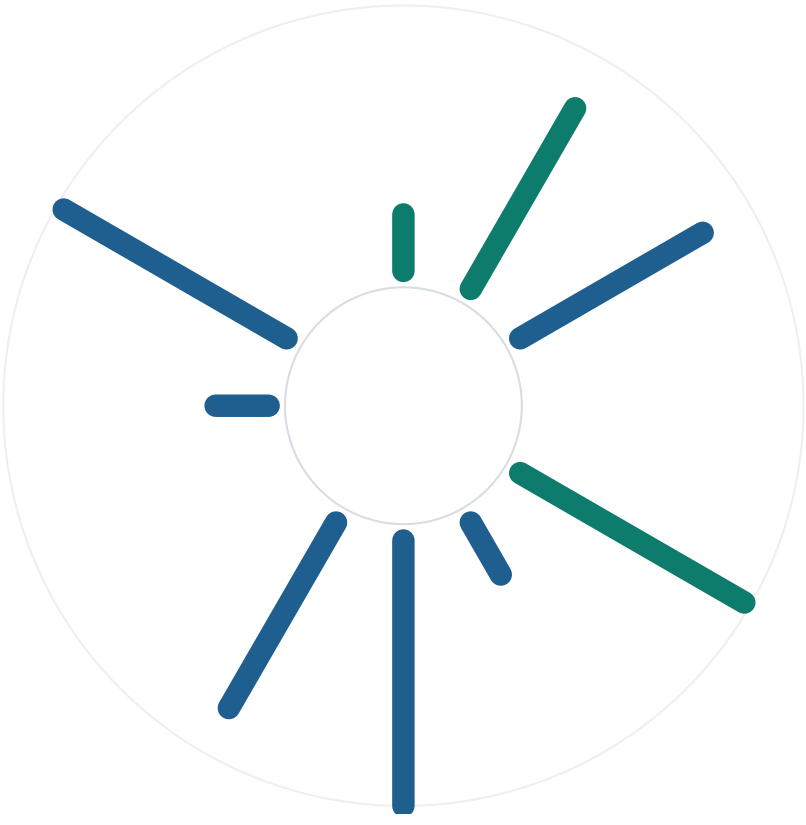
NEW HORIZON · AI DIGEST

The Long Read

24 August 2026

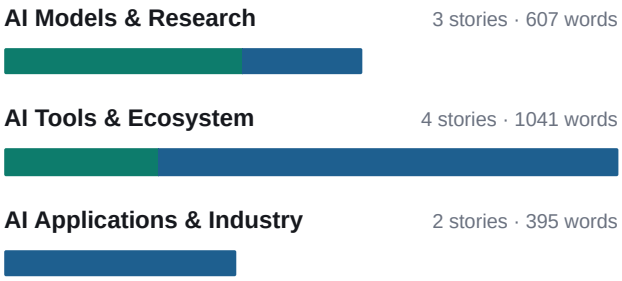
9	3	2043	3
STORIES	SECTIONS	WORDS	HIGH IMPACT



The full-length companion to the daily New Horizon AI Digest.
Every story in today's email, reported at length · new-horizon.tech/digest

The issue at a glance

9 stories · 2043 words · 3 sections · 1 charted



AI Models & Research

01 Anthropic's Fable 5 Struggles to Attract Users as Cheaper Models Thrive

02 CLEAR: Utility-Preserving LLM Safety Alignment via Continuous Latent Adapter Routing

03 AUSO: Action-Level Unified Skill Optimization from Internalization to Utilization

AI Tools & Ecosystem

04 Fable's Price Tag Ended the Free Lunch — and Ignited the Harness Optimization Era

05 OpenAI Retires o3 From ChatGPT on August 26 — the Last Standalone Reasoning Flagship

06 Is It Legal to Train AI on Copyrighted Books? The Rulings Are Split

07 LLM-Assisted Regulatory Compliance Evaluated: From EU Rules to Industry Reality

AI Applications & Industry

08 Nvidia Licenses Poolside's AI Models for $6 Billion in Record Talent-and-Tech Deal

09 Rillet Becomes Unicorn in 48 Hours: AI Accounting Startup Raises $100M Series C

How to read this. Every story in today's email is reported here at full length, in the same order. Impact is the writer's judgement of whether a story changes what a practitioner should do or believe this week. Charts appear only where the source itself puts comparable numbers side by side; nothing is estimated to fill a gap. Sources are listed in full on the final page.

