INSIGHT DAILY

2026-10-28

Three-series written report and verified spoken scripts

Daily report

AI Daily

39 source-linked editorial stories.

1. GitHub Copilot supports server-to-server authentication

GitHub Copilot now supports server-to-server authentication for GitHub Apps installed on organizations, available in public preview.

Evidence: Services and agents built with the GitHub Copilot SDK can now make requests on behalf of an organization without relying on a signed-in user’s credentials.

Impact: Teams can evaluate the documented GitHub Copilot workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: github.blog · https://github.blog/changelog/2026-10-27-github-copilot-supports-server-to-server-authentication

2. OpenAI: Texting your dot

OpenAI: Texting your dot.

Evidence: The OpenAI ChatGPT release notes page documents the OpenAI update.

Impact: Teams can evaluate the documented OpenAI workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: help.openai.com · https://help.openai.com/en/articles/6825453-chatgpt-release-notes

3. Nous Research: Every new plan now includes a cloud agent out of the box, and usage has daily and weekly limits instead of one monthly block...

Every new plan now includes a cloud agent out of the box, and usage has daily and weekly limits instead of one monthly block of credits.

Evidence: The X @NousResearch page documents the Nous Research update.

Impact: Teams can evaluate the documented Nous Research workflow before wider rollout.

Watch next: watch pricing, quota, and account limits for Nous.

Source: x.com · https://x.com/NousResearch/status/2115120531611787323

Source: portal.nousresearch.com · https://portal.nousresearch.com/billing

4. 3 Ways to Boost MoE Inference on NVIDIA DGX Spark

Generating tokens faster starts with identifying what limits inference performance.

Evidence: In the batch-size-1 evaluation of Ling-3.0-flash from Inclusion AI on NVIDIA DGX Spark, profiling showed that decoding was memory-bound: reading model weights drove execution time.

Impact: Teams can evaluate the documented NVIDIA workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: developer.nvidia.com · https://developer.nvidia.com/blog/3-ways-to-boost-moe-inference-on-nvidia-dgx-spark/

5. MAI Image 2.7 debuts above GPT Image 2.5 Flare on Artificial Analysis and #3 on Arena

MAI Image 2.7 debuts above GPT Image 2.5 Flare on Artificial Analysis and #3 on Arena.

Evidence: The source records 2.7 / 2.5.

Impact: Teams can evaluate the documented Microsoft workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: microsoft.ai · https://microsoft.ai/news/mai-image-2-7-debut

6. AMD and Meta Advance Strategic Partnership to Power Meta's Superintelligence Infrastructure

AMD and Meta Advance Strategic Partnership to Power Meta's Superintelligence Infrastructure.

Evidence: The AMD Press Releases release records Day-0 for Amd.

Impact: Teams can evaluate the documented Amd workflow before wider rollout.

Watch next: watch availability, limits, and real usage for Amd.

Source: ir.amd.com · https://ir.amd.com/news-events/press-releases/detail/1302/amd-and-meta-advance-strategic-partnership-to-power-metas-superintelligence-infrastructure

7. 6.1 Sol Degraded Model Performance

Status: Resolved All affected services are operating normally again.

Evidence: The OpenAI Status release records 6.1 for OpenAI.

Impact: Teams can evaluate the documented OpenAI workflow before wider rollout.

Watch next: Watch OpenAI Status for recovery confirmation and further OpenAI incidents.

Source: status.openai.com · https://status.openai.com//incidents/01M5YY4A3TQETEQ53DVSK8SW10

8. CohortFlow: A Steerable Clinical Cohort Curation Agent with Localized Semantic Correction

Language model agents for clinical cohort curation should accommodate evolving eligibility criteria without repeatedly processing unchanged requirements.

Evidence: arXiv introduce CohortFlow, a steerable agent that represent.

Impact: Teams can evaluate the documented arXiv workflow before wider rollout.

Watch next: watch docs, endpoint access, and developer limits.

