

DAILY BRIEF · WHAT AI ENGINEERS ARE BUILDING

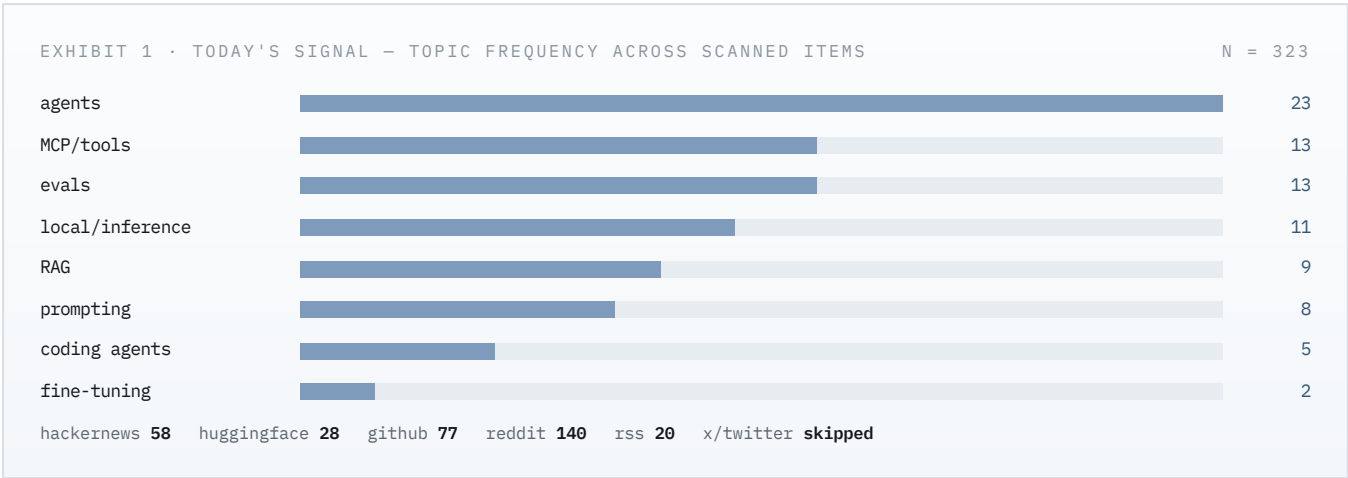
Today in AI engineering

2026-09-15

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 323 items scanned mode: **mechanical fallback**

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



DEEP DIVES

Show HN: Kinesis – Control your Mac with the Meta Neural Band

Show HN

a couple hours ago i had astra take on a 10 month old repo of mine <https://github.com/callbacked/neural-band-poc/> that involved finding a way to use the meta neural band independently from the glasses to read the sEMG data off of it. i really think it is a cool piece of technology that largely gets

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/callbacked/kinesis>

Show HN: Authorize MCP tool calls without giving agents the credentials

[Show HN](#)

Hey HN, Adam here. We built a small MCP example around a single problem: How do you let an agent use a tool that needs credentials without giving the credentials to the agent? The demo has two tools. One is open. One is protected. When the agent calls the protected tool, Keydris checks the policy fi

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/keydrisLabs/mcp-auth-keydris-template>

OpenArch – PyTorch implementations of modern LLM architectures

[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/anuj0456/OpenArch>

Show HN: Rebuno - An open-source runtime for production agents

[Show HN](#)

I have been building Rebuno, an open-source execution runtime for operating AI agents in production. It manages execution state and guardrails across agents built with different frameworks. Agents run as HTTP services. Rebuno dispatches work through signed webhooks and keeps execution state in Postg

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/rebuno/rebuno>

deepseek-ai/DeepSeek-V4.1-Flash

[HF model](#)

Trending model on Hugging Face — image-text-to-text, transformers; 2563 likes, 325,712 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: Nari Qwen3-TTS and Qwen3-ASR – High accuracy, low latency and cost

[HN](#)

Hey HN, Toby from Nari Labs here. We've been working on making OSS speech models super-fast. Last year, we built Dia, the first OSS text-to-speech model capable of doing natural dialogue. Since then, so many more great speech models have been released to the public. But the marke

