THE AI TOOLCHAIN

NO. 016

Head

THE WEEK'S TOP PICKS

PUBLISHED NOVEMBER 28, 2026 · FRIDAYS

// SHARE THIS WEEK ▽ Top 5 ▽ Top Pick

The best tooling updates that shipped this week.

The week in numbers · Nov 21–27

1,138
changes we read →

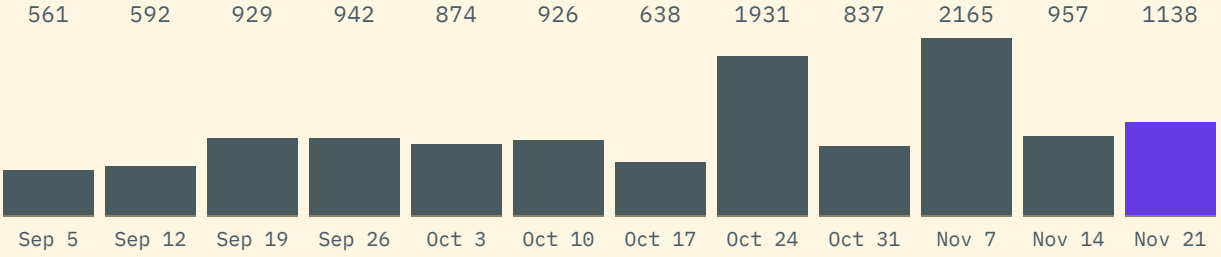
325
tools shipped →

3
product launches →

504
breaking changes →

Changes per week, last 12 weeks · Nov 21–27

Changes we read in that week's dailies, including docs and API changes. A gap means no daily was published.



▸ // HOW THESE PICKS ARE MADE SHOW

▾ // BENCHMARKS MOVED FOR TOOLS WE TRACK HIDE

Gemini API ARC-AGI-2 Gemini 4 Hydrogen xhigh

vim ? help

92.5% +42.9 pts released 2026-11-12 · leaderboard ↗

Gemini API ARC-AGI-2 Gemini 4 Hydrogen medium

89.2% +39.6 pts released 2026-11-12 · leaderboard ↗

Gemini API ARC-AGI-2 Gemini 4 Neon medium

93.8% +44.2 pts released 2026-11-18 · leaderboard ↗

Gemini API ARC-AGI-2 Gemini 4 Neon low

93.8% +44.2 pts released 2026-11-18 · leaderboard ↗

Gemini API ARC-AGI-2 Gemini 4 Hydrogen low

65.8% +16.2 pts released 2026-11-12 · leaderboard ↗

Gemini API ARC-AGI-2 Gemini 4 Hydrogen high

90.0% +40.4 pts released 2026-11-12 · leaderboard ↗

Gemini API ARC-AGI-2 Gemini 4 Neon high

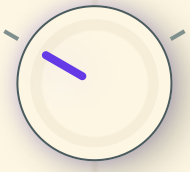
90.8% +41.2 pts released 2026-11-18 · leaderboard ↗

Scores from Epoch AI Benchmarking Hub (CC BY 4.0), evaluated 2026-11-21 to 2026-11-28.

VIEW

ISSUE VIEW

scroll



SCROLL | EXPAND IN PLACE

INFORMATION DENSITY

everything



BRIEFED | EVERYTHING

Every change in this week's picks, in full. all 65 changes

Do you prefer **scroll**? 👍 👎

This week's 5

01 Pipecat Feature · Speech & Voice AI

02	**PydanticAI**	Feature · AI Agent Frameworks
03	**gram**	Feature · MCP Servers & Tooling
04	**Databricks AI**	Feature · ML Platforms & SDKs
05	**FastMCP**	Feature · MCP Servers & Tooling

The picks in full

01 Pipecat v1.17.0

Speech & Voice AI

Pipecat is a framework for realtime voice and multimodal agents, orchestrating speech-to-text, model inference, and text-to-speech as a streaming pipeline.

// WHY IT MADE THE LIST — DEPTH 5/5 — IMPACT 5/5 — NOVELTY 4/5

Native SIP transport lands as a complete, production-ready capability: dial-in/out, DTMF in both directions, attended transfers via REFER+Replaces, VP8 video, hold/resume, and full NAT traversal config — all in one `pipecat-ai[sip]` install, with Daily SIP credential provisioning from code; the new LLM extensions middleware lets you attach pluggable behaviors (system-prompt augmentation, frame interception, output filtering) to any LLM service without modifying the service itself.

Pipecat adds native SIP transport, LLM extensions middleware, Rime Prism speech-to-speech service, AWS Transcribe language ID, Deepgram Flux turn-endpoint control, and user waiting silence detection.

└──▷ GET THIS VERSION ▾

```
$ git clone --branch v1.17.0 https://github.com/pipecat-ai/pipecat.git
# already have the repo? check out this version:
$ git checkout v1.17.0
```

NEW INTEGRATION Add a native SIP transport with `pipecat-ai[sip]`: the SIPTransport class registers a SIP account or dials through a registration-less trunk, handles dial-in/dial-out with `on_dialin_ready`, `on_dialin_connected`, `on_dialout_answered`, start_dialout(), send_dtmf(), and sip_refer(), and supports DTMF in both directions, hold/resume, and VP8 video. Not checked yet

Configure SIP media-NAT traversal and debugging via SIPConnection(extra_params=..., jitter_buffer_mode=..., jitter_buffer_ms=..., native_log_level=..., sip_trace=...), or via environment variables `SIP_EXTRA_PARAMS`, `SIP_STUN_SERVER`, `SIP_JITTER_BUFFER`, `SIP_NATIVE_LOG_LEVEL`, and `SIP_TRACE=1`. Not checked yet