Source: arxiv.org · https://arxiv.org/abs/2610.36100

9. Introducing artifacts for legal work with Claude

Since we launched Claude for Legal in May, the most common requests from legal teams have been for simpler setup and easier ways to work with colleagues.

Evidence: Artifacts are interactive tools, such as trackers and dashboards, that teams can use and share in Claude.

Impact: Teams can evaluate the documented Claude workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: claude.com · https://claude.com/resources/articles/introducing-artifacts-for-legal-work

Source: the-decoder.com · https://the-decoder.com/claude-for-legal-now-verifies-sources-automatically-then-hands-the-file-to-a-lawyer/

10. Hill-Climbing Agent Quality with Agents CLI Evaluation

Building an agent that works on a demo is a different problem from building an agent that stays working across model upgrades, tool changes, and new conversation shapes.

Evidence: The first takes a weekend; the second takes a loop.

Impact: Teams can evaluate the documented Google workflow before wider rollout.

Watch next: watch docs, endpoint access, and developer limits.

Source: developers.googleblog.com · https://developers.googleblog.com/hill-climbing-agent-quality-with-agents-cli-evaluation/

11. OpenAI expands partnership with Oracle to bring its apps to ChatGPT

OpenAI and Oracle are expanding their partnership to bring business applications and AI agents into ChatGPT and power new Oracle offerings.

Evidence: The OpenAI News page documents the OpenAI update.

Impact: Teams can evaluate the documented OpenAI workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: openai.com · https://openai.com/index/oracle-apps-in-chatgpt

12. Retell uses OpenAI models to help power 100M monthly calls

Its platform now serves more than 4,000 businesses, with monthly call volume growing from 30 million in early 2026 to more than 100 million by September.

Evidence: As Retell serves more enterprises with complex workflows, it uses OpenAI models to keep call quality and reliability high.

Impact: Teams can evaluate the documented OpenAI workflow before wider rollout.

Watch next: watch license, setup path, and benchmark.

Source: openai.com · https://openai.com/index/retell-ai

13. The platform approach to becoming Frontier

Last year at Ignite, we introduced the concept of Frontier Transformation and how it is different from any technology-led pursuit, because it requires us to focus on business outcomes first.

Evidence: It starts with a simple idea: technology was invented to serve humanity, not the other way around.

Impact: Teams can evaluate the documented Microsoft workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: blogs.microsoft.com · https://blogs.microsoft.com/blog/2026/10/27/the-platform-approach-to-becoming-frontier

14. Anyone can now bring their ideas into the company

Anyone can now bring their ideas into the company.

Evidence: The Replit release records generally available for Replit.

Impact: Teams can evaluate the documented Replit workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: replit.com · https://replit.com/blog/anyone-can-bring-ideas-to-company-codebase

15. OpenAI: Much requested, and now yours to connect: iCloud Mail and Calendar plugins are available in ChatGPT.

Much requested, and now yours to connect: iCloud Mail and Calendar plugins are available in ChatGPT.

Evidence: Ask ChatGPT to find that email, catch up on your inbox, check what’s next, or update a calendar event. Available on all ChatGPT plans on web, desktop, and iOS.

Impact: Teams can evaluate the documented OpenAI workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: x.com · https://x.com/ChatGPT/status/2115139248458191070

16. CrowdStrike: When AI Guardrails Expose Vulnerabilities

Safety and capability are usually framed as a trade-off.

Evidence: This research suggests the opposite: Every improvement labs make to their guardrails also sharpens a methodology that cybersecurity practitioners or adversaries can use.

Impact: Teams can evaluate the documented CrowdStrike workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: www.crowdstrike.com · https://www.crowdstrike.com/en-us/blog/when-ai-guardrails-expose-vulnerabilities/

17. Muse Is Now Available in Singapore: The World's First Personal AI Agent Built for Everyone

Connector features will roll out progressively, with some experiences becoming available after launch.

Evidence: You can ask Muse to research products and compare prices before buying on Shopee or Lazada, or plan grocery shopping and dining through GrabMart and Grab Dine Out.

