

AI Features Shipped

Hi Alex, here's your personalized Deep Research Briefing on **AI Features Shipped**.

Quick summary: Agent quality monitoring has become table stakes this week, while the bigger shift is merchants and platforms moving agents from recommendations into actual transaction completion.

Product teams across the industry shipped a wide range of AI features this week, spanning agent quality monitoring, voice interfaces, commerce automation, and developer tooling. Below is a breakdown organized by the underlying problem each feature addresses.

Agent Reliability and Quality Control

A cluster of releases this week targets the same problem: AI agents that work in testing but degrade or misbehave once deployed.

Company	Feature	Problem solved
LaunchDarkly	AI Judge	Detects agent accuracy drift and toxicity spikes in milliseconds, catching silent quality failures before they compound in production
LaunchDarkly	Control Gap research	Surfaces that 91% of engineers distrust AI-generated code in production despite 94% saying it ships faster, framing the need for post-deploy safeguards
GitLab	Agent Session Blame (proof of concept)	Lets developers trace which Duo agent session wrote a given line of code, addressing demand for accountability as agent-generated commits scale
GitHub	Copilot agent PR growth response	Automated code review, validation, and rollback are being adopted as agent-written pull requests have grown 9x in size over eight months
Vercel	shadcn/lint	Enforces design-system rules on AI agents with explanatory errors, so agents stop generating UI that

violates approved components and themes

Runway	Real-time streaming safety moderation	Classifies video frames in under 500ms during live generation, closing the gap where users could see harmful output before a safety check completes
--------	---	---

Voice and Multimodal Interfaces

Several teams pushed real-time voice capability into agents and assistants this week.

Company	Feature	Problem solved
Google	Gemini 3.8 Live and Live Extended Thinking	Adds near-real-time visual processing and 97-language mid-conversation switching so voice agents can reason while speaking without breaking conversational flow
Google	Universal Speech Model and Gemini 3.5 Live Translate	Targets under-resourced languages via cross-lingual transfer, aiming to make speech AI work in markets with little training data
Vercel	Gemini 3.8 Live models added to AI Gateway	Gives developers WebSocket-based access to real-time audio agents without building custom audio pipelines
SpaceXAI	Grok Bot voice	Extends Grok Bot from text-only to spoken interaction, matching the shift toward voice as a primary agent modality
ElevenLabs	Reception by ElevenAgents	A 24/7 AI phone receptionist that answers business calls, books appointments, and supports 70+ languages, aimed at small businesses currently reliant on human receptionists or missed calls
Synthesia	Interactive Avatar Realtime API	Embeds conversational avatars that listen and respond live on websites, for support and training use cases where static video no longer suffices
Apple	Siri AI	A rebuilt assistant with personal context, onscreen awareness, and Visual Intelligence, addressing the gap between Siri and multimodal competitors like ChatGPT
Google	CC household agent	Coordinates calendars, tasks, and permission slips across up to six family members with granular sharing controls, solving multi-person household logistics rather than single-user assistance

Agentic Commerce and Payments

The clearest theme this week is agents completing real transactions, not just recommending them.

Company	Feature	Problem solved
OpenAI / Shopify	ChatGPT Ads with Shopify catalog sync ^{1,2}	Lets merchants appear in ChatGPT shopping conversations with zero-install catalog sync, addressing friction in manual ad creation for a new discovery channel
OpenAI	Sponsored Agents, Ads Manager , HubSpot integration	Lets shoppers converse with business agents after clicking an ad, and lets advertisers build campaigns via natural language instead of manual ad tools
Stripe	Link CLI plus Browserbase, endorsed as a solved pattern by Stripe's Jeff Weinstein ^{3,4}	Enables agents to complete web purchases end-to-end (demonstrated with a real restaurant reservation) using one-time-use cards to limit fraud exposure
Atom / Stripe / Tempo	Machine Payments Protocol domain purchasing	Lets agents search, appraise, and buy domains via an HTTP 402 payment flow with no checkout page, removing humans from a routine B2B transaction
Google	Business Agent for YouTube ads, Universal Commerce Protocol updates	Lets viewers ask product questions inside a YouTube ad without leaving the video, and adds cart-transfer support ahead of the holiday shopping season