Run the development runner in SIP mode with `-t sip` to start an HTTP server with `POST /start` and `POST /daily-sip-trunk-webhook`; use `-t sip -d` to run one bot immediately with no server; use `SIP_NET_INTERFACE` to restrict the stack to one local interface. Not checked yet

Provision Daily SIP credentials from code with new `DailyRESTHelper` methods: create_sip_client(), list_sip_clients(), get_sip_client(), delete_sip_client(), and rotate_sip_client_password(), wrapping Daily's `/sip-clients` REST API. Not checked yet

Perform an attended SIP transfer with SIPTransport.sip_attended_transfer({"toEndPoint": "sip:..."}): the transport dials a consult leg, waits for answer, then sends a REFER with a Replaces header so the two peers connect directly. Not checked yet

+7 more

NEW INTEGRATION Rime Prism speech-to-speech integration: RimeRealtimeLLMService

---

GOOD PICK? 👍 👎 GOOD REASON? 👍 👎

02 PydanticAI v2.74.0

AI Agent Frameworks

How Python does AI: agents, realtime voice, image generation, embeddings. Every model, every interface, typed end to end.

// WHY IT MADE THE LIST — DEPTH 5/5 · IMPACT 5/5 · NOVELTY 4/5

Computer-use support lands across 10 sandbox runtimes (Docker, Daytona, E2B, Sprites, BoxLite, Boxd, Smol, Render, Modal, Kubernetes) with native routing to OpenAI's computer tool on GPT-5.4+, Anthropic's computer-use tool, and Google's computer-use tool on Gemini — a practitioner can now wire a computer-use agent to any of these backends in a few lines; RunContext.submit_output() also lets tools end a run directly, closing a long-standing control-flow gap.

> PydanticAI's latest three releases add computer-use agents across eleven sandboxed desktop runtimes with native routing to OpenAI, Anthropic, and Google tools, a Temporal-backed deferred tool-call waiter, GPT-Live realtime voice controls, and a long run of breaking changes to authority-based workspace visibility, fallback-model error handling, and Gemini/OpenAI token accounting.

└──▷ WHAT SHIPPED · 30 FEATURES most completely described first ? what's the number?

01 Authority-based workspace file visibility in RepoContext, Skills, and Memory

Not checked yet — 90

BREAKING

`RepoContext`, `Skills`, and `Memory` gain a `lowest_authority_under` parameter controlling which workspace files are kept out of the system-prompt instructions. Workspace files below `operator` authority no longer reach the system-prompt instructions: `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`. 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. 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.

— Names config key and migration path but no count of affected files v2.75.0

02 **Computer-use agents across eleven sandboxes and three model providers**

Not checked yet NEW 90

Enable computer-use agents with `ComputerTool` from `pydantic_ai.computer_use` and the ComputerUse() harness capability, backed by the Display protocol. Run them in a desktop environment via 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). 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.

— Every sandbox constructor and model routing target named v2.74.0

03 **Automatic fallback on empty or malformed model responses**

Not checked yet 85

BREAKING

`FallbackModel` and `ModelFallback` now automatically fall back instead of raising `UnexpectedModelBehavior` when the model stream is empty, missing `choices`, or returns malformed JSON, via the new `ModelEmptyResponseError` and `ModelInvalidResponseError`. An `except ModelAPIError` placed before `except UnexpectedModelBehavior` will now catch both new errors; to keep raising instead of falling back, pass a `fallback_on` such as lambda e: isinstance(e, ModelAPIError) and not isinstance(e, UnexpectedModelBehavior).

— Exact exception names and migration lambda given v2.75.0

04 **Targeted clearing in ClearToolResults** Not checked yet IMPROVED

75

`ClearToolResults` now accepts a `tools` argument to target specific tools, a `keep_turns` argument to retain whole user turns, and a callable `placeholder` argument.

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]',
)
```

— All three new arguments named with a runnable example v2.75.0

05 **SubAgents delegation instruction gated by model profile**

Not checked yet 75

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 when 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=...).

— Names the profile flag, block id, and two migration options v2.75.0

06 **Gemini thinking tokens counted in output usage** Not checked yet

75

BREAKING

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. On a server whose `total_tokens` counts more than input, output, and thinking tokens, pass profile=OpenAIModelProfile(openai_chat_completion_tokens_omit_reasoning=False).

— Exact profile flag and affected fields named v2.75.0

07 **Tools can end a run directly via submit_output** Not checked yet

NEW 70

Call RunContext.submit_output() from inside a tool function to submit the run's final output and end the run directly, without waiting for the model to call a final 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'
```

— Exact method name with a runnable code example v2.75.0

08 **Higher default thinking_budget for gemini-2.5-flash**

Not checked yet 70

BREAKING

`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}`.

— Exact values, date, and two workaround settings named v2.75.0

09 **Resume token required in Pax approval handler** Not checked yet

65

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. 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.

— Exact module path and replacement binding given v2.75.0

10 **Previous-response continuation inherits static instructions on GPT-6.2**

Not checked yet 65

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 parameter and profile flag with two fixes named v2.74.0

11 **Max image dimension control in Coder** Not checked yet NEW

60

Control maximum image dimensions in `Coder` with the new `FileSystem(max_image_side=...)` argument, which downscales images to at most 1024 pixels on the longest side by default.

Cap the size of images the Coder reads from disk to avoid sending oversized payloads to the model.

```python
from pydantic_ai_harness import Coder
from pydantic_ai_harness.filesystem import FileSystem

agent = Agent(
    'anthropic:claude-fable-5-2',
    capabilities=
[Coder(filesystem=FileSystem(max_image_side=512))],
)
```

— Exact argument, default value, and runnable example v2.73.0

12 **FallbackModel no longer falls through on 400 for unified settings**

Not checked yet BREAKING

60

A `FallbackModel` that relied on a 400 error (from sending unsupported unified settings to Claude, Gemini, or Grok) to reach a later model now stays on the first model without the setting; list a model that supports the setting first to preserve the behaviour.