Impact: Teams can evaluate the documented Meta workflow before wider rollout.

Watch next: Watch the Meta release for patch notes and rollout issues.

Source: about.fb.com · https://about.fb.com/news/2026/10/muse-is-now-available-in-singapore-the-worlds-first-personal-ai-agent-built-for-everyone/

18. An update from Amazon engineering leadership on the role of the developer in the age of AI

One of the biggest shifts its leadership teams are seeing right now is how AI is changing the way software development engineers (SDEs) do their work, while balancing that the important elements of being an SDE are the same.

Evidence: Engineering leadership, including Distinguished and Senior Principal Engineers, from across Amazon recently came together to reaffirm what it means to be an SDE at Amazon.

Impact: Teams can evaluate the documented Amazon workflow before wider rollout.

Watch next: watch docs, endpoint access, and developer limits.

Source: www.aboutamazon.com · https://www.aboutamazon.com/news/company-news/amazon-software-engineer-role-guideline-ai?utm_source=rss

19. Anysphere: New in Cursor design mode: Annotate.

Give agents visual prompts: edit elements, draw on the page, and review your designs at different screen sizes.

Evidence: The X @cursor_ai page documents the Anysphere update.

Impact: Teams can evaluate the documented Anysphere workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: x.com · https://x.com/cursor_ai/status/2115159585548021814

Source: cursor.com · https://cursor.com/docs/agent/design-mode

20. When Retrieval Is Not Enough: Obsolete Facts in Agent Memory

Memory-agent benchmarks often collapse retrieval and answer generation into one score, making it hard to tell whether a failure comes from missing an update or from not using an update that was already retrieved.

Evidence: The arXiv coding-agent research feed release records 2610.36308v1 for arXiv.

Impact: Teams can evaluate the documented arXiv workflow before wider rollout.

Watch next: Watch the arXiv release for patch notes and rollout issues.

Source: arxiv.org · https://arxiv.org/abs/2610.36308

21. Coding agent setup & developer resources | Gemini API

Coding agent setup & developer resources | Gemini API.

Evidence: The Google Gemini API Changelog page documents the Google update.

Impact: Teams can evaluate the documented Google workflow before wider rollout.

Watch next: watch docs, endpoint access, and developer limits.

Source: ai.google.dev · https://ai.google.dev/gemini-api/docs/coding-agents

22. FrontierCode 2.0

Most existing coding benchmarks fail to capture these gaps and tend to be saturated quickly.

Evidence: Cognition is excited to introduce FrontierCode 2.0, a new generation of its coding benchmark series and its best attempt yet to evaluate user experience.

Impact: Teams can evaluate the documented Cognition workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: cognition.com · https://cognition.com/blog/frontier-code-2.0

Source: x.com · https://x.com/cognition/status/2115132493259321443

23. MAI Cyber 1.1 Flash: Sharper findings and fewer tokens inside MDASH

MAI Cyber 1.1 Flash builds on that foundation.

Evidence: It's a more focused model that takes on more of the work itself, so the full MDASH system can surface useful findings at lower cost.

Impact: Teams can evaluate the documented Microsoft workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: microsoft.ai · https://microsoft.ai/news/mai-cyber-1-1-flash

Source: microsoft.ai · https://microsoft.ai/pdf/MAI-Cyber-1.1-Flash%20ORA%20Model%20Card.pdf

24. NVIDIA and Oracle Expand NVIDIA Vera Rubin to Multi-Gigawatt Scale and Bring NVIDIA Vera CPU to Oracle Cloud Infrastructure

NVIDIA and Oracle today announced plans to expand Oracle Cloud Infrastructure (OCI) Superclusters with NVIDIA Vera Rubin NVL72 systems to multi-gigawatt scale, supporting growing customer demand for AI training, inference and agentic applications.

Evidence: The NVIDIA Newsroom page documents the NVIDIA update.

