

DAILY BRIEF · WHAT AI ENGINEERS ARE BUILDING

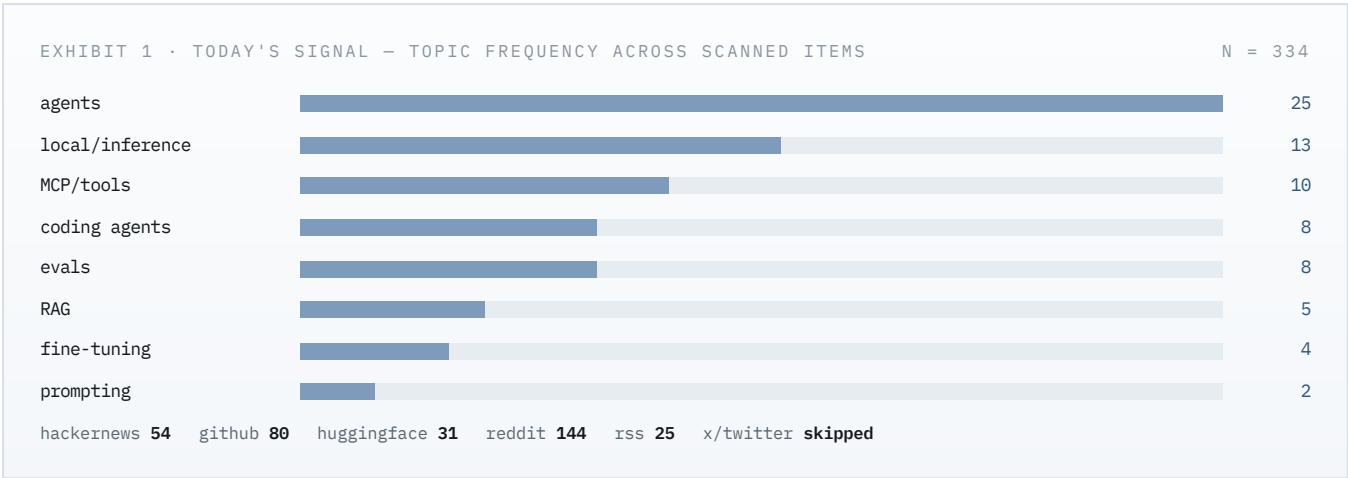
Today in AI engineering

2026-08-26

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 334 items scanned mode: **mechanical fallback**

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



DEEP DIVES

Show HN: I made a Raspberry with Qwen my local car AI

HN

Found that you can actually run a 35B Qwen model on a Pi with very impressive intelligence and stability. Built connectors for car ODB to read all about car internals, and manufacturer's cloud service for stuff like changing AC or opening/ locking doors. Gave it info such as the full car manual. And

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/ThinkOffApp/CarWatch>

ApodexAI/FrontierAgent

[GitHub](#)

🌱 FrontierAgent, our agent framework, open-sourced alongside it — native command-line TUI, ReAct and Agent Team modes, one command on macOS and Linux, no preinstall, no hard Docker dependency.

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/ApodexAI/FrontierAgent>

duty1g/x64dbg-mcp-server

[GitHub](#)

x64dbg-MCP Server is a native MCP (Model Context Protocol) plugin for x64dbg that exposes the debugger's full functionality over HTTP. Connect any MCP-compatible AI assistant and control x64dbg programmatically: set breakpoints, step through code, read memory, dump registers, and more. Built with Z

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/duty1g/x64dbg-mcp-server>

Show HN: TeXbrain, a LaTeX editor that runs pdfTeX in the browser via WASM

[HN](#)

I'm a master's engineering student and a big fan of LaTeX, which I used for my thesis and research articles. I have used Overleaf and that was fine until I wanted to git sync, which unfortunately sits behind a paywall. Since I didn't want to pay subscriptions for things that should simply just work,

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/swimmingbrain/texbrain>

Show HN: LatticeDB – Like SQLite but for graph databases

[Show HN](#)

We have been using graph DBs more and more at work. I found them painful to work with locally and decided to try and build something better.