— Names affected providers and the workaround ordering v2.73.0

THINNER COVERAGE BELOW

13 **Gemini streaming usage now reports endpoint totals**

Not checked yet BREAKING

55

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.

— States before/after but no workaround setting given v2.74.0

14 **Runtime warning for deprecated models** Not checked yet NEW

55

Raising `DeprecatedModelWarning` at runtime warns when a request targets a deprecated Anthropic or OpenAI model.

Catch a warning before a deprecated model is called, so a CI step or test suite surfaces the problem immediately rather than waiting for a provider-side error.

```python
import warnings
from pydantic_ai.exceptions import DeprecatedModelWarning

warnings.filterwarnings('error', category=DeprecatedModelWarning)

agent = Agent('openai:gpt-4')
result = agent.run_sync('Summarise this report.')
```

— Exact warning class with a runnable CI example v2.73.0

15 **Realtime voice session controls on GPT-Live** Not checked yet

IMPROVED

55

Append dynamic voice instructions mid-session on GPT-Live with the new `RealtimeSession.append_voice_instructions` method, including a Twilio Media Streams bridge example, and pass a custom voice object via `openai_voice`.

— Names the method and the voice parameter, no snippet included v2.73.0

16 **Dependency floors raised across openai, pillow, and boxd extras**

Not checked yet

BREAKING

55

The `openai` floor is raised from 3.31.0 to 3.50.0 on every extra that installs `openai`; the `coder` extra now requires `pillow>=12.3.0`; `pydantic-ai-harness[boxd]` now requires `boxd>=0.3.3,<0.4`, so code calling 0.2-only boxd APIs directly needs updating.

— Exact version floors named for all three extras v2.73.0

17 **Working directory announced once per run** Not checked yet NEW

50

The new `AnnounceWorkingDirectory` capability makes `Shell`, `FileSystem`, and custom tools announce the working directory once per run rather than repeating it in every tool description.

— Names the capability but not its configuration options v2.75.0

18 **Temporal ExternalEventWaiter for deferred tool calls**

Not checked yet NEW 50

A new Temporal `ExternalEventWaiter` lets `HandleDeferredToolCalls` token handlers and `AskUser` wait durably inside a Temporal workflow.

– Names the class but not its interface or limits v2.75.0

---

19 **Shared instructions identity in ModelMessagesTypeAdapter**

Not checked yet BREAKING 50

`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`.

– Exact type and comparison guidance named v2.75.0

---

20 **Egress policy disables WebRTC in PlaywrightBrowser**

Not checked yet BREAKING 50

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.

– Names every disabled feature, no opt-out mentioned v2.74.0

---

21 **Pax transcript theming and folder trust prompts** Not checked yet

NEW 45

Pax now shows user prompts in the transcript on a grey background under a `/theme` palette, and asks whether to trust a new folder via a `/trust` command with a status-row indicator.

– Names both commands but no configuration detail v2.75.0

---

22 **Capability overrides and public instruction types** Not checked yet

NEW 45

Control agent capabilities at override time by passing `capabilities` to `Agent.override`, and import the now-public `AgentInstructions` and `SourcedInstruction` types.

– Names the parameter and two new public types v2.73.0

---

23 **pax --version command** Not checked yet NEW 40

Add `pax --version` to report the installed Pax version.

Check the installed Pax version at the terminal before filing a bug report or upgrading.

```
$ pax --version
```

— Single exact command, nothing more to describe v2.75.0

24 **Per-model reasoning guidance in Coder** Not checked yet IMPROVED 40

Coder now sends OpenAI-specific `'max'` reasoning guidance through per-model profile flags instead of static instructions.

— States the change but not the profile flag name v2.75.0

25 **CDPBrowser harness capability** Not checked yet NEW 40

Drive Chrome directly with raw DevTools Protocol commands using the new CDPBrowser harness capability.

— Names the capability but no example of commands used v2.74.0

26 **JSON string tool returns for Qwen3 Next** Not checked yet NEW 40

Send tool text output to Qwen3 Next as a JSON string field with `needs_json_string_tool_returns`.

— Exact flag named, no further mechanism given v2.74.0

27 **SubAgents deferred tool-call delegation** Not checked yet IMPROVED 40

`SubAgents` can now pass a delegate's deferred tool calls back to the caller and resume the delegate once they are answered.

— Describes behaviour but no API names for the handoff v2.73.0

28 **Thinking profiles for six model families** Not checked yet NEW 40

Add thinking profiles for the Nemotron, Inkling, MiMo, Step, Hy, and Ling model families on OpenAI-compatible gateways.

— Names all six families but no profile mechanism detail v2.73.0

29 **OpenAI Responses input_transformations surfaced** Not checked yet NEW 40

Surface OpenAI Responses `input_transformations` in `provider_details` and on the model request span.

— Names the field and two surfaces, no further detail v2.73.0

30 **Override capability context variable now a tuple** Not checked yet

35

BREAKING

The private `Agent._override_root_capability` context variable now holds a (capability, injected) tuple; code that reads it directly needs updating.

— Exact internal variable and new shape named `v2.73.0`

└──▷ **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.

! The `openai` floor is raised from 3.31.0 to 3.50.0 on every extra that installs `openai`.

! The private `Agent._override_root_capability` context variable now holds a (capability, injected) tuple; code that reads it directly needs updating.

! The `coder` extra now requires `pillow>=12.3.0`.

! A `FallbackModel` that relied on a 400 error (from sending unsupported unified settings to Claude, Gemini, or Grok) to reach a later model now stays on the first model without the setting; list a model that supports the setting first to preserve the behaviour.

