

tech Gazette

The Monthly

September 2026

Software and AI, straight from the source.

The stories that moved it	2
What stops working next	3
Exploited, and down	4
Models, and what developers starred	5
Research worth the month	6

5,589 PIECES PRINTED	1,081 SOURCES	10 EXPLOITED VULNERABILITIES	52 DEADLINES AHEAD
--	-------------------------------------	--	--

Every line in this report is a piece a primary source published, in its words, with its link: the labs, companies, projects and people who did the thing. No algorithm chose it; nothing here is sponsored.

THE MONTH

The stories that moved it

The Weekly Reviews' top stories of the month, ranked by how many sources reported them and how much they change. Each links to where it was first said.

1**Claude Opus 5.5**

CLAUDE API RELEASE NOTES · AND 10 MORE SOURCES

We've launched Claude Opus 5.5 (claude-opus-5-5), a model for long-running agentic coding and knowledge work. It has a 1M token context window by default, 128k max output tokens, and always-on adaptive thinking, at \$4 /...

2**Released GPT-6 Sol (gpt-6-sol) and GPT-6 Luna (gpt-6-luna).**

OPENAI API CHANGELOG · AND 7 MORE SOURCES

Feature · gpt-6-sol · gpt-6-luna · v1/responses · v1/chat/completions. Released GPT-6 Sol (gpt-6-sol) and GPT-6 Luna (gpt-6-luna). These reasoning models accept text and image inputs and generate text through the Respons...

3**Gemini 3.8 Flash TTS and Gemini 3.8 Flash-Lite TTS generally available (GA)**

GEMINI API CHANGELOG · AND 3 MORE SOURCES

Gemini 3.8 Flash TTS and Gemini 3.8 Flash-Lite TTS generally available (GA): Released our next-generation text-to-speech (TTS) audio models and the Gemini API Voices endpoint (/v1beta/voices): Gemini 3.8 Flash TTS (gemin...

4**Grok 4.7**

XAI · AND 1 MORE SOURCE

Grok 4.7, SpaceXAI's frontier model for coding, agentic tasks, and knowledge work, is now available on the xAI API as grok-4.7. It has a 500k context window, text and image inputs with text-only output, and no text outpu...

5**Cache diagnostics is out of beta on the Claude API and no longer requires the cache-diagnosis-2026-04-07 beta header.**

CLAUDE API RELEASE NOTES

Cache diagnostics is out of beta on the Claude API and no longer requires the cache-diagnosis-2026-04-07 beta header. Include the diagnostics object on a Messages request to opt in; requests that still send the header wo...

6**Designing agent-first platforms: What changes when agents do the work**

MICROSOFT AZURE

The organizations pulling ahead are not simply adding AI to what they already have. They are designing for a different kind of software. The post Designing agent-first platforms: What changes when agents do the work appe...

7**Confused Deputy: The Old Bug That AI Agents Keep Reintroducing**

AUTHO

A category of vulnerability we aimed to eradicate is re-emerging with the rise of AI agents.

8**Build your way: Use any AI agent of your choice in Android Studio**

ANDROID DEVELOPERS

Posted by Matthew Warner, Product Manager, Android Developer Experience AI-powered developer tools have become an essential multiplier for engineering productivity, with teams adopting specialized AI coding agents, custo...

ACT ON THIS

What stops working next

Models being shut down and versions reaching end of support in the 90 days after Sep 30, with the replacement each vendor recommends. Only dates they state.

WHEN	WHAT STOPS	USE INSTEAD	STATED BY
Sep 30	labs-leanstral-1-5 shuts down		Mistral AI
Sep 30	mistral-ocr-4-0 shuts down	OCR4.1	Mistral AI
Oct 1	gpt-5.4-cyber shuts down	The most capable cyber model available to you.	OpenAI
Oct 1	Python 3.10 reaches end of support		Python (PEP 602 release cycle)
Oct 5	antigravity-preview-05-2026 shuts down	antigravity-preview-09-2026	Google (Gemini API)
Oct 20	gemini-2.5-flash shuts down	gemini-3.8-flash or gemini-3.5-flash-lite or gemini-3.1-flash-lite	Google Cloud (Vertex AI)
Oct 20	gemini-2.5-flash-lite shuts down	gemini-3.8-flash or gemini-3.1-flash-lite or Gemma 4	Google Cloud (Vertex AI)
Oct 20	gemini-2.5-pro shuts down	gemini-3.8-flash or gemini-3.5-flash	Google Cloud (Vertex AI)
Oct 22	gemini-omni-flash-preview shuts down	gemini-omni-1.1-flash	Google (Gemini API)
Oct 22	veo-3.1-fast-generate-preview shuts down	gemini-omni-1.1-flash	Google (Gemini API)
Oct 22	veo-3.1-generate-preview shuts down	gemini-omni-1.1-flash	Google (Gemini API)
Oct 22	veo-3.1-lite-generate-preview shuts down	gemini-omni-1.1-flash	Google (Gemini API)
Oct 23	deep-research-pro-preview-12-2025 shuts down	deep-research-preview-04-2026	Google (Gemini API)
Oct 23	ft-babbage-002 shuts down	gpt-5.6-terra	OpenAI
Oct 23	ft-davinci-002 shuts down	gpt-5.6-terra	OpenAI
Oct 23	ft-gpt-3.5-turbo shuts down	gpt-5.6-terra	OpenAI
Oct 23	ft-gpt-4 shuts down	gpt-5.6-sol	OpenAI
Oct 23	ft-gpt-4.1-nano-2025-04-14 shuts down	gpt-5.6-luna	OpenAI