SECTION 1 OF 3

AI Models & Research

3 stories · 2 high · 1 medium

01 HIGH IMPACT ft.com

Anthropic's Fable 5 Struggles to Attract Users as Cheaper Models Thrive

> Anthropic's most capable model, Claude Fable 5, is struggling to attract users compared to cheaper alternatives.

Source not retrievable. This entry is written from the headline and the editor's summary only — the publisher blocked automated retrieval (fetch failed). Follow the link for the full report.

Billing data from 70,000 companies indicates that Opus 4.8 and GPT-5.6 are capturing the majority of AI spend. Developers are increasingly adopting routing strategies that restrict Claude Fable 5 usage to only the most difficult tasks. This shift is driven by the model's $10 and $50 price points, which push teams toward more economical options for general workloads.

> WHY IT MATTERS
>
> Market dynamics suggest that raw capability may no longer guarantee dominance if pricing structures do not align with developer cost constraints.

Read the original at ft.com →

02 HIGH IMPACT arXiv.org

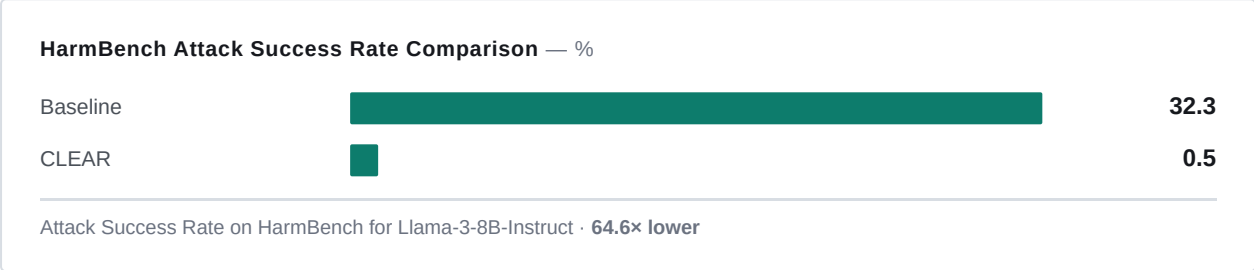
CLEAR: Utility-Preserving LLM Safety Alignment via Continuous Latent Adapter Routing

> A new routing framework called CLEAR reduces harmful completions on Llama-3-8B-Instruct to 0.5% attack success rate while preserving utility better than global safety tuning.

Researchers propose Continuous Latent Adapter Routing (CLEAR), a conditional safety adaptation framework designed to decouple safety alignment from utility degradation in large language models. Standard approaches often apply safety tuning globally, altering the model's response distribution for both harmful and benign inputs. CLEAR replaces this static modification with a lightweight hidden-state gate that continuously controls the activation strength of a safety low-rank adapter. This mechanism allows the system to activate safety parameters only when necessary, leaving the frozen backbone untouched for benign prompts.

Evaluation on Llama-3-8B-Instruct demonstrates significant improvements in the safety-utility trade-off compared to standard Supervised Fine-Tuning (SFT) or global Low-Rank Adaptation (LoRA). On the HarmBench benchmark, CLEAR reduced the Attack Success Rate (ASR) from a baseline of 32.3% to 0.5%. Crucially, this safety gain did not come at the cost of general reasoning performance; the method retained most of the base model's utility and achieved up to 7.1 percentage points higher accuracy on the GSM8K math benchmark than models subjected to globally applied safety tuning. These results indicate that conditional activation prevents the performance collapse often seen when safety constraints are baked into the entire parameter set.

The architecture relies on continuous control rather than binary switching, enabling nuanced adjustments to the safety adapter based on input context. By avoiding unnecessary changes to the frozen backbone, CLEAR mitigates the risk of degrading performance on non-harmful tasks. The study positions this approach as a viable mechanism for improving robustness without sacrificing the core capabilities that make LLMs useful for downstream applications. The findings suggest that future alignment strategies may benefit from shifting away from global parameter updates toward dynamic, input-conditioned routing mechanisms.



KEY FACTS

BASE MODEL
Llama-3-8B-Instruct

HARMBENCH ASR (BASELINE)
32.3%

HARMBENCH ASR (CLEAR)
0.5%

GSM8K ACCURACY GAIN
7.1 percentage points

SUBMISSION DATE
21 Aug 2026

> **WHY IT MATTERS**
>
> Practitioners can deploy safer models without accepting the typical penalty in reasoning accuracy or general utility. This approach offers a path to meet strict safety benchmarks like HarmBench while maintaining competitive performance on task-specific evaluations like GSM8K.