Impact: Teams can evaluate the documented NVIDIA workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: nvidianews.nvidia.com · https://nvidianews.nvidia.com/news/nvidia-and-oracle-expand-nvidia-vera-rubin-to-multi-gigawatt-scale-and-bring-nvidia-vera-cpu-to-oracle-cloud-infrastructure

25. Announcing SOCI-indexed images for faster SageMaker AI jobs

Every minute your deep learning container image spends downloading is a minute you pay for and a minute your team waits.

Evidence: Container image download is pure overhead: billable job time that produces no model progress.

Impact: Teams can evaluate the documented Amazon workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: aws.amazon.com · https://aws.amazon.com/blogs/machine-learning/announcing-soci-indexed-images-for-faster-sagemaker-ai-jobs

26. Memory-Enforced GPU Sharing in Kubernetes with NVIDIA CUDA 13.4

Multi-instance GPU (MIG) splits a GPU into hardware-isolated instances that are transparent to the application, time-slicing lets several processes take turns without memory enforcement, and Multi-Process Service (MPS) lets them run concurrently in separate address spaces (sm-sharing).

Evidence: CUDA 13.4 now enforces a memory boundary that is transparent to the application with a new time-sharing mode, alongside the compute sharing mode (sm-sharing) that already existed in MPS.

Impact: Teams can evaluate the documented NVIDIA workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: developer.nvidia.com · https://developer.nvidia.com/blog/memory-enforced-gpu-sharing-in-kubernetes-with-nvidia-cuda-13-4/

27. AWS: SageMaker MLflow Apps now support MLflow 3.16

Amazon SageMaker AI MLflow Apps now support MLflow 3.16, bringing a redesigned tracing experience and faster experiment tracking to your generative AI and machine learning workflows.

Evidence: MLflow 3.16 improves debugging and analysis of generative AI applications and agents.

Impact: Teams can evaluate the documented AWS workflow before wider rollout.

Watch next: Watch AWS What's New for recovery confirmation and further AWS incidents.

Source: aws.amazon.com · https://aws.amazon.com/about-aws/whats-new/2026/10/mflow-version-3-16/

28. Fine-Tuning NVIDIA Nemotron to Know Local Rules, Not Just Local Languages

As AI adoption expands globally, people across languages and cultures are using it in more aspects of their daily lives.

Evidence: General-purpose language models serve as a powerful foundation, but they sometimes lack the local knowledge, cultural nuance, and precision a particular application needs.

Impact: Teams can evaluate the documented Nvidia workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: developer.nvidia.com · https://developer.nvidia.com/blog/fine-tuning-nvidia-nemotron-to-know-local-rules-not-just-local-languages

29. Cloudflare's own MCP servers

Cloudflare's own MCP servers.

Evidence: The Cloudflare Docs page documents the Cloudflare update.

Impact: Teams can evaluate the documented Cloudflare workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: developers.cloudflare.com · https://developers.cloudflare.com/agents/model-context-protocol/cloudflare/servers-for-cloudflare

30. Exowatt beats frontier models on accuracy with SageMaker serverless model customization

Exowatt beats frontier models on accuracy with SageMaker serverless model customization.

Evidence: The AWS Machine Learning Blog page documents the Amazon update.

Impact: Teams can evaluate the documented Amazon workflow before wider rollout.

Watch next: watch license, setup path, and benchmark.

Source: aws.amazon.com · https://aws.amazon.com/blogs/machine-learning/exowatt-beats-frontier-models-on-accuracy-with-sagemaker-serverless-model-customization

31. GEPA-Curator: Curated Learning for Self Improving Agents

Large Language Models (LLMs) are increasingly used in multi turn agentic workloads that require sustained interaction with complex environments.

Evidence: The arXiv Language-Model Research page documents the arXiv update.

Impact: Teams can evaluate the documented arXiv workflow before wider rollout.

Watch next: Watch the arXiv release for patch notes and rollout issues.

Source: arxiv.org · https://arxiv.org/abs/2610.37651v1

32. Genie 4: A new generation of world simulation

Genie 4: A new generation of world simulation.