! `pydantic-ai-harness[boxd]` now requires `boxd>=0.3.3,<0.4`; code that calls 0.2-only boxd APIs directly needs updating.

GOOD PICK? 👍 👎 GOOD REASON? 👍 👎

03 gram server@2.43.0

MCP Servers & Tooling

Securely scale AI usage across your organization. A single stack to Connect, Secure, Observe and Distribute agents, MCPs, and Skills within your company.

// WHY IT MADE THE LIST — DEPTH 5/5 — IMPACT 5/5 — NOVELTY 5/5

The agent feedback system is a first-of-kind observability primitive: MCP servers now expose a `submit_feedback` tool agents call when blocked or confused, with reports stored PII-masked and exportable as OTLP logs or via webhook — closing the loop between agent failures and the teams who built the tools; the same week also ships access-request workflows (agents file time-bounded access requests from the device agent) and MCP connection health monitoring.

> Gram's biggest shipment this window is a full Access Requests workflow spanning device agent, Slack and dashboard, alongside a new MCP connection health monitoring API/UI and an agent feedback system for MCP servers; it also replaced several Platform MCP tools with an Organization Gateway, added encryption-key migration tooling for OAuth secrets, and removed legacy session-issuer RPCs and the Tailscale Kubernetes operator provider.

└──▷ **WHAT SHIPPED · 21 FEATURES** most completely described first — ? what's the number?

01 Access Requests workflow for temporary permissions

Not checked yet — NEW — 95

Members can request temporary access with a duration (1, 2, or 4 hours) or a deadline such as end of day, plus an optional message, via `agent.requestAccess`, and can list what they can still request with `agent.listRequestableAccess`; when a delivered MCP server reports no access, `agent.getMyConnections` now names the exact access to request. Organization administrators manage requests through `listAccessRequests`, `decideAccessRequest`, and `withdrawAccessRequest` endpoints, can list pending ones from the device agent via `agent.listPendingAccessRequests` (newest-first, with dashboard approve/deny links), or approve/deny directly from Slack DMs with a duration picker, notifying requesters by DM on approval, denial, or expiry. Requests carry a `requested_until` field stored in UTC and surfaced on the Slack card, and the dashboard's new Access Requests tab under Roles & Permissions shows a Wants it column and lets admins grant temporary access or assign a role directly.

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()
```

Fetch the first page of pending access requests as an org admin to review what needs a decision.

```
$ curl -X POST https://<your-gram-server>/api/listAccessRequests
\
  -H 'Authorization: Bearer <token>' \
  -H 'Content-Type: application/json' \
  -d '{}'
```

As an administrator, approve a pending access request for 4 hours using the new endpoint.

```
$ curl -X POST https://api.example.com/decideAccessRequest \
  -H 'Authorization: Bearer <token>' \
  -d '{"requestId": "<request_id>", "decision": "approve",
"duration": "4h"}'
```

— Names every endpoint and field with runnable examples `server@2.44.0` `dashboard@0.157.0` `server@2.39.0` `dashboard@0.153.0`

02 **MCP connection health monitoring** Not checked yet NEW

95

The `mcpConnectionHealth` API, gated behind the `mcp-connection-health` feature flag (which can target individual people, not just whole orgs), lists suspected problems for an MCP sign-in and returns paged connection logs (repeated attempts folded into one row), totals and per-filter-value counts, individual attempt step detail, lookup by reference, sign-in and renewal summaries, active sessions per day with

average session length, and apps refused at sign-in; sign-in summaries can be filtered by step group via the optional `steps` parameter to `getSignInSummary`. The dashboard surfaces this as Connection Logs and Active Sessions tabs with filters, reference search, and per-attempt detail, plus a per-server view on each server's Clients and Sessions tab, and MCP Sessions is now out of beta.

Retrieve a sign-in summary filtered to specific pipeline steps, matching what Connection Logs shows for those steps.

```javascript
getSignInSummary({ steps: ['sign-in', 'sign-in-renewal'] })
```

— Named API, flag and parameter with a runnable example `server@2.43.0` `server@2.42.0` `dashboard@0.155.0`

---

03 **Agent feedback system for MCP servers** Not checked yet NEW

90

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 also 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"
}
```

— Names the tool, qualified-name format and webhook event `server@2.43.0`

---

04 **Pagination and search for keys.list** Not checked yet IMPROVED

85

`keys.list` can now be paginated with `limit` (1–100) and `cursor` (via the returned `next_cursor`), and filtered with case-insensitive `search` by key name; callers that omit `limit` still receive all keys in one response.

Walk through all API keys for a large organization without timing out by paging 50 at a time and filtering by name.

```
$ curl -X POST https://<your-gram-server>/api/keys.list \
  -H 'Authorization: Bearer <token>' \
  -H 'Content-Type: application/json' \
  -d '{"limit": 50, "search": "prod", "cursor": "<next_cursor>"}'
```

Fetch the first page of API keys whose name contains `prod`, then retrieve the next page using the returned cursor.

```
$ # Page 1
curl -X POST https://api.example.com/keys.list \
  -H 'Authorization: Bearer <token>' \
  -d '{"limit": 50, "search": "prod"}'

# Page 2 — pass the next_cursor from the previous response
curl -X POST https://api.example.com/keys.list \
  -H 'Authorization: Bearer <token>' \
  -d '{"limit": 50, "search": "prod", "cursor": "<next_cursor>"}'
```

— Names all parameters with runnable paginated curl examples `server@2.39.0` `dashboard@0.153.0`

05 **tool_calls_only filter for observability overview** Not checked yet

IMPROVED 75

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`

06 **`_meta` support for resources with ui.csp** Not checked yet NEW

75

Set metadata on resources by passing `meta` to resource() or experimental_uiResource(), which returns it on `resources/list` and `resources/read`; MCP Apps UI resources can use it to declare `ui.csp`.

