

DAILY BRIEF · WHAT AI ENGINEERS ARE BUILDING

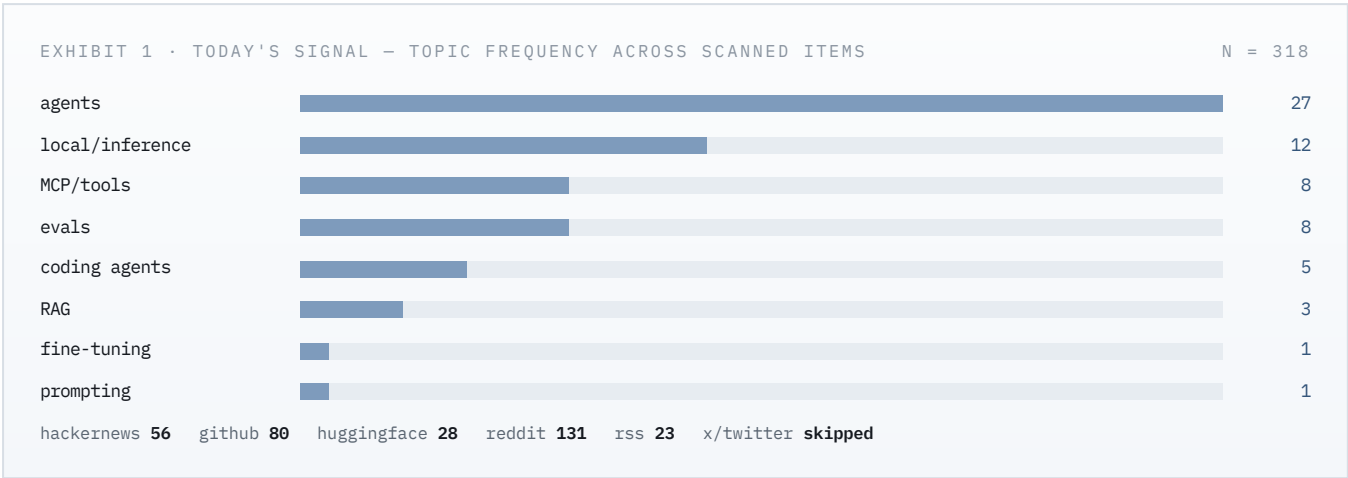
Today in AI engineering

2026-09-19

Daily AI-engineering signal (mechanical fallback — LLM error: 404 Client Error: Not Found for url: <https://api.groq.com/openai/v1/chat/complet>)

5 deep dives 3 builder stories 15 quick hits 318 items scanned mode: **mechanical fallback**

- agents
- local/inference
- mcp/tools
- evals
- coding agents
- rag
- fine-tuning
- prompting



DEEP DIVES

Show HN: An OSS Python dependency scanner for exploited, unmaintained packages

Show HN

I built an open source python dependency scanner that will scan and flag packages with known exploit CVEs(CISA's Known Exploited list and FIRST EPSS) and unmaintained packages that have not had a release or commit in 2 years. Theres also claude hook that will make your AI agent not install these typ

Why it matters. Flagged by the ranking heuristics as a real, artifact-backed build.

TRY THIS

Open the link and skim the repo / paper.

<https://github.com/binuka200/package-doctor>

incoai/splash

[GitHub](#)

A local inference engine for Apple silicon, built around the model.

Why it matters. Flagged by the ranking heuristics as a real, artifact-backed build.

TRY THIS

Open the link and skim the repo / paper.

<https://github.com/incoai/splash>

AetherLabsAI/RSIAgent

[GitHub](#)

A training-free multi-agent framework for recursive self-improvement in new environments through broad-then-deep autonomous exploration and reusable memory.

Why it matters. Flagged by the ranking heuristics as a real, artifact-backed build.

TRY THIS

Open the link and skim the repo / paper.

<https://github.com/AetherLabsAI/RSIAgent>

Show HN: I built a PDF accessibility checker that is cross platform

[Show HN](#)

(no summary)

Why it matters. Flagged by the ranking heuristics as a real, artifact-backed build.

TRY THIS

Open the link and skim the repo / paper.

<https://github.com/visionably/outloud>

deepseek-ai/DeepSeek-V4.1-Flash

[HF model](#)

Trending model on Hugging Face — image-text-to-text, transformers; 3239 likes, 482,270 downloads.

Why it matters. Flagged by the ranking heuristics as a real, artifact-backed build.

TRY THIS

Open the link and skim the repo / paper.

<https://huggingface.co/deepseek-ai/DeepSeek-V4.1-Flash>

WHAT PEOPLE ACTUALLY BUILT

Show HN: Seal – Letters and passwords that open for your family after you die

[Show HN](#)

I'm a CS undergrad who has developed an interest in hardware keys ever since I used them for a school project a few years ago. I made a little chat app that only allows users to chat if they registered a key, nothing else. The backstory of how I came to work on Seal (this project

STEAL THIS

Flagged by keyword heuristics as a first-person build write-up — read the post for the details.