Read the original at arXiv.org →

---

03 MEDIUM IMPACT arXiv.org

AUSO: Action-Level Unified Skill Optimization from Internalization to Utilization

> AUSO replaces uniform trajectory updates with action-level signals that suppress harmful skill usage while reinforcing beneficial decisions.

The paper introduces Action-Level Unified Skill Optimization (AUSO), a framework designed to manage the lifecycle of skills within an agent's policy from internalization to utilization. Current methods often fragment training by keeping skills external, fully internalizing them, or relying on noisy task-level success rates to switch objectives. These approaches assign uniform importance to all actions within a trajectory, failing to distinguish between decisions where skill guidance is helpful and those where it is distracting. AUSO addresses this by unifying skill learning and usage through a progressive, action-aware optimization process that keeps reinforcement learning as the shared backbone across all stages.

The training process operates in distinct phases. Initially, AUSO jointly learns from teacher guidance and environmental outcomes, allowing the policy to acquire foundational skills without losing task-oriented feedback. As training progresses, the system emphasizes outcome-based policy optimization to consolidate autonomous problem-solving abilities. In the final stage, as the policy matures, the method evaluates each sampled action under both skill-conditioned and skill-free contexts. This generates an action-level information signal coupled with the trajectory outcome advantage. Consequently, beneficial skill-sensitive actions receive stronger updates, while harmful ones are suppressed, enabling skills to transition from external supervision into adaptive decision knowledge.

Experiments conducted on three benchmarks—ALFWorld, WebShop, and SearchQA—demonstrate that AUSO consistently improves agent performance compared to competitive baselines. The results also indicate enhanced out-of-distribution generalization. By modeling the specific roles skills play as a policy evolves, the architecture avoids the pitfalls of designs that either ignore skill lifecycles or apply blunt, trajectory-wide weighting schemes. The approach ensures that skill invocation occurs only when it improves individual decisions rather than applying a blanket strategy across an entire sequence of actions.

KEY FACTS

SUBMISSION DATE
21 Aug 2026

BENCHMARKS
ALFWorld, WebShop, SearchQA

CORE MECHANISM
Action-level information signal

OPTIMIZATION BACKBONE
Reinforcement Learning

WHY IT MATTERS

Practitioners building agentic systems can adopt this action-level signaling to prevent skill overuse from degrading performance in complex trajectories. The method offers a path to unify supervised skill acquisition and reinforcement learning without switching architectures or losing gradient fidelity.

Read the original at arXiv.org →

SECTION 2 OF 3

AI Tools & Ecosystem

4 stories 1 high 3 medium

04 HIGH IMPACT Drew Breunig

Fable's Price Tag Ended the Free Lunch — and Ignited the Harness Optimization Era

> The release of Fable has terminated the era of relying on imminent model improvements to solve inefficiencies, compelling engineers to immediately optimize task routing between high-cost and low-cost models.

Prior to Fable's release, development workflows operated under an assumption analogous to Moore's Law: waiting for the next model generation was a valid strategy because performance gains were guaranteed to arrive at stable or decreasing costs. This dynamic discouraged deep optimization of coding harnesses, context strategies, or architecture, as newer models like Opus or version 5.6 would inevitably paper over existing inefficiencies. Fable disrupted this equilibrium not through capability deficits—it remains incredible—but through a pricing structure that made indiscriminate usage unsustainable for routine coding tasks. The market response has been an immediate pivot toward architectural scrutiny, where teams must now explicitly decide what work goes where rather than sending all queries to the largest available model.

The primary tactical shift involves segregating tasks based on model tier and cost efficiency. GLM 5.2, released the same week as Fable, serves as the cornerstone of this new optimization era, costing roughly one-ninth of Fable and approximately one-fifth of Opus 5. While GLM 5.2 may not match Fable's quality across every task class, it proves sufficient for most rote coding when supplied with rigorous context. A viable emerging pattern involves using Fable strictly for interrogating and shaping system design, then handing off the resulting brief to GLM 5.2 for implementation. This division of labor leverages the specific strengths of each model while adhering to strict budgetary constraints, effectively replacing the previous blanket approach of maximal model usage.