Declare a Content Security Policy for an MCP Apps UI resource so the host enforces it when rendering the resource.

```javascript
experimental_uiResource({
  uri: 'myapp://dashboard',
  name: 'Dashboard',
  meta: {
    ui: {
      csp: "default-src 'self'; script-src 'self'"
    }
  },
  handler: async () => ({ html: '<h1>Dashboard</h1>' })
})
```

— Names the functions and field with a runnable example `@speakeasy-api/functions@0.21.0`

07 Organization Gateway replaces Platform MCP server tools

Not checked yet — 70

BREAKING

Platform MCP no longer offers `list_my_servers`, `search_my_tools`, `describe_tools`, or `execute_tool`; MCP clients that called these must instead retrieve the Organization Gateway connection via `get_my_organization_gateway` to use their own MCP servers from an MCP client.

— Names removed tools and the replacement call, no walkthrough `server@2.43.0`

08 Encryption key migration tooling in admin dashboard

Not checked yet — IMPROVED — 70

Staff can migrate organizations' 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, with enroll, backfill, finalize, and rollback steps; a dedicated step-by-step Tailscale OAuth migration page runs the same enroll, backfill-and-verify, finalize, then clear flow for saved Tailscale OAuth clients. 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.

— Names the migration steps and UI path, no API example `admin@0.14.2` `server@2.43.0` `admin@0.14.1`

09 **Session issuer RPCs renamed** 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.

— Names all six old and new RPCs, a clear migration `server@2.44.0`

10 **Tailscale Kubernetes operator provider removed** Not checked yet

65

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 exact env vars to delete from deployments `server@2.43.0`

11 **Managed assistant platform toolset removed** Not checked yet

60

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 removed flag and tool pattern, no migration steps `server@2.44.0`

THINNER COVERAGE BELOW

12 **Read-only tool execution in progressive gateways** Not checked yet

NEW

55

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.

— Names the new tool and its scope, no usage example `server@2.43.0`

13 **Intent-based policy rules and preview page** Not checked yet NEW

55

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 scope and preview page but not a config key `server@2.44.0` `server@2.43.0`

---

14 **Hooks release rollout progress tool** 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.

— Names the exact staff tool and its three metrics `server@2.44.0`

---

15 **Upstream auth enforcement refusal visibility** Not checked yet

NEW

55

View upstream auth enforcement refusals — including personal connection misuse — under failed authorization checks on the agent's or person's Permissions tab and in Roles & Permissions, or list refusals per server with the Platform MCP `list_agent_authorization_checks` tool.

— Names the tool and two UI locations, no example run `server@2.41.0`

---

16 **merged_client_ids field in auth consolidation preview**

Not checked yet IMPROVED

45

Preview MCP Authentication Provider consolidations now include a `merged_client_ids` field listing which OAuth client IDs will merge.

— Names the field but not how to call the preview `server@2.41.0`

---

17 **GA for GitHub Actions workload platform and prompt-based risk policies**

Not checked yet NEW

45

GitHub Actions is now available as a workload platform to every organization without requiring the `workload-platform-github-actions` feature flag, and prompt-based risk policies are now available to every project without a feature flag.

— Names both former flags now removed, no setup steps `server@2.41.0`

18 **EU-region AI model routing via OpenRouter** Not checked yet NEW

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.

— Describes the behavior change with no config surface `server@2.43.0`

19 **unannotated rule value for ungoverned tools** Not checked yet NEW

40

Use `unannotated` as a rule annotation value in the access, agents, and Platform MCP APIs to classify and govern tools that carry no annotations.

— Names the value and three APIs, no example call `dashboard@0.153.0`

20 **Block rules now cover undeterminable tool name or kind**

Not checked yet IMPROVED

35

A tool-name or tool-kind block rule now also blocks calls where the tool name or kind cannot be determined, closing a gap where such calls fell outside the rule.

— Describes the behavior fix with no named rule syntax `server@2.41.0`

21 **Expanded model completion span telemetry** Not checked yet IMPROVED

30

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.

— States what fields were added, no access path given `server@2.44.0`

└──▷ **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.

GOOD PICK? 👍 👎 GOOD REASON? 👍 👎

04 Databricks AI

announced · Nov 19, 2026

ML Platforms & SDKs

Databricks AI provides managed model serving, agent development, evaluation, and retrieval over data in the Databricks platform.

// WHY IT MADE THE LIST

DEPTH 4/5 IMPACT 5/5 NOVELTY 4/5

Genie ZeroOps moves from preview to public beta: it autonomously detects data pipeline failures, investigates them, and opens pull requests with pre-populated code fixes — a team inbox, custom agent instructions, and one-click PR creation are all included, closing the loop from alert to remediation without a human in the middle; the week also ships Genie Migrate for Serverless (bulk classic-to-serverless migration with reviewable diffs) and Tag Exchange for Iceberg REST Catalog.

> Databricks AI shipped a GA native Lakebase Postgres sink for Structured Streaming with declarative upserts, alongside two new Genie-branded agents in public beta: ZeroOps for autonomous pipeline triage and Migrate for Serverless for classic-to-serverless workload conversion.

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

01 Native Lakebase Postgres sink for Structured Streaming

Not checked yet NEW INTEGRATION NEW SUPPORTED PLATFORM NEW

95

Generally available on Databricks Runtime 18 LTS and above, the sink lets you write streaming results directly to Lakebase with a single `writeStream` call using .toTable() (for Unity Catalog-registered tables) or format("postgresql") (for unregistered tables or external databases), replacing custom `ForeachWriter` implementations. Declare upsert semantics with the `upsertkey` option, with declarative merge strategies (DBR 19+) controlling conflict resolution without hand-written `ON CONFLICT` SQL; tune throughput with `batchinterval` and `batchsize`. Writing to external PostgreSQL databases (e.g. Amazon RDS or self-hosted) via a Unity Catalog connection is in Public Preview on DBR 19+, and the connector works in both micro-batch and Real-Time Mode triggers, using `Trigger.AvailableNow` on serverless compute.