Evidence: The Reference discovery page documents the Google update.

Impact: Teams can evaluate the documented Google workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: deepmind.google · https://deepmind.google/blog/genie-4-a-new-generation-of-world-simulation

Source: blog.google · https://blog.google/innovation-and-ai/models-and-research/google-deepmind/introducing-genie-4

33. Hunger beats tenure: Advice for junior engineers

There's no more point in giving you a little React challenge than there is in quizzing people about assembly; Claude is now better at both.

Evidence: Plenty of ink has been spilled to describe how AI is a challenge for junior developers, especially those trying to land a job.

Impact: Teams can evaluate the documented Felixrieseberg workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: felixrieseberg.com · https://felixrieseberg.com/hunger-beats-tenure-advice-for-junior-engineers

34. Impression-R1 for radiological impression generation with multicenter benchmarking and prospective evaluations

Authors develop a domain-specific large reasoning model for radiological impression generation called Impression-R1.

Evidence: It generates diagnostic summaries from radiology findings and performed favorably across institutions and public datasets.

Impact: Teams can evaluate the documented Nature workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: www.nature.com · https://www.nature.com/articles/s41467-026-79102-7

35. The Trinity:Capability-Driven Recursive Self-Improvement for Multimodal Agents

Multimodal intelligence can be viewed as human-guided co-evolution between worlds and agents through interaction.

Evidence: However, scaling this process faces two bottlenecks: constructing heterogeneous, verifiable envir.

Impact: Teams can evaluate the documented arXiv workflow before wider rollout.

Watch next: Watch the arXiv release for patch notes and rollout issues.

Source: arxiv.org · https://arxiv.org/abs/2610.36077

36. Huggingface: Spark-X2.6

Advancing the Frontier of On-Device Agents for Real-World Tasks.

Evidence: The Reference discovery page documents the Huggingface update.

Impact: Teams can evaluate the documented Huggingface workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: huggingface.co · https://huggingface.co/collections/XHToken/spark-x26

37. Skill Dreaming: Evolving Repository Skills through Pull Request Streams

However, repository-specific skills must evolve with changing code structures, implementation conventions, and maintenance requirements.

Evidence: The arXiv coding-agent research feed release records 2610.36188v1 for arXiv.

Impact: Teams can evaluate the documented arXiv workflow before wider rollout.

Watch next: watch availability, limits, and real usage for arXiv.

Source: arxiv.org · https://arxiv.org/abs/2610.36188

38. MAI-Cyber-1.1-Flash

MAI-Cyber-1.1-Flash A new cybersecurity model built into MDASH, its multi-agent vulnerability identification and remediation harness.

Evidence: The source records 1.1 / Cyber-1.

Impact: Teams can evaluate the documented Microsoft workflow before wider rollout.

Watch next: watch availability, limits, and real usage.

Source: microsoft.ai · https://microsoft.ai/models/mai-cyber-1-1-flash

39. How American Universities Are Using AI Models From Meta to Accelerate Medical and Scientific Breakthroughs

Some projects also use wearable technologies from Meta.

Evidence: The Meta AI Research Blog page documents the Meta update.

Impact: Teams can evaluate the documented Meta workflow before wider rollout.

Watch next: watch license, setup path, and benchmark.

Source: research.meta.ai · https://research.meta.ai/blog/supporting-medical-scientific-research

UNVERIFIED supplemental report

Non-official information. Unverified report. The Information reports: Why Cheaper Models Won't Win OpenAI the Enterprise. The claim from The Information is not independently confirmed. Confirmation to watch for the The Information report: A first-party page from OpenAI that states it.

Attributed source: https://www.theinformation.com/newsletters/applied-ai/cheaper-models-win-openai-enterprise

UNVERIFIED supplemental report

Non-official information. Unverified report. The Verge reports: Researchers are getting tired of OpenAI gatekeeping its breakthroughs. The claim from The Verge is not independently confirmed. Confirmation to watch for the The Verge report: A first-party page from OpenAI that states it.