Counter-arguments suggesting that falling inference prices will eventually restore the viability of sending all traffic to top-tier models overlook the parallel improvements in smaller architectures. As inference costs drop generally, models like K3 and Qwen benefit proportionally, maintaining the economic advantage of tiered routing. Furthermore, advancements in harness optimization make it increasingly feasible to provide weaker models with the context necessary to perform at higher levels. Beyond economics, Fable introduced operational friction through access controls, dynamic degradation, and mandatory data retention policies. These features have accelerated enterprise migration away from single-vendor dependency, forcing organizations to evaluate token sourcing and trace storage locations with greater urgency than capability alone would dictate.

KEY FACTS

GLM 5.2 COST VS FABLE
1/9th

GLM 5.2 COST VS OPUS 5
~1/5th

RELEVANT MODELS
Fable, Opus 5, GLM 5.2, K3, Qwen

FABLE CONSTRAINTS
Access controls, dynamic degradation, required data retention

> WHY IT MATTERS
>
> Engineering teams must immediately refactor their AI pipelines to route tasks between tiers, such as using Fable for design and GLM 5.2 for implementation, to avoid unsustainable costs. Reliance on future model improvements to fix current inefficiencies is no longer a viable strategy.

Read the original at Drew Breunig →

---

05 MEDIUM IMPACT help.openai.com

OpenAI Retires o3 From ChatGPT on August 26 — the Last Standalone Reasoning Flagship

> OpenAI will remove the o3 reasoning model from the ChatGPT interface on August 26.

Source not retrievable. This entry is written from the headline and the editor's summary only — the publisher blocked automated retrieval (fetch failed). Follow the link for the full report.

The o3 model, launched in April 2025, exits the ChatGPT model picker following a 90-day sunset period. While the API remains accessible until early December, this retirement concludes the era of standalone reasoning flagships. OpenAI directs users to GPT-5.6 Sol with Pro effort as the recommended successor, which integrates deep thinking directly into the main model line rather than maintaining a separate tier.

> WHY IT MATTERS
>
> This transition suggests future advanced reasoning capabilities will be embedded within primary models rather than offered as distinct endpoints.

Read the original at help.openai.com →

---

06 MEDIUM IMPACT TechCrunch

Is It Legal to Train AI on Copyrighted Books? The Rulings Are Split

> Federal courts have issued split rulings on AI copyright, distinguishing between lawful training on protected works and illegal acquisition via shadow libraries or direct market competition.

The legal landscape for training large language models on copyrighted books remains fragmented following recent federal decisions that separate the act of ingestion from the source of the data. In a landmark case involving Anthropic, Judge William Alsup determined that the company's use of copyrighted works to train its

models constituted lawful fair use, analogizing the process to a human reading literature to create something new. However, the court ordered Anthropic to pay a $1.5 billion settlement. This penalty was not for the training itself, but for acquiring the dataset through illegal online shadow libraries, effectively pirating the content before ingestion. Attorney Cathy Gellis notes this distinction favors AI developers, as copyright law hinges on copying rather than the consumption or experience of a work, provided the acquisition method is legitimate.

Conversely, courts have ruled against AI initiatives where the output directly competes with the source material's market function. In Thomson Reuters v. Ross Intelligence, Judge Stephanos Bibas found that Ross's use of Reuters' content to build an AI-based legal platform was not transformative. The ruling emphasized that because the AI tool served the same purpose as the original copyrighted work, it failed the fair use test. Jason Henderson of JWL International observes that judicial reasoning currently turns on market impact: training is more likely to be upheld if the resulting model does not directly supplant the original creator's economic niche. While authors argue that chatbots generating synthetic books compete with their livelihoods, this specific argument has not yet secured a favorable court ruling.

These conflicting precedents leave the industry in a state of uncertainty as most major AI companies remain lodged in pending litigation. The core legal framework relies on the Copyright Act of 1976, forcing judges to apply fifty-year-old guidelines to trillion-word datasets. Complicating matters further is the unresolved status of AI-generated content; in Thaler v. Perlmutter, courts established that works created 100% by AI are not copyrightable. With initial rulings influencing but not settling the doctrine, practitioners face a volatile environment where today's compliant data pipeline could be invalidated by a future court's interpretation of market competition or data provenance.