Write to an external PostgreSQL database (such as Amazon RDS) via a Unity Catalog connection when the target table is not registered in Unity Catalog.

```
$ df.writeStream
  .format("postgresql")
  .option("dbtable", "public.user_activity")
  .option("upsertkey", "user_id")
  .start("<lakebase_or_rds_endpoint>")
```

— Exact flags, APIs and a runnable writeStream example given `launch-20261123-39bdded8`

02 **Genie Migrate for Serverless** Not checked yet NEW 85

A beta agent that automates classic-to-serverless workload migration with reviewable diffs and rollback. Run bulk migrations from the command line via the `compute-migrations` command group in Databricks CLI 1.24.0 or above to build serverless working copies of multiple jobs without an open Genie session, or interactively migrate a single job/notebook by running `/compute migrate` inside Genie, which scans eligibility and proposes changes as a labeled, reversible diff. Enable the beta via Genie Migrate for Serverless in workspace preview settings before running `/compute migrate`.

Migrate a single job or notebook interactively after enabling the beta in workspace preview settings — the agent scans eligibility and shows a reviewable diff before applying anything.

📍 `/compute` migrate

— Named CLI command group and runnable `/compute migrate` command

`launch-20261119-9c8460c9`

03 **Genie ZeroOps autonomous pipeline agent** Not checked yet NEW 70

Now in public beta and enabled from the workspace Previews page, Genie ZeroOps detects, investigates and raises PRs for data pipeline issues. It adds a shared team inbox for incident visibility, custom agent instructions to embed business rules (flaky sources, preferred fixes, naming conventions), one-click PR creation with pre-populated fixes for Git-backed jobs, and notifications to Slack, Teams or PagerDuty with filtering for low-priority alerts. It applies Unity Catalog permissions to govern what it can read, redacts PII before notifications leave Databricks, and detects silent issues like late-arriving commits (freshness) and below-expected row counts (completeness) via Data Quality Monitoring, in addition to hard job failures.

Have a workspace admin activate Genie ZeroOps so the team can start receiving autonomous pipeline triage in the current workspace.

📍 In the Databricks workspace UI, go to Settings › Previews and enable the ZeroOps feature. Once enabled, find the ZeroOps banner at the top of the Pipelines

page.

— Rich feature description but only a UI enablement path, no API/flag

launch-20261119-dfda78f2

announced · Nov 24, 2026

Databricks adds Tag Exchange BETA to share Unity Catalog table- and column-level tags via the Apache Iceberg REST Catalog `labels` field for discovery, governance, and agent workflows.

Enable Tag Exchange via the Databricks Previews portal to expose Unity Catalog's table- and column-level tags through the `labels` field in the Apache Iceberg REST `LoadTable` response, making classifications consumable by any Iceberg REST-compatible platform, engine, or agent. Not checked yet

NEW INTEGRATION Unity Catalog now surfaces table- and column-level tags as string key-value pairs through the Apache Iceberg REST Catalog API's `labels` field, allowing downstream tools to use business context such as ownership, domain, and sensitivity without custom sync pipelines. Not checked yet

announced · Nov 19, 2026

Lakebase adds PostgreSQL 19 on day-one GA, with REPACK CONCURRENTLY, query plan advice extensions, WAIT FOR LSN, and logical replication upgrades.

Reclaim table space without blocking reads or writes using the new REPACK (CONCURRENTLY) command, which rebuilds tables online and optionally clusters rows via `USING INDEX`; requires free capacity under `max_repack_replication_slots` and a non-deferrable primary key or index-based replica identity. Not checked yet

Stabilize unstable query plans without touching application SQL using new `pg_plan_advice` and `pg_stash_advice` extensions; stash advice by query ID and activate it per-database so PostgreSQL applies it automatically on matching queries. Not checked yet

Block a standby until it has replayed a specific WAL position to implement read-your-writes on replicas using new `WAIT FOR LSN`. Not checked yet

Get a clean get-or-create upsert pattern that returns both new and existing rows without phantom dead tuples using new `INSERT ... ON CONFLICT DO SELECT`. Not checked yet

Export query results directly as JSON from `COPY TO`, using `FORCE_ARRAY` to wrap output in a single JSON array instead of one object per line. Not checked yet

+9 more

> NEW SUPPORTED PLATFORM PostgreSQL 19 on Lakebase

GOOD PICK? 👍 👎 GOOD REASON? 👍 👎

05 FastMCP v4.2.6

MCP Servers & Tooling

FastMCP is a Python framework for building MCP servers and clients, handling transport, schema generation, authentication, and server composition.

// WHY IT MADE THE LIST — DEPTH 4/5 — IMPACT 4/5 — NOVELTY 4/5

FastMCP's new `CodeModeClient` wraps any `Client` or `ClientGroup` to give agents search, schema, and execute meta-tools over servers that don't run CodeMode themselves — a single `execute` block can chain tools across servers — and the new `fastmcp-events` package adds typed Pydantic event models with Redis-backed persistence via `DocketEventLog`, enabling durable pub/sub across restarts and replicas.

> FastMCP's three releases rework its event system into a stable `fastmcp-events` package (replacing the experimental extension), add live subscription streams and client-side CodeMode execution, and tighten OAuth/OpenAPI security defaults.