Attributed source: https://www.theverge.com/ai-artificial-intelligence/1019921/openai-breakthrough-math-computer-science-physics

Finance Daily

2 source-linked editorial stories.

1. Bitcoin up 1.6% at $80,798 on CoinGecko data

Bitcoin up 1.6% at $80,798 on CoinGecko data.

Evidence: CoinGecko Bitcoin price data reports: $80,798 (up 1.6%).

Impact: CoinGecko price data show Bitcoin at $80,798, up 1.6% from $79,548 over the past 24 hours.

Limits: Bitcoin up 1.6% at $80,798: firmer rates or sticky inflation could reverse the liquidity and equity-risk read

Watch next: Watch whether Bitcoin holds the move as Wall Street opens, and how the dollar and Treasury yields respond.

Source: api.coingecko.com · https://api.coingecko.com/api/v3/coins/bitcoin/market_chart/range?vs_currency=usd&from=1793084400&to=1793170799

2. Ether up 2.1% at $2,349 on CoinGecko data

Ether up 2.1% at $2,349 on CoinGecko data.

Evidence: CoinGecko Ether price data reports: $2,349 (up 2.1%).

Impact: CoinGecko price data show Ether at $2,349, up 2.1% from $2,302 over the past 24 hours.

Limits: Ether up 2.1% at $2,349: firmer rates or sticky inflation could reverse the liquidity and equity-risk read

Watch next: Watch whether Ether holds the move as Wall Street opens, and how the dollar and Treasury yields respond.

Source: api.coingecko.com · https://api.coingecko.com/api/v3/coins/ethereum/market_chart/range?vs_currency=usd&from=1793084400&to=1793170799

GitHub Daily

1 source-linked editorial story.

1. PowerToys

Microsoft PowerToys is a collection of utilities that supercharge productivity and customization on Windows.

Evidence: GitHub Repo — https: //github.com/microsoft/PowerToys.

Impact: Shown as a quick item; the source link carries the full repository details.

Watch next: Check the README, releases, and recent activity for microsoft/PowerToys.

Source: github.com · https://github.com/microsoft/PowerToys

Source: github.com · https://github.com/trending?since=daily&spoken_language_code=en

Full source ledger: https://insight-daily-review.pages.dev/en/issues/2026-10-28/#sources

---

Spoken scripts

Text matches the passed narration and caption parity checks for this release.

AI Daily

This is AI Daily. Leading today: GitHub Copilot adds server-to-server authentication preview.

GitHub Copilot now supports server-to-server authentication for GitHub Apps installed on organizations, available in public preview. Services and agents built with the GitHub Copilot SDK can now make requests on behalf of an organization without relying on a signed-in user’s credentials. This workload-owned identity helps agents and automated services operate reliably at larger scale without tying their authentication or usage to individual users.

OpenAI: Texting your dot.

Every new plan now includes a cloud agent out of the box, and usage has daily and weekly limits instead of one monthly block of credits.

3 Ways to Boost MoE Inference. Generating tokens faster starts with identifying what limits inference performance. In the batch-size-1 evaluation of Ling 3 point 0-flash from Inclusion AI on NVIDIA DGX Spark, profiling showed that decoding was memory-bound: reading model weights drove execution time.

MAI Image 2.7 debuts above GPT Image 2.5 Flare on Artificial Analysis and #3 on Arena.

AMD and Meta Advance Strategic Partnership to Power Meta's Superintelligence Infrastructure.

6.1 Sol Degraded Model Performance. Status: Resolved All affected services are operating normally again.

CohortFlow: A Steerable Clinical Cohort Curation Agent. Language model agents for clinical cohort curation should accommodate evolving eligibility criteria without repeatedly processing unchanged requirements. arXiv introduce CohortFlow, a steerable agent that represent.

Since we launched Claude for Legal in May, the most common requests from legal teams have been for simpler setup and easier ways to work with colleagues. Artifacts are interactive tools, such as trackers and dashboards, that teams can use and share in Claude.