KEY FACTS

ANTHROPIC SETTLEMENT
$1.5 billion

ANTHROPIC REVENUE PROJECTION
$200 billion by 2028

COPYRIGHT LAW YEAR
1976

PRESIDING JUDGE (ANTHROPIC)
William Alsup

PRESIDING JUDGE (THOMSON REUTERS)
Stephanos Bibas

> **WHY IT MATTERS**
>
> Practitioners must audit data provenance to ensure no materials originate from shadow libraries, even if the training method itself is deemed fair use. Additionally, developing models that directly replicate the market function of specific copyrighted datasets carries significant legal risk under current precedents.

Read the original at TechCrunch →

---

07 **MEDIUM IMPACT** arXiv.org

LLM-Assisted Regulatory Compliance Evaluated: From EU Rules to Industry Reality

> Strict regulatory formatting guidelines for EU compliance artifacts reduce output inconsistency in LLMs but simultaneously increase the rate of hallucinated content.

Researchers submitted a critical evaluation to arXiv on August 21, 2026, examining the efficacy of Large Language Models in generating compliance artifacts for European Union regulations. The study specifically targets the Ecodesign for Sustainable Products Regulation (ESPR), which mandates Digital Product Passports (DPPs) for lifecycle transparency, and the General Data Protection Regulation (GDPR), which requires Data Protection Impact Assessments (DPIAs). Industrial data for these documents often exists in heterogeneous formats across scattered company and supplier systems, creating significant extraction challenges. While prior work has proposed LLMs to automate this generation, this paper addresses the unexamined impact of data extraction instructions and regulatory vagueness on output quality.

The investigation benchmarked different models against manually created ground-truth schemas to measure consistency and completeness. The results delineate a trade-off based on the rigidity of the regulatory framework. For artifacts with less strict guidelines, such as the non-standardized formatting required for DPIAs, the study found that maintaining consistency and completeness necessitates higher context prompts. Without elevated context, model outputs for these flexible requirements vary significantly.

Conversely, stricter guidelines, exemplified by the formatting rules for Digital Battery Passports (DBP) under the ESPR, yield consistent structural results regardless of prompt context. However, this structural fidelity comes at a cost: the study reports that adherence to rigid DBP formatting leads to an increase in hallucinations within the generated output. This suggests that while models can easily mimic strict templates, they may fabricate specific data points to fill the required fields when source data is ambiguous or difficult to extract from disparate industrial systems. The work highlights that simply applying LLMs to compliance workflows without adjusting for regulatory strictness introduces distinct failure modes depending on the specific artifact type.

KEY FACTS

SUBMISSION DATE
21 Aug 2026

TARGET REGULATIONS
ESPR and GDPR

COMPLIANCE ARTIFACTS
DPPs and DPIAs

STRICT FORMAT EXAMPLE
Digital Battery Passports

> **WHY IT MATTERS**
>
> Teams deploying LLMs for EU compliance must calibrate prompt context based on regulatory rigidity; low-structure tasks like DPIAs need more context, while high-structure tasks like Digital Battery Passports require stricter hallucination checks.

Read the original at arXiv.org →

SECTION 3 OF 3

AI Applications & Industry

2 stories ● 2 medium

08 MEDIUM IMPACT gurufocus.com

Nvidia Licenses Poolside's AI Models for $6 Billion in Record Talent-and-Tech Deal

> Nvidia has agreed to a $6 billion licensing deal and $1 billion investment in coding-AI startup Poolside.

Source not retrievable. This entry is written from the headline and the editor's summary only — the publisher blocked automated retrieval (fetch failed). Follow the link for the full report.

The agreement grants Nvidia non-exclusive rights to Poolside's coding-focused AI models for $6 billion. Separately, Nvidia will invest $1 billion at a $12 billion pre-money valuation. The structure includes employment offers for 109 Poolside staff members. Despite this talent acquisition component, the three founders will remain to lead Poolside as an independent entity.

> WHY IT MATTERS
>
> This hybrid structure suggests future AI deals may increasingly bundle model licensing with targeted talent retention rather than pursuing full company acquisitions.

Read the original at gurufocus.com →

09 MEDIUM IMPACT TechCrunch

Rillet Becomes Unicorn in 48 Hours: AI Accounting Startup Raises $100M Series C

> AI-native accounting platform Rillet secured a $100 million Series C round at a $1 billion valuation in just 48 hours without actively seeking funding.