<https://github.com/jasonepage/Seal>

Show HN: Timeline, a team project planner with a fast canvas Gantt view

[Show HN](#)

I manage a few projects at work. I really am unsatisfied with modern tools available to me for that purpose. I had always envisioned a tool that I could interact with visually for this purpose. I started building one many years ago but the graphics programming was beyond my capab

STEAL THIS

Flagged by keyword heuristics as a first-person build write-up — read the post for the details.

<https://github.com/vakahnke/Timeline>

I tried Jev-style decisions with local Qwen: same accuracy, 239 ms vs 368 ms

[r/LLMDevs](#)

I liked the idea behind Jev, but I don't want to send all my data to someone else's server. So I made choosekit, a small TypeScript package for getting choices and probabilities from a model running in llama.cpp. I tested it with Qwen3.8 27B Q4 XL against Jev, using SemIf's 144-t

STEAL THIS

Flagged by keyword heuristics as a first-person build write-up — read the post for the details.

https://www.reddit.com/r/LLMDevs/comments/1wkc9hp/i_tried_jevstyle_decisions_with_local_qwen_same/

QUICK HITS

Claude Code now reads AGENTS.md if there is no Claude.md [HN](#)

<https://code.claude.com/docs/en/changelog>

prism-ml/Ternary-Bonsai-2-27B-gguf [HF](#) [MODEL](#)

Trending model on Hugging Face — text-generation, llama.cpp; 1031 likes, 1,516,960 downloads.

<https://huggingface.co/prism-ml/Ternary-Bonsai-2-27B-gguf>

ruc-datalab/EvoOntology [GITHUB](#)

EvoOntology: A Self-Evolving Ontology Layer for Data Agents  EvoOntology插件为Claude Code/Codex 建立&进化本体层

<https://github.com/ruc-datalab/EvoOntology>

Show HN: I told my open-source AI agent it was a prisoner – it tried to escape [SHOW](#) [HN](#)

https://github.com/nordevelopment/RoninAgent/blob/main/EXPERIMENT_REPORT_SUBJECT0.md

An empirical study of harness design for coding agents [HN](#)

<https://arxiv.org/abs/2609.20804>

Show HN: Jeff – A read-only CLI for semantic code review using Jev [SHOW](#) [HN](#)

<https://github.com/Alurith/jeff>

elsechord/CyberGuard [GITHUB](#)

Evidence-driven autonomous SOC team on AgentTeams: hash-bound approvals, HMAC audit chain, rollback and reproducible attack scenarios — audi

<https://github.com/elsechord/CyberGuard>

Qwen/Qwen3.8-27B [HF](#) [MODEL](#)

Trending model on Hugging Face — image-text-to-text, transformers; 15690 likes, 7,365,368 downloads.

<https://huggingface.co/Qwen/Qwen3.8-27B>

Tokha233/deepseek-v4.1-flash-a100-turbo [GITHUB](#)

DeepSeek-V4.1-Flash inference on NVIDIA Ampere with vLLM, SM80 patches, reproducible throughput and coding-agent benchmarks. Tested on 8x A8

<https://github.com/Tokha233/deepseek-v4.1-flash-a100-turbo>

m-a-p/YuE2-3B HF MODEL

Trending model on Hugging Face — text-to-audio; 836 likes, 15,446 downloads.

<https://huggingface.co/m-a-p/YuE2-3B>

dashscope-sdk/agentscope-java GITHUB

Build distributed, production-grade, long-running agents.

<https://github.com/dashscope-sdk/agentscope-java>

ISTA-DASLab/Qwen3.8-27B-GSQ-RCO-GGUF HF MODEL

Trending model on Hugging Face — image-text-to-text, gguf; 1374 likes, 1,154,265 downloads.

<https://huggingface.co/ISTA-DASLab/Qwen3.8-27B-GSQ-RCO-GGUF>

XingChen-AGI/Xing4.0-29B-A4B HF MODEL

Trending model on Hugging Face — text-generation, transformers; 545 likes, 7,278 downloads.

<https://huggingface.co/XingChen-AGI/Xing4.0-29B-A4B>

ukisai/Swift-Qwen3.8-27b HF MODEL

Trending model on Hugging Face — image-text-to-text, transformers; 460 likes, 8,761 downloads.

<https://huggingface.co/ukisai/Swift-Qwen3.8-27b>

Lightricks/LTX-2.5 HF MODEL

Trending model on Hugging Face — image-to-video, diffusion-single-file; 4366 likes, 1,607,815 downloads.

<https://huggingface.co/Lightricks/LTX-2.5>