STEAL THIS

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

<https://narilabs.com/blog/nari-labs-leads-coval-voice-ai-benchmarks/>

Show HN: I built a browser-native agent for automating job application forms

[Show HN](#)

(no summary)

STEAL THIS

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

<https://github.com/maazkhanxo/jobeasyapply-linkedin-auto-apply>

Show HN: I built Otis, a minimal AI agent that runs local models out of the box

[Show HN](#)

Hi everyone, Been working on Otis, an open-source ai agent that gives you one minimal experience across local and hosted open-weight models, privacy-focused by design. On setup it recommends a local model based on the hardware Otis is running on, downloads it and runs it through

STEAL THIS

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

<https://triangllabs.ai/otis>

QUICK HITS

Show HN: Macros with a Behringer FCB1010 MIDI Pedalboard in macOS [SHOW](#) [HN](#)

<https://github.com/JamesRyanATX/fcbnerd>

Edge0/Edge0-35B-A3B-preview [HF](#) [MODEL](#)

Trending model on Hugging Face — text-generation, mlx; 2317 likes, 17,853 downloads.

<https://huggingface.co/Edge0/Edge0-35B-A3B-preview>

youngyangyang04/llm-master [GITHUB](#)

大模型（LLM）全栈学习路线与中文教程🔥：覆盖 Prompt Engineering、RAG、AI Agent、MCP、微调、模型部署、Transformer、AI 编程与大厂面试，从入门到生产实践。

<https://github.com/youngyangyang04/llm-master>

Show HN: A skill that makes people think you're using Claude Code [SHOW](#) [HN](#)

Built a skill that makes people think you're using Claude Code. Sometimes when browsing repositories on GitHub, you might notice "Claude" li

<https://github.com/Juwan-Hwang/claude-mimic>

panda-sandeep/dearagent [GITHUB](#)

Email inbox API for AI agents, self-hosted on Cloudflare Workers. Open source, with an MCP server. AgentMail alternative.

<https://github.com/panda-sandeep/dearagent>

Show HN: ChaosTree – A zero-dependency Java tree library (AVL,RBT,B-Tree,B+ Tree) [SHOW](#) [HN](#)

What is ChaosTree? ChaosTree is a zero-dependency Java Sorted Set/Map library built around multiple tree implementations. It provides implem

<https://github.com/Chaos-vy/ChaosTree>

openbmb/MiniCPM5-2B HF MODEL

Trending model on Hugging Face — text-generation, transformers; 1419 likes, 271,754 downloads.
<https://huggingface.co/openbmb/MiniCPM5-2B>

Show HN: Fly.exe – An EON systems like virtual fruit fly uploaded to computer SHOW HN

<https://github.com/Ibtisam-Mohammad/Fly.exe>

Show HN: Pelican-bicycle alternatives HN

In November and December 2025, inspired by Simon Willison's pelican-riding-a-bicycle benchmark, I had some then-current LLMs create SVGs from
<https://gally.net/temp/20260914pelican-alternatives/index.html>

NativePHP/superstack GITHUB

Open-source Laravel starter with Filament, NativePHP SuperNative, and MCP
<https://github.com/NativePHP/superstack>

nex-agi/Nex-N2.5-mini HF MODEL

Trending model on Hugging Face — text-generation, transformers; 793 likes, 5,202 downloads.
<https://huggingface.co/nex-agi/Nex-N2.5-mini>

Qwen/Qwen3.8-27B HF MODEL

Trending model on Hugging Face — image-text-to-text, transformers; 15192 likes, 7,702,543 downloads.
<https://huggingface.co/Qwen/Qwen3.8-27B>

nex-agi/Nex-N2.5-Pro HF MODEL

Trending model on Hugging Face — text-generation, transformers; 640 likes, 30,881 downloads.
<https://huggingface.co/nex-agi/Nex-N2.5-Pro>

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

Trending model on Hugging Face — text-to-audio; 512 likes, 6,716 downloads.
<https://huggingface.co/m-a-p/YuE2-3B>

AetherLabsAI/RSIAgent GITHUB

A training-free multi-agent framework for recursive self-improvement in new environments through broad-then-deep autonomous exploration and
<https://github.com/AetherLabsAI/RSIAgent>