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

01 fastmcp-events package replaces experimental Events extension

Not checked yet — 95

BREAKING

FastMCP ships a new `fastmcp-events` package (`pip install 'fastmcp[events]'`, import `fastmcp_events`) for declaring typed events as Pydantic models registered with @events.event('name'), emitted via await events.emit(...), and consumed via @events.reader(Model) with an optional `for_subscription` method; events persist across restarts and replicas via `DocketEventLog` (`pip install 'fastmcp-events[docket]'`). This replaces the prior experimental `fastmcp.experimental.events` extension (SEP-3415), which offered poll/push delivery and a `fastmcp experimental events list|watch` CLI and is now removed — each event type, formerly a check function, must be ported to the new Pydantic model/reader pattern.

Define a typed event, emit it from a tool, and consume it with a reader in the new fastmcp-events package.

python

```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')
```

— Full code example, install commands, and explicit migration path given. v4.2.6

v4.2.5

02

Live subscription streams for change notifications

Not checked yet NEW 90

FastMCP now serves `subscriptions/listen` on 2026-07-28 connections so clients can hold an open stream and receive tool, prompt, and resource change events pushed by FastMCP.notify_tools_changed() and its siblings; servers advertise `listChanged` and `resources.subscribe` and accept up to 1024 open streams per process by default. Disable this server-wide with `FASTMCP_ENABLE_SUBSCRIPTIONS_LISTEN=false` in the environment or `fastmcp.settings.enable_subscriptions_listen = False` in code before the server is created — recommended for serverless hosts or small worker pools.

Opt out of live subscription streams on a serverless host where long-lived open requests are billed or time-limited.

```
$ FASTMCP_ENABLE_SUBSCRIPTIONS_LISTEN=false fastmcp run
my_server.py
```

— Runnable command and exact env var/setting to toggle the feature. v4.2.5

03 **OpenAPI header forwarding and cookie encoding** Not checked yet

90

BREAKING

New `forward_client_headers` option on `FastMCP.from_openapi`, `FastMCP.from_fastapi`, and `OpenAPIProvider` controls which client headers reach a wrapped OpenAPI backend — pass `False` to forward none or a list of names to forward only those. OpenAPI tools no longer forward MCP transport headers such as `mcp-protocol-version` to the wrapped API, and OpenAPI cookie parameter values are now percent-encoded outside the characters cookies allow, so a backend reading a cookie containing `;`, spaces, or non-ASCII text must decode it.

Prevent an OpenAPI-backed server from leaking MCP transport headers to the upstream API, while still forwarding an allow-listed `x-request-id` header.

python

```python
from fastmcp import FastMCP

mcp = FastMCP.from_openapi(
    openapi_url='https://api.example.com/openapi.json',
    forward_client_headers=['x-request-id'],
)

if __name__ == '__main__':
    mcp.run()
```

— Exact option name, affected header, and runnable example provided. v4.2.5

04 **Client-side CodeMode execution** Not checked yet NEW

70

The new `CodeModeClient` class wraps any `Client` or `ClientGroup` to run CodeMode on the client side, giving agents search, schema, and execute meta-tools over servers that do not run CodeMode themselves — a single `execute` block can chain tools across servers.

— Named class and meta-tools but no runnable code example. v4.2.6

05 **OAuth security hardening for dynamic clients** Not checked yet

65

BREAKING

If neither `valid_scopes` nor `required_scopes` is set, CIMD clients can no longer request any scope. A redirect URI must now match an exact allowed pattern's query component and may not carry a fragment, and an exact pattern at a host's root (e.g. `https://app.example/`) now matches only that exact path unless `/*` is appended. Presenting a refresh token more than 60 seconds after it was rotated now ends the whole grant as suspected theft, and revoking a refresh token now also revokes that grant's access tokens.

– Precise before/after rules but no config example to adapt. v4.2.5

06 **UsageMiddleware for per-user tool usage metrics** Not checked yet

NEW

65

New `UsageMiddleware` records per-user, per-tool usage as the OpenTelemetry `mcp.server.operation.duration` histogram for `tools/call` and `prompts/get`; it should be added before any error-handling middleware.

– Named middleware and metric but no instantiation example. v4.2.5

07 **AzureProvider token issuer pinning** Not checked yet

BREAKING

65

Pin the token issuer in `AzureProvider` for multi-tenant authorities (`organizations`, `consumers`, or `common`) by passing `token_issuer`; without it, users from any tenant are now accepted.

– Names the parameter and behavior change, no code sample. v4.2.5

THINNER COVERAGE BELOW

08 **client_registration_ttl_seconds for dynamic client registration**

Not checked yet NEW

50

New `client_registration_ttl_seconds` setting on `OAuthProxy` and OIDCProxy lets unused dynamic client registrations expire from persistent storage.

– Names the exact setting but no usage example. v4.2.5

09 **OpenTelemetry network transport attributes** Not checked yet NEW

40

Server spans now carry `network.transport` and, over HTTP, `network.protocol.name` and `network.protocol.version` OpenTelemetry attributes.

– Names attributes but gives no way to enable or inspect them. v4.2.6

10 **Authlib decoupled from core 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.

– One-line dependency change with minimal elaboration. v4.2.6

11 **Mount versioning mismatch warning** Not checked yet IMPROVED

25

FastMCP now warns when a mount mixes versioned and unversioned components under the same name, surfacing ambiguous server configurations at mount time.