Meta's Chief AI Officer Alexandr Wang used the week to publicly back a long list of Muse agent transactions rather than announce new features outright: an auto insurance switch saving \$3,500/year [in about five minutes](#), a telecom bill negotiation saving \$85.30/month over [a live phone call](#), an IKEA return [handled end to end](#), a dependent-care [FSA claim filed autonomously](#), [flight booking](#), [Facebook Marketplace shopping](#), CDL licensing [research and study planning](#), household logistics spanning shopping, [travel](#), [and language learning](#), and a compiled set of 100 use cases totaling [\\$9,649.71 in documented savings](#). These are Wang's endorsements of user-reported outcomes rather than distinct product launches, so they should be read as his perspective on what Muse can do rather than as separate shipped features.

Enterprise Agent Platforms

Several vendors expanded agents from single tasks into orchestrated, multi-system workflows.

Company	Feature	Problem solved
Salesforce	Agentforce Operations, with Marshall and Systems Integration Agent	Automates multi-step back-office processes (invoice matching, approvals) deterministically across disconnected systems, and learns browser-based apps without APIs in about 90 minutes, targeting compliance-critical work that general agents cannot audit
Salesforce	Long-horizon agents (Hunter)	Gives agents persistent memory and plan durability across multi-week goals, so an outbound sales agent can course-correct when an email bounces instead of restarting
Salesforce	Builder Central	A no-code workspace for building apps and agents in plain language, aimed at the 88% of IT leaders who say they cannot meet internal app demand safely
Salesforce	Koa reasoning model	A task-specific model built on Nvidia Nemotron and trained on synthetic data, offering lower token cost and data sovereignty as an alternative to routing every sales and support task through a frontier model
Salesforce	Alforce, with Claudeforce and Slackforce	Carries Salesforce permissions and governance into external AI interfaces like Claude and Slack, so agents built elsewhere still respect existing security rules
Zendesk	Specialized Agents	Purpose-built agents for specific workflows, addressing CEO Tom Eggemeier's view that generic AI automates without understanding domain constraints
Asana	Agentic Work Management, 30+ AI Teammates, Asana Dash	Bundles prebuilt agents into every paid tier and adds a personal "chief of staff" agent, lowering the barrier to agent adoption inside existing workflows
Okta	Dex internal agent	An orchestrating agent for IT, people, travel, and expense support, registered in Okta's own directory so access and actions stay auditable; reached 25% internal adoption in two weeks

Monday.com	Notebook interface for agents ⁵ , Org Agent Directory ⁶ , programmatic Vibe app generation ⁷	Together these address agent transparency (showing reasoning), duplication (letting teams discover and reuse existing agents instead of rebuilding), and static tooling (letting agents generate custom interfaces on the fly)
ClickUp	Super Agent Artifacts ⁸ , Brain² Outlook integration ⁹ , cloneable Super Agents ¹⁰	Extends agents from task automation into generating custom software, bridges ClickUp work into Outlook scheduling and inbox triage, and lets teams template agent patterns instead of rebuilding them