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/jeffhajewski/latticedb>

WHAT PEOPLE ACTUALLY BUILT

Show HN: Slash-tokens – know LLM cost before the call leaves your machine

[Show HN](#)

I kept finding out what a call cost after I had already paid, so I built /Slash -- slash-tokens is a 4.8KB WASM that estimates tokens FAST and can block a wasted trip, or switch to cheaper same-provider model before the request leaves your machine. Try it MIT, no account needed:

STEAL THIS

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

<https://github.com/Wolfe-Jam/slash-tokens>

Show HN: I built self-hosted deployment automation tool for Windows and IIS

[Show HN](#)

Seven years in the making (I admit that I did get a kid at some point which slowed down progress and brought it to a halt quite a few times), I decided to finally release it, with a rather drastic change to the now (I believe) generous licensing model. Unlike current major players

STEAL THIS

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

<https://fdeploy.com/>

Show HN: Keenable – A different web search API for AI agents

[Show HN](#)

Hey HN! We built <https://keenable.ai>, a different web search API for AI agents. Keenable searches our own 100B+ page index. We are focused on low cost and latency (p95 <https://keenableai.github.io/needle>): a live benchmark that compares Keenable with other search APIs on fresh

STEAL THIS

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

<https://keenable.ai/>

QUICK HITS

Show HN: Screen memory without screenshots, just text to Markdown HN

It's a macOS menu bar app that reads the text of your focused window every few seconds through the Accessibility API. No screenshots, no vid
<https://github.com/dragthelake/ambient-context>

Show HN: Lightweight system monitor for Linux VPS written in Go SHOW HN

<https://github.com/leodeim/vpsmon>

Qwen/Qwen3.8-27B HF MODEL

Trending model on Hugging Face — image-text-to-text, transformers; 12753 likes, 2,945,415 downloads.
<https://huggingface.co/Qwen/Qwen3.8-27B>

unsloth/Qwen3.8-27B-GGUF HF MODEL

Trending model on Hugging Face — model; 2928 likes, 7,334,695 downloads.
<https://huggingface.co/unsloth/Qwen3.8-27B-GGUF>

ArihantDeva/heimdall GITHUB

Persistent memory for AI coding agents — one verified kb_search replaces the grep/find/ls orientation loop. Cross-repo, CPU-only, zero token
<https://github.com/ArihantDeva/heimdall>

OBLITERATUS/Qwen3.8-27B-OBLITERATED HF MODEL

Trending model on Hugging Face — text-generation, mlx; 764 likes, 389,747 downloads.
<https://huggingface.co/OBLITERATUS/Qwen3.8-27B-OBLITERATED>

orcarouter/Qwen3.8-27B-Uncensored-MLX HF MODEL

Trending model on Hugging Face — image-text-to-text, mlx; 1106 likes, 68,855 downloads.
<https://huggingface.co/orcarouter/Qwen3.8-27B-Uncensored-MLX>

d4ncboz/github-farm GITHUB

Production-grade, AI-Agent-friendly multi-platform OAuth harvesting and session management framework for AI Gateways.
<https://github.com/d4ncboz/github-farm>

orcarouter/Qwen3.8-27B-Uncensored-FP8 HF MODEL

Trending model on Hugging Face — image-text-to-text, transformers; 1154 likes, 249,744 downloads.
<https://huggingface.co/orcarouter/Qwen3.8-27B-Uncensored-FP8>

Show HN: Open-source AMDGCN kernels for optimizing LLM inference SHOW HN

<https://github.com/NetraRuntime/netra-kernel>

ornith-ai/Ornith-1.5-35B-A3B HF MODEL

Trending model on Hugging Face — text-generation, transformers; 429 likes, 70,158 downloads.
<https://huggingface.co/ornith-ai/Ornith-1.5-35B-A3B>

HauhauCS/Qwen3.8-27B-Uncensored-HauhauCS-Aggressive-MTP-GGUF HF MODEL

Trending model on Hugging Face — image-text-to-text; 628 likes, 832,185 downloads.
<https://huggingface.co/HauhauCS/Qwen3.8-27B-Uncensored-HauhauCS-Aggressive-MTP-GGUF>

Lightricks/LTX-2.5 HF MODEL

Trending model on Hugging Face — image-to-video, diffusion-single-file; 1809 likes, 833,845 downloads.
<https://huggingface.co/Lightricks/LTX-2.5>

Show HN: JeopardyGithub: GitHub Repo to Jeopardy Game SHOW HN

Introducing <http://jeopardygithub.com/> convert any github repo into a multiplayer classic jeopardy board!
<https://jeopardygithub.com>

Show HN: Orbit - One agent across many repos: real worktrees, no index SHOW HN

<https://github.com/orbcli/orbit>