AND 34 MORE DEADLINES AT [TECHGAZETTE.COM.BR/DEADLINES/](https://techgazette.com.br/deadlines/)

SECURITY DESK

Exploited, and down

Exploited in the wild

CVE-2026-7273: Zyxel GS1900 Series Switches Stack-Based Buffer Overflow Vulnerability

CISA KEV · SEP 21

CVE-2026-93952: Arista VeloCloud Orchestrator Improper Input Validation Vulnerability

CISA KEV · SEP 22

CVE-2026-94127: F5 BIG-IP APM Heap-based Buffer Overflow Vulnerability

CISA KEV · SEP 22

CVE-2026-93616: Check Point Multiple Products Path Traversal Vulnerability

CISA KEV · SEP 22

CVE-2026-85102: Check Point Multiple Products Improper Certificate Validation Vulnerability

CISA KEV · SEP 22

CVE-2026-5430: WSO2 Multiple Products Path Traversal Vulnerability

CISA KEV · SEP 24

CVE-2026-71362: Adobe Commerce and Magento Incorrect Authorization Vulnerability

CISA KEV · SEP 24

CVE-2026-67279: Mikrotik RouterOS Improper Enforcement of Behavioral Workflow Vulnerability

CISA KEV · SEP 25

CVE-2026-65660: Microsoft SharePoint Code Injection Vulnerability

CISA KEV · SEP 25

CVE-2026-87902: WordPress Core Remote File Inclusion Vulnerability

CISA KEV · SEP 25

The longest incidents

Kimi K3 Model API outage

BASETEN · SEP 20

Errors accessing Netlify-hosted sites

NETLIFY · SEP 21

Confluent Cloud Console Degraded (All Regions) and Multiple Services Impacted in AWS us-west-2

CONFLUENT CLOUD · SEP 21

Elevated inference 500s for a subset of models deployed in GCP central1

BASETEN · SEP 21

Elevated rate of failed iOS submissions

EXPO · SEP 21

Elevated rate of failed iOS submissions

EXPO · SEP 21

Elevated errors for multiple models

CLAUDE · SEP 22

Aura-2 TTS Elevated Error Rates for English Voices

DEEPGRAM · SEP 22

Elevated error rates in Azure Germany West Central region

CONFLUENT CLOUD · SEP 22

[Serverless][AWS][eu-west-1] 5xx on read requests

PINECONE · SEP 22

WHAT SHIPPED, WHAT ROSE

Models, and what developers starred

Models released (74)

New model: HeyGen: HeyGen Video

OPENROUTER · SEP 30

New model: Together: Tev1 4B Experimental

OPENROUTER · SEP 30

Open model: Cactus-Compute/whistle

HUGGING FACE HUB · SEP 30

New model: VoyageAI by MongoDB: rerank-3-lite

OPENROUTER · SEP 30

New model: VoyageAI by MongoDB: rerank-3

OPENROUTER · SEP 30

Open model: RunningHubAI/rh-qwen-image-2.1-lora-2104918997757157378

HUGGING FACE HUB · SEP 30

New model: OpenAI: GPT-6.1 Sol Pro

OPENROUTER · SEP 29

New model: OpenAI: GPT-6.1 Sol

OPENROUTER · SEP 29

Open model: kyutai/ovie-512

KYUTAI · SEP 29

Open model: PSRben/VisionHOPE

HUGGING FACE HUB · SEP 29

Open model: internlm/AdvancedMathBench-AutoVerifier

INTERNLN (SHANGHAI AI LAB) · SEP 29

New model: Anthropic: Claude Sonnet 5.5

OPENROUTER · SEP 28

AND 62 MORE MODELS AT
[TECHGAZETTE.COM.BR/MODELS/](https://techgazette.com.br/models/)

Rising on GitHub

GitHub's Trending weeks for this month aren't kept yet.

FROM THE PAPERS

Research worth the month

1,770 papers printed this month. A selection, each with its own abstract's words; the week's papers by field are at techgazette.com.br/research.

ChunkRank: Model-Aware Text Chunking and Abstention-Aware Answer Selection for LLM Pipelines

ARXIV CS.CL · SEP 24

LLM Agents Can Easily Tamper With Their Own Traces

ARXIV CS.CR · SEP 24

To Trust or Not to Trust: Retrieval-Augmented Fact Checking in Speech

ARXIV CS.LG · SEP 24

Instrumental Monitor Evasion Emerges Under Ordinary Task Pressure

ARXIV CS.CR · SEP 24

VietPrism: A large-scale Vietnamese speech and deepfake corpus with diverse dialects and code-switching

ARXIV CS.CL · SEP 24

tech Gazette

The paper, every morning. Software and AI news from the people who made it. Free, no algorithm, no account. It ends.

Read it at techgazette.com.br, or get it in your inbox:
techgazette.com.br/newsletter.

The Monthly is made from the paper's own archive: 5,589 pieces from 1,081 sources in September 2026. Titles are the sources' own. Share it freely, whole and unchanged.