Developer Tooling and Coding Agents

Company	Feature	Problem solved
Anthropic	Claude Code AGENTS.md fallback	Aligns Claude Code with an emerging community convention for agent configuration files, easing migration for teams not using CLAUDE.md
Anthropic	Docs and slides as shareable artifacts	Extends Claude's design-artifact capability into documents and presentations, so teams can draft a spec, gather feedback, and build without leaving the chat interface
Anthropic	Claude Plugin Eval	Lets plugin developers score performance against test suites and a no-plugin baseline, closing a gap in validating whether a plugin actually helps
Anthropic	Claude Mods	An extensibility system for community-built modifications, moving Claude toward a platform model rather than a closed tool
Anthropic	Salesforce plugin, 37 sales skills	Brings CRM pipeline data directly into Claude for call prep and forecasting, so sellers stop switching tools mid-workflow
Vercel	AI SDK harness layer native subscription auth	Lets developers switch between eight coding agents (Claude Code, Cursor, GitHub Copilot, and others) without changing code, keeping credentials on the host boundary rather than per-agent
Vercel	v0 model-agnostic routing	CEO Guillermo Rauch frames hiding model choice as a source of developer confusion; v0 now routes across Claude, GPT, Kimi, GLM, Grok, and DeepSeek via AI Gateway

Vercel	mcp-handler WebMCP support	Lets browser-based agents call MCP tools via a single script tag without browser-side OAuth, easing in-page agent integrations
Vercel	fx safety classifier benchmark	Found a candidate replacement model 5 to 18 times faster than the current classifier that reviews every command in fx's auto mode
Vercel	GLM 5.3 FlashX added to AI Gateway	Adds a roughly 200 tokens/second inference option aimed at reducing latency in coding agents and interactive tool loops
GitHub	AI Scan decoupled from CodeQL	Removes the requirement to run full CodeQL setup just to get AI-driven vulnerability detection on pull requests
GitHub	Copilot impact dashboard	Gives enterprise admins per-feature adoption data so they can see which Copilot capabilities are used and which need promotion
GitHub	npm stage-only token permission	A narrower token scope lets automation stage package versions for review without granting full publish rights
GitLab	Open-weight models on Duo Agent Platform	CEO Bill Staples argues the real unit of cost is now "cost per accepted change," not per-token price; new models deliver up to 4x more calls per credit
Atlassian	Bitbucket Cloud code search (open beta)	Catches Bitbucket up to table-stakes workspace-wide code search already common on GitHub and GitLab
Atlassian	BearQ testing agent in Jira	An assignable agent that connects test execution to work planning, targeting high-maintenance, fragile test suites
Atlassian	CLI for design system context	Cuts token costs and speeds agent tasks 8% by giving agents design-system context from the terminal alongside the existing MCP server
Zed	Delta public beta , shared thread links in browser	Replaces pull requests with threaded, real-time collaboration; Zed's own team disabled PRs internally and landed 570 changes to main through Delta instead

Data, Analytics, and Infrastructure

Company	Feature	Problem solved
---------	---------	----------------

Snowflake	Catalog-Linked Database	Queries external Iceberg data through Snowflake compute without copying it, letting teams keep engineering on Databricks/Spark while using Snowflake for analytics and AI
Snowflake	Per-user AI quotas (GA)	Adds automatic spend blocking at the individual level across AI Functions, Cortex Agents, and CoWork, closing a cost-governance gap for shared AI platforms
Snowflake	Redpoint Identity Studio	A transparent, tunable identity-resolution Native App that runs on the customer's own compute, positioned against opaque vendor tools for regulated industries
WorkOS	SCIM directory Logs tab	Surfaces 30 days of provisioning requests with debug messages, addressing the problem that SCIM failures happen between the identity provider and WorkOS where neither side alone can see the full picture
Cohere	Model Vault confidential computing	Encrypts inference while it runs, not just at rest or in transit, targeting regulated industries where standard cloud inference is a compliance blocker
PostHog	Twelve Data Warehouse sources reach GA	Moves integrations like Close, Pipedrive, and SendGrid out of prerelease, giving teams stable SLAs for data consolidation

