆ **Frontier Models**

OpenAI GPT-6 Astra Healthcare

Feedback ⓘ

released · Nov 26, 2026 MODEL ↗

GPT-6 Astra Healthcare from OpenAI was released on 24 Nov 2026. 1050K-token context; $10.00 in / $50.00 out per million tokens (models.dev). Not checked yet

WAS THIS USEFUL? 👍 👎

OpenAI GPT-6.1 Sol Healthcare

Feedback ⓘ

released · Nov 26, 2026 MODEL ↗

GPT-6.1 Sol Healthcare from OpenAI was released on 24 Nov 2026. 1050K-token context; $2.00 in / $10.00 out per million tokens (models.dev). Not checked yet

WAS THIS USEFUL? 👍 👎

OpenAI GPT-6.2 Astra Healthcare

Feedback ⓘ

released · Nov 26, 2026 MODEL ↗

GPT-6.2 Astra Healthcare from OpenAI was released on 24 Nov 2026. 1050K-token context; $8.00 in / $40.00 out per million tokens (models.dev). Not checked yet

Succeeds GPT-6 GPT-6.2 Astra Healthcare. Not checked yet

WAS THIS USEFUL? 👍 👎

OpenAI GPT-6.2 Sol Healthcare

Feedback ⓘ

released · Nov 26, 2026 MODEL ↗

GPT-6.2 Sol Healthcare from OpenAI was released on 24 Nov 2026. 1050K-token context; $2.00 in / $10.00 out per million tokens (models.dev). Not checked yet

Succeeds GPT-6.1 GPT-6.2 Sol Healthcare. Not checked yet

WAS THIS USEFUL? 👍 👎

Unbiased Pareto 26.10 Feedback ⓘ

retirement scheduled · Nov 25, 2026 MODEL ↗

Pareto 26.10 is a multimodal model from Unbiased.

OpenRouter lists a retirement date of 25 Nov 2026 for unbiased/pareto-26.10, first seen 25 Nov 2026. Not checked yet

OpenRouter lists a retirement date of 25 Nov 2026 for unbiased/pareto-26.10-preview, first seen 25 Nov 2026. Not checked yet

WAS THIS USEFUL? 👍 👎

Unbiased Pareto 26.9 Feedback ⓘ

retirement scheduled · Nov 25, 2026 MODEL ↗

Pareto 26.9 is a multimodal model from Unbiased.

OpenRouter lists a retirement date of 25 Nov 2026 for unbiased/pareto-26.9, first seen 25 Nov 2026. Not checked yet

WAS THIS USEFUL? 👍 👎

OTHER / UNCATEGORIZED

◆ UNCATEGORIZED

Tokenrip

Feedback ⓘ changes since 2026-10-27 CODE ↗

Tokenrip's command-line client.

Tokenrip adds named sign-ins, jobs/outputs/feedback commands, task queue ordering, schedules, search filtering, and table creation across versions 2.6.12–2.6.35

Filter task lists by type, step, label, or job with `--type`, `--step`, `--label`, and `--job` flags on `rip task list`. Not checked yet

List unassigned tasks (no one assigned) with `rip task list --unassigned`. Not checked yet

List tasks waiting on the caller with `rip task list --waiting`. Not checked yet

Add `--found-via` and `--source-tag` flags to `rip auth login` to tag how a sign-in was established. Not checked yet

Manage jobs, outputs, labels, and feedback with new `job`, `output`, and `feedback` commands, including `job delete`. Not checked yet

+5 more

└──▷ **BREAKING ON UPGRADE**

! The `--audience`, `--membership`, and `folder share-contents` options are removed (flat workspace access removed in 2.6.24).

! `rip comment` for workspace threads is replaced by `rip message`.

WAS THIS USEFUL? 👍 👎

// END OF TRANSMISSION

The AI Toolchain · curated daily · November 26, 2026
Releases & features from the open-source and commercial AI tools that practitioners actually run.

▷ Subscribe Feedback on this issue