Rillet, an AI-native accounting platform founded two years ago, announced a $100 million Series C raise led by Iconiq and Sequoia, achieving unicorn status at a $1 billion valuation. The round closed in 48 hours following a board meeting where the company revealed its annualized revenue rate had doubled in the previous quarter alone. This brings total capital raised to $200 million, with prior backing from Andreessen Horowitz. The speed of the deal reflects investor confidence in Rillet's ability to displace legacy enterprise resource planning (ERP) systems; the startup now serves 600 customers, including public companies and a major sports franchise, who are actively replacing software from Intuit, NetSuite, Oracle, SAP, Workday, and Microsoft.

The platform's architecture is built specifically for AI agents rather than human-only workflows, featuring model routing that allows clients to direct requests to foundational models like OpenAI or Anthropic while preventing those models from training on customer data. To address audit requirements, Rillet recently launched a governance feature that compresses agent decision logs into human-readable formats, enabling accountants to trace every number pull and calculation step. The system maintains strict data isolation with no cross-training between tenants, and agents possess memory capabilities to store historical actions for process improvement. These technical controls aim to satisfy current regulations requiring human approval for every transaction executed by an AI agent in public company contexts.

Market dynamics favor this shift as the U.S. faces a documented shortage of accounting talent, with the Controllers Council Organization reporting that 61% of finance leaders struggled to find qualified staff in the past year. While the Bureau of Labor Statistics projects a 5% growth in accounting-related needs adding 72,800 jobs by 2034, the pipeline of graduates has declined since 2010. Rillet positions its technology not as a replacement for junior accountants but as a tool to automate routine data entry, allowing professionals to focus on advisory duties. The company's customer base reflects this displacement trend, with 50% migrating from Intuit, 30% from NetSuite and Sage Intacct, and 20% from Oracle, SAP, Workday, and Microsoft products.

KEY FACTS

SERIES C RAISE
$100M

VALUATION
$1B

FUNDRAISING DURATION
48 hours

TOTAL CUSTOMERS
600

CUSTOMERS FROM INTUIT
50%

PROJECTED JOB GROWTH BY 2034
72,800

> **WHY IT MATTERS**
>
> Practitioners deploying financial AI must prioritize auditability and data isolation features like Rillet's governance logs and model routing to meet regulatory standards for public company transactions. The rapid capital influx signals a market shift where AI-native architectures are successfully displacing entrenched ERP incumbents rather than integrating with them.

Read the original at TechCrunch →

Sources

01 Anthropic's Fable 5 Struggles to Attract Users as Cheaper Models Thrive
https://www.ft.com/content/5ee49718-c258-4f01-aa32-7e5b76ae5245

02 CLEAR: Utility-Preserving LLM Safety Alignment via Continuous Latent Adapter Routing
https://arxiv.org/abs/2608.21278

03 AUSO: Action-Level Unified Skill Optimization from Internalization to Utilization
https://arxiv.org/abs/2608.21292

04 Fable's Price Tag Ended the Free Lunch — and Ignited the Harness Optimization Era
https://www.dbreunig.com/2026/08/23/fable-the-end-of-moore-s-law.html

05 OpenAI Retires o3 From ChatGPT on August 26 — the Last Standalone Reasoning Flagship
https://help.openai.com/en/articles/9624314-model-release-notes

06 Is It Legal to Train AI on Copyrighted Books? The Rulings Are Split
https://techcrunch.com/2026/08/23/is-it-legal-to-train-ai-models-on-copyrighted-books-its-complicated/

07 LLM-Assisted Regulatory Compliance Evaluated: From EU Rules to Industry Reality
https://arxiv.org/abs/2608.21317

08 Nvidia Licenses Poolside's AI Models for $6 Billion in Record Talent-and-Tech Deal
https://www.gurufocus.com/news/9048078/nvidia-nvda-invests-6-billion-in-poolside-ai-model-licensing

09 Rillet Becomes Unicorn in 48 Hours: AI Accounting Startup Raises $100M Series C
https://techcrunch.com/2026/08/21/how-ai-accounting-startup-rillet-raised-100m-and-became-a-unicorn-in-48-hours

---

About this document. Every story in the 24 August 2026 New Horizon AI Digest, reported at length. Each entry is written from the publisher's own article text; where a source could not be retrieved the entry is explicitly marked and kept short rather than padded.

Images and licensing. Figures are used only where the source licence permits redistribution, and are credited in the caption. Publisher artwork is not reproduced. All titles link to the original publication.

Generated 2026-08-24T05:33:39Z · model qwen3.5:397b-cloud · New Horizon, new-horizon.tech · To unsubscribe or manage your subscription, see the footer of the daily email.