Vertical AI Tools

Company	Feature	Problem solved
Harvey	User memory, Agentic Vault Search, Review Table Agent Actions , multi-source queries	Stores individual drafting and negotiating preferences, searches entire data rooms with page-level citations, and restructures large document review sets via natural language for non-technical legal staff
Gusto	Expanded Claude connector	Moves the Claude integration from read-only payroll lookups into write actions like bonus setup, timesheet checks, and salary calculations
Ramp	Accounting Bench	A 137-task benchmark showing the best frontier model hits only 21% accuracy on real accounting workflows, establishing that reliable agentic accounting is still unsolved
Netflix	ja Java tooling suite	Composable command-line tools (jig, jfmt, jist, jdocserver) that give coding agents dependency

and documentation context without requiring an IDE or language server

HeyGen	MCP connector for Claude	Lets Claude trigger HeyGen video generation directly from a prompt, removing manual export and file-transfer steps between script and finished video
ElevenLabs	Expanded MCP with voice , music , image , video	Consolidates ElevenLabs's creative stack into a single connector usable from Claude, ChatGPT, Cursor, and Hermes, so a single brief can produce a script, voiceover, music, and video without leaving the chat
Cursor / Meta	WhatsApp Business Tools MCP	Lets coding agents automate WhatsApp Business account setup and template creation, consolidating tasks previously scattered across several Meta consoles
Google	MCP server for Google Home	Lets third-party agents like Claude and ChatGPT control connected home devices and review camera summaries via natural language, opening Google Home beyond its own Gemini assistant

Creative and Design Tools

Company	Feature	Problem solved
Figma	Weave tools publishable to Community	Lets users share generative AI workflows with others inside Figma Design, similar to a plugin marketplace, though only image and vector tools currently work inside Design itself
Runway	Real-time sketch animation app	Generates video frames live as a user draws, demonstrating that real-time video inference can hit interactive latency, not just batch generation
Runway	Fall 2026 model collection	Bundles third-party models (Fish Audio, MiniMax, Cartesia, Flux) alongside Runway's own tools, positioning Runway as a distribution hub rather than a single-model product
Runway	Unified credit pricing across web and Dev API	Merges billing so enterprise customers use one credit pool across the visual interface and the programmatic API instead of managing two separate contracts
Notion	Native iOS Calendar rebuild	Improves cold-start performance and adds day/week/month/agenda views, replacing a web-based implementation that lagged native competitors

Notion	Sub-agent delegation in Custom Agents	Lets a primary agent hand off focused work to sub-agents with their own model and access settings, so cheaper models handle simple steps and costlier ones are reserved for reasoning
Notion	DeepSeek V4.1 Flash added as a model option	Expands open-weight model choice inside Personal and Custom Agents
Linear	Markdown and text files as embedded content	Removes friction from managing documentation and code snippets inside issues by embedding.md and.txt files as content instead of attachments

Reflections From Product Leaders (Perspective, Not Announcements)

A few voices this week described strategic direction rather than shipped features. Zapier co-founder Mike Knoop pointed to AI-native programming languages like Probably as a step toward primitives designed for LLM reasoning [rather than retrofitted logic](#). Harvey co-founder Gabriel Pereyra said the company is evaluating the Jev model as a replacement for its internal routing and citation classifiers because of [better-calibrated probability estimates](#). Runway co-CEO Cristóbal Valenzuela argued that video generation is heading [toward real-time world simulation](#) and warned that AI meeting-recording tools risk creating "slop listening," where attendees disengage and shift cognitive work downstream to whoever processes the summary later. Coinbase President Emilie Choi described restructuring toward flatter, IC-heavy teams to ship faster with AI while [guarding against quality loss](#). Glean founder Arvind Jain argued that enterprise AI's value depends on grasping organizational context, not just [explicit documents and tickets](#). None of these reflect a specific feature shipped this week, but they frame why several of the launches above were built the way they were.

[Read the full report](#)



Researched and written by AI on the Department of Product Intelligence platform. It cites its sources, and they are worth checking before you act on anything here.

You're receiving this because a teammate set up a recurring research briefing ("AI features shipped") and added you as a recipient. [Unsubscribe from this briefing](#) · [Manage briefings](#)