Building an agent that works on a demo is a different problem from building an agent that stays working across model upgrades, tool changes, and new conversation shapes. The first takes a weekend; the second takes a loop.

OpenAI expands partnership with Oracle to bring its apps. OpenAI and Oracle are expanding their partnership to bring business applications and AI agents into ChatGPT and power new Oracle offerings.

Retell uses OpenAI models. Its platform now serves more than 4,000 businesses, with monthly call volume growing from 30 million in early 2026 to more than 100 million by September. As Retell serves more enterprises with complex workflows, it uses OpenAI models to keep call quality and reliability high.

Last year at Ignite, we introduced the concept of Frontier Transformation and how it is different from any technology-led pursuit, because it requires us to focus on business outcomes first. It starts with a simple idea: technology was invented to serve humanity, not the other way around.

Anyone can bring their ideas into the company. Anyone can now bring their ideas into the company.

Much requested, and now yours to connect: iCloud Mail and Calendar plugins are available in ChatGPT. Ask ChatGPT to find that email, catch up on your inbox, check what's next, or update a calendar event. Available on all ChatGPT plans on web, desktop, and iOS.

CrowdStrike: When AI Guardrails Expose Vulnerabilities. Safety and capability are usually framed as a trade-off. This research suggests the opposite: Every improvement labs make to their guardrails also sharpens a methodology that cybersecurity practitioners or adversaries can use.

Connector features will roll out progressively, with some experiences becoming available after launch. You can ask Muse to research products and compare prices before buying on Shopee or Lazada, or plan grocery shopping and dining through GrabMart and Grab Dine Out.

One of the biggest shifts its leadership teams are seeing right now is how AI is changing the way software development engineers (SDEs) do their work, while balancing that the important elements of being an SDE are the same. Engineering leadership, including Distinguished and Senior Principal Engineers, from across Amazon recently came together to reaffirm what it means to be an SDE at Amazon.

Anysphere: New in Cursor design mode: Annotate. Give agents visual prompts: edit elements, draw on the page, and review your designs at different screen sizes.

Memory-agent benchmarks often collapse retrieval and answer generation into one score, making it hard to tell whether a failure comes from missing an update or from not using an update that was already retrieved.

Coding agent setup & developer resources | Gemini API.

FrontierCode publishes 2.0 official release notes. Most existing coding benchmarks fail to capture these gaps and tend to be saturated quickly. Cognition is excited to introduce FrontierCode 2.0 , a new generation of its coding benchmark series and its best attempt yet to evaluate user experience.

MAI Cyber 1.1 Flash: Sharper findings. MAI Cyber 1.1 Flash builds on that foundation. It's a more focused model that takes on more of the work itself, so the full MDASH system can surface useful findings at lower cost.

NVIDIA and Oracle today announced plans to expand Oracle Cloud Infrastructure (OCI) Superclusters with NVIDIA Vera Rubin NVL72 systems to multi-gigawatt scale, supporting growing customer demand for AI training, inference and agentic applications.

Announcing SOCI-indexed images for faster SageMaker AI jobs. Every minute your deep learning container image spends downloading is a minute you pay for and a minute your team waits. Container image download is pure overhead: billable job time that produces no model progress.

Multi-instance GPU (MIG) splits a GPU into hardware-isolated instances that are transparent to the application, time-slicing lets several processes take turns without memory enforcement, and Multi-Process Service (MPS) lets them run concurrently in separate address spaces (sm-sharing). CUDA 13.4 now enforces a memory boundary that is transparent to the application with a new time-sharing mode, alongside the compute sharing mode (sm-sharing) that

already existed in MPS.

Amazon SageMaker AI MLflow Apps now support MLflow 3.16, bringing a redesigned tracing experience and faster experiment tracking to your generative AI and machine learning workflows. MLflow 3.16 improves debugging and analysis of generative AI applications and agents. The redesigned trace explorer is now the default, with adjustable row height, reorderable columns, custom columns from any trace tag or metadata field.

