DAILY BRIEF · WHAT AI ENGINEERS ARE BUILDING

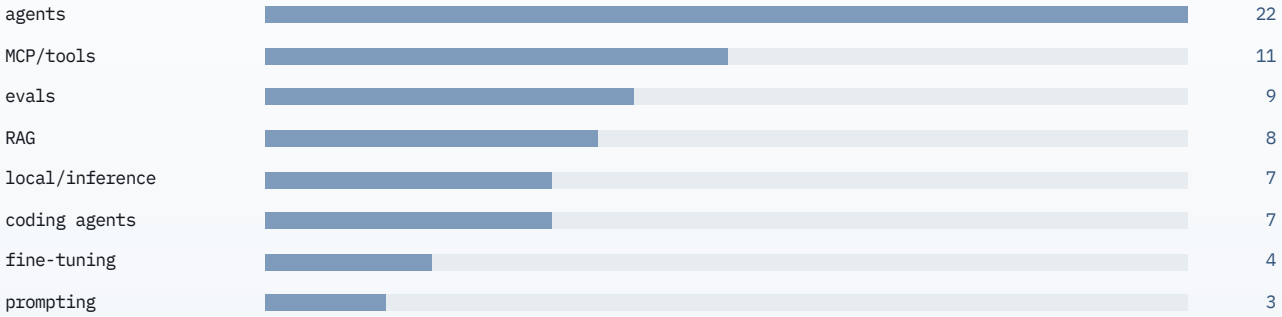
Today in AI engineering
2026-11-08

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 **319** items scanned mode: **mechanical fallback**

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

EXHIBIT 1 · TODAY'S SIGNAL — TOPIC FREQUENCY ACROSS SCANNED ITEMS N = 319



hackernews **55** github **80** huggingface **29** reddit **133** rss **22** x/twitter **skipped**

DEEP DIVES

Show HN: Coldframe, a honeypot that poses as a supermarket fridge controller

Show HN

I built Coldframe, a honeypot that pretends to be a supermarket refrigeration controller. Inspired by the events on industrial fridge installations in the past weeks, I wanted to see what people would try against a convincing fake. To learn more about this I built a honeypot that resembled perfectly

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/hedgerow-dev/coldframe

Show HN: We compressed a 31B video+audio diffusion model to 14B, 6 steps, INT8

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/Video-Rebirth/hyperflow-h3-flash

ruvnet/ruvllm

GitHub

Governed neural weight editing and coding model adaptation. Native Rust patches, exact rollback, LoRA, Claude Code routing, and reproducible benchmarks.

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/ruvnet/ruvllm

Show HN: Ryph – Self-Hosted, Open-Source Coding Agents for Incident Response

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/ryph-hq/ryph

beipo-ai/NorthCUA

GitHub

Turn intent into work. Open-source Desktop/Web agent harness with long-horizon memory, live evaluation and recoverable batches.

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/beipo-ai/NorthCUA

WHAT PEOPLE ACTUALLY BUILT

Show HN: An online shop where your friends can hurry the courier

Show HN

I built an emulator of a flash-sale fashion outlet with no registration and fake checkout that never takes money. You can browse items that are on sale, watch in real time as they go out of stock one by one, and get notified when they are back on sale (possibly with different pri

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

https://www.dopaminechic.com/

Show HN: Where Were We

Show HN

A little web game I made as an anniversary present for my wife. Upload vacation photos and make them into a game. WWW uses EXIF to determine where and when the photo was taken. This is then turned into a game you can send to anyone.

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

https://wherewerewe.app/

Show HN: We made a new theme to increase engagement on our website

Show HN

Backstory: We run an old-internet-style website called ZenMemes. Recently, end-user testing has shown that many users find this site disorienting, because there is no cookie banner, and at no point are you told the experience is better in the app. To help them adapt, we built a n

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

https://zenmemes.com/threads/thread-20261107173646-27caa60f/

QUICK HITS

shanselman/sandbox-mcp GITHUB

Fresh, throwaway Windows machines for testing your apps: an MCP server and CLI that automate Windows Sandbox

https://github.com/shanselman/sandbox-mcp

Glen-Web-App/best-agent-memory-mcp GITHUB

Best Agent Memory MCP servers, ranked and compared: long-term memory for Claude Code, Codex, Cursor and any MCP client. Glen, Mem0, Supermem

https://github.com/Glen-Web-App/best-agent-memory-mcp

gogoymh/auto-ontology-mcp GITHUB

NVIDIA Auto Ontology's semantic layer and a personal knowledge base as an MCP server for Claude, ChatGPT and other MCP clients. No LLM API k

https://github.com/gogoymh/auto-ontology-mcp

speridlabs/idol HF MODEL

Trending model on Hugging Face — image-to-3d, idol; 1263 likes, 22,854 downloads.

https://huggingface.co/speridlabs/idol

phonny40rowle/rag-search GITHUB

A self-hosted retrieval-augmented search API: web search, content extraction, embeddings, and answer generation.

https://github.com/phonny40rowle/rag-search

didilili/mianstack GITHUB

🚀面栈 MianStack | 程序员面试题库天花板！Java、Python、数据库、算法、大模型、AI Agent、RAG 一栈备齐。真实面经关联高频考点，按公司、岗位定向备考；图文题解、精华专题、项目源码拆解，全系标配。面试官出题，面栈出招！免费开源，持续更新。

https://github.com/didilili/mianstack

liangjianshu/multimodal-image-retrieval GITHUB

多模态检索实战：意图识别、Prompt 工程、Qwen LoRA、EmbeddingGemma 2 与 InfoNCE 对比学习。含真实演示、自有数据接入、训练评测及 RAG 检索组件学习指南。

https://github.com/liangjianshu/multