

Daily AI Release Roundup — September 17, 2026

This is an automated daily roundup of notable AI model and product releases from frontier and open-weight labs, covering the last 24 hours (September 16–17, 2026). Findings are drawn from live web research; no releases or benchmark figures are invented.

- **GPT-6 Astra (OpenAI)** — First OpenAI model classified 'Critical' for cybersecurity capability under OpenAI's Preparedness Framework; Microsoft made it generally available in Azure AI Foundry Models the same day (Sep 17). Pricing: \$10/\$50 per million tokens (input/output) standard tier, rising to \$20/\$75 on long context; 10% premium for US Data Zone.
- **DeepSeek-V4.1-Flash (DeepSeek)** — New multimodal Mixture-of-Experts model with a novel asymmetric Causal Encoder-Decoder architecture, 1M-token context window, MIT-licensed open weights. Scored 74.2 on DeepSWE v1.1 and 90.6 on Terminal-Bench 2.1; API priced at \$0.15/\$0.60 per million tokens off-peak (50% of peak rates).
- **'One Claude' interface + Claude Docs/Slides (Anthropic)** — Anthropic merged its Claude Chat and Cowork products into a single interface (auto-routes tasks, no more tab-switching), integrated Claude Design app-wide, and launched Claude Docs and Claude Slides in beta — real-time collaborative documents/presentations exportable to Word/PowerPoint/PDF. Rolling out first to Pro/Max users on web, desktop, mobile.
- **HiDream-O1-Video-1.0 (HiDream.ai, China)** — New native omnimodal video generation model supporting text/image/video input, generating 1080p video 5–20 seconds long with synchronized audio and content-adaptive duration. Ranked #4 globally on the Artificial Analysis Image-to-Video leaderboard (with audio) and #8 on Arena.ai's Image-to-Video leaderboard. Currently in internal testing.
- **Koa (Salesforce + Nvidia)** — New CRM-specialized reasoning model for Agentforce, built by post-training Nvidia's Nemotron 3 Super open model on synthetic data modeled on ~30 years of Salesforce CRM deployments. Unveiled at Dreamforce 2026; in pilot now, broader US availability expected winter 2026.
- **Grok Bot Galaxy / Grok 4.8 (xAI)** — xAI ran a 3-day 'Grok Bot Galaxy' event (Sep 15–17) showcasing agentic 'Grok Bot' workflows for sales, engineering, marketing teams; separately, xAI is finishing pre-training on a next model (reportedly ~4.8, a step above Grok 4.7) — not yet released, still in training/RL phase.

NOTE: No notable same-day (last 24h) releases were confirmed from Google DeepMind, Meta, Mistral, Alibaba/Qwen, or Tencent — their most recent shipped models (Gemini 3.8 Flash, Muse Spark 1.3, Qwen3.8-Max, etc.) are more than 24 hours old as of this report.