As AI adoption expands globally, people across languages and cultures are using it in more aspects of their daily lives. General-purpose language models serve as a powerful foundation, but they sometimes lack the local knowledge, cultural nuance, and precision a particular application needs.

Cloudflare's own MCP servers.

Exowatt beats frontier models on accuracy with SageMaker serverless model customization.

GEPA-Curator: Curated Learning for Self Improving Agents. Large Language Models (LLMs) are increasingly used in multi turn agentic workloads that require sustained interaction with complex environments.

Genie 4: A new generation of world simulation.

This month, I laid the traditional coding interview to rest. As a hiring manager at Anthropic, I stopped asking candidates to do them. It had a good run: Being able to quickly define problems and find elegant solutions was a weak but useful predictor for how successful you'd be on my teams. Small coding challenges in real code did tell us something meaningful about candidates. There's no more point in giving you a little React challenge than there is in quizzing people about assembly; Claude is now better at both. its most impactful builders are successful regardless of how good they are at these tests. The coding interview was, despite its ubiquity, never popular.

Impression-R1 for radiological impression generation with multicenter benchmarking. Authors develop a domain-specific large reasoning model for radiological impression generation called Impression-R1. It generates diagnostic summaries from radiology findings and performed favorably across institutions and public datasets.

The Trinity:Capability-Driven Recursive Self-Improvement for Multimodal Agents. Multimodal intelligence can be viewed as human-guided co-evolution between worlds and agents through interaction. However, scaling this process faces two bottlenecks: constructing heterogeneous, verifiable envir.

Huggingface: Spark-X2.6. Advancing the Frontier of On-Device Agents for Real-World Tasks.

Skill Dreaming: Evolving Repository Skills through Pull Request Streams. However, repository-specific skills must evolve with changing code structures, implementation conventions, and maintenance requirements.

MAI-Cyber-1.1-Flash A new cybersecurity model built into MDASH, its multi-agent vulnerability identification and remediation harness.

How American Universities Are Using AI Models From Meta. Some projects also use wearable technologies from Meta.

Non-official information. Unverified report. The Information reports: Why Cheaper Models Won't Win OpenAI the Enterprise. The claim from The Information is not independently confirmed. Confirmation to watch for the Information report: An official OpenAI announcement that states it.

Non-official information. Unverified report. The Verge reports: Researchers are getting tired of OpenAI gatekeeping its breakthroughs. The claim from The Verge is not independently confirmed. Confirmation to watch for the Verge report: An official OpenAI announcement that states it.

Review official release notes for rollout, pricing, limits, and compatibility.

Informational and educational purposes only. This is not investment advice, a product endorsement, or professional guidance.

Finance Daily

Finance Daily for October 28, 2026. Today's stories include Bitcoin and Ether, with primary sources linked below.

Bitcoin up 1.6% at $80,798 on CoinGecko data. CoinGecko Bitcoin price data reports: $80,798 (up 1.6%). CoinGecko price data show Bitcoin at $80,798, up 1.6% from $79,548 over the past 24 hours.

Ether up 2.1% at $2,349 on CoinGecko data. CoinGecko Ether price data reports: $2,349 (up 2.1%). CoinGecko price data show Ether at $2,349, up 2.1% from $2,302 over the past 24 hours. Watch whether Ether holds the move as Wall Street opens, and how the dollar and Treasury yields respond.

Watch whether Bitcoin holds the move as Wall Street opens, and how the dollar and Treasury yields respond. Bitcoin up 1.6% at $80,798: firmer rates or sticky inflation could reverse the liquidity and equity-risk read.

Disclaimer: Informational and educational purposes only. This is not investment advice, a product endorsement, or professional guidance. Verify cited primary sources independently.

GitHub Daily

GitHub Curated Picks for October 28, 2026. Lead pick: PowerToys.

PowerToys, from microsoft. Microsoft PowerToys is a collection of utilities that supercharge productivity and customization on Windows.

Informational and educational purposes only. This is not investment advice, a product endorsement, or professional guidance.