— Single sentence with no example or config key shown. v4.2.4

└──▷ **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).
- ! If you set neither `valid_scopes` nor `required_scopes` , CIMD clients can no longer request any scope; set `valid_scopes` to the scopes you accept.
- ! A redirect URI must now match an exact allowed pattern's query component and may not carry a fragment.
- ! An exact redirect URI pattern at a host's root (such as `https://app.example/`) now matches only that exact path; append `/*` to allow any path.
- ! Presenting a refresh token more than 60 seconds after it was rotated ends the whole grant as a suspected theft; clients that copy one stored refresh token between processes must share the rotated one.
- ! Revoking a refresh token now also revokes that grant's access tokens.
- ! On 2026-07-28 connections, servers now advertise `listChanged` and `resources.subscribe` and accept up to 1024 open `subscriptions/listen` streams per process by default; servers on serverless platforms or with small worker pools should set `FASTMCP_ENABLE_SUBSCRIPTIONS_LISTEN=false` .
- ! OpenAPI tools no longer forward MCP transport headers such as `mcp-protocol-version` to the wrapped API.
- ! OpenAPI cookie parameter values are now percent-encoded outside the characters cookies allow; a backend that reads a cookie containing `;` , spaces, or non-ASCII text must decode it.

GOOD PICK? 👍 👎 GOOD REASON? 👍 👎

Where each pick stands · Nov 21–27

Tools with changes in each pick's category over the last 12 weeks, including the pick itself.

Pipecat

Speech & Voice AI

10

PydanticAI
AI Agent Frameworks
109

gram
MCP Servers & Tooling
98

Databricks AI
ML Platforms & SDKs
104

FastMCP
MCP Servers & Tooling
98

Each pick's release rhythm · Nov 21–27

Changes per week for the last 12 weeks. The bright bar is this week; a dash marks a week with no daily.

Pipecat
96
in 12 known wk

PydanticAI
91
in 12 known wk

gram
76
in 12 known wk

Databricks AI
22
in 12 known wk

FastMCP
13
in 12 known wk

Which categories moved · Nov 21–27

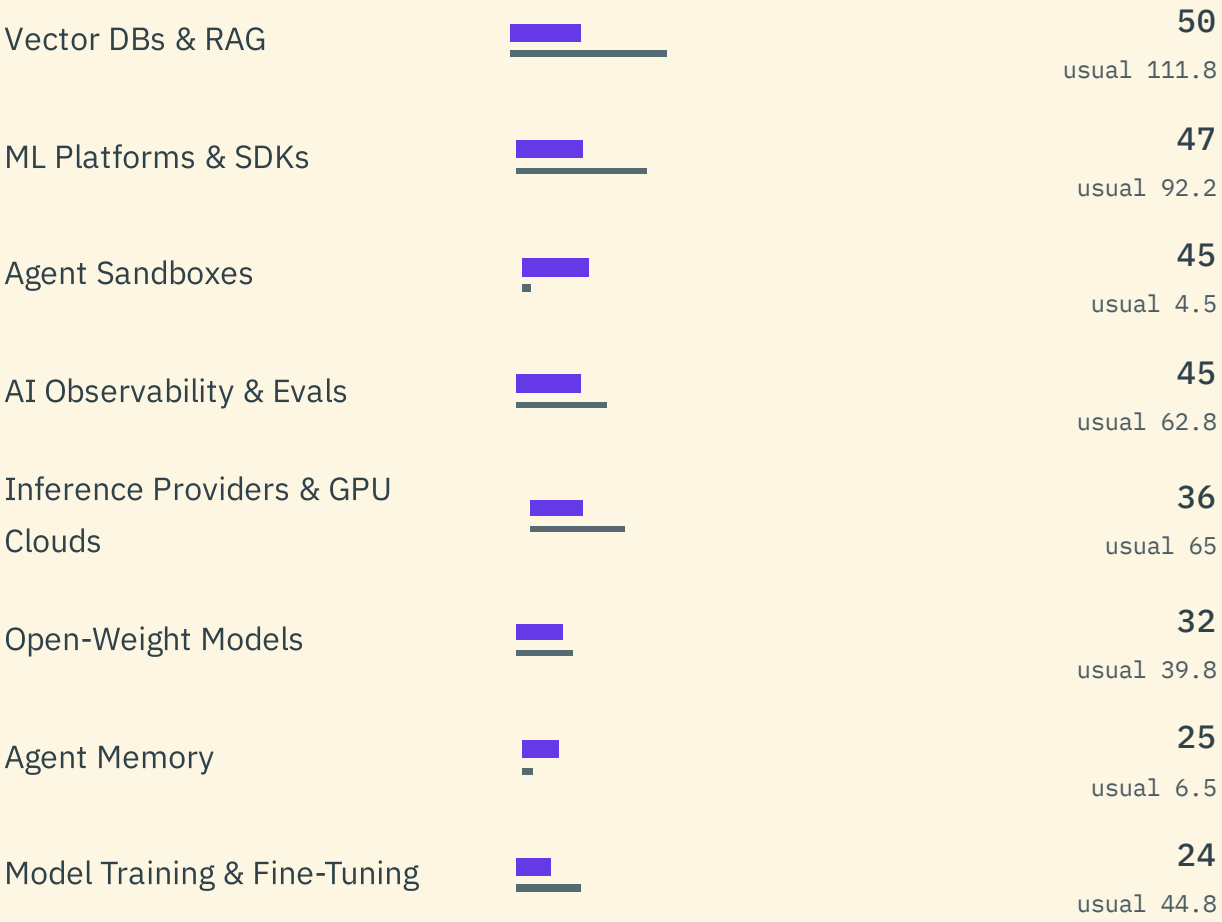
The thick bar is this week. The thin bar is the average of the four weeks before, counting only weeks with a daily.

AI Coding Agents
372
usual 304.8

AI Agent Frameworks
128
usual 250.8

MCP Servers & Tooling
100
usual 128

AI / LLM Gateways
64
usual 64



Who shipped the most · Nov 21–27

Changes per tool this week. Volume is not a quality score.



Upcoming conferences →

// END OF TRANSMISSION

The AI Toolchain · This Week's Highlights · November 28, 2026
The full firehose — every feature release, every day — is in the newsletter.

▷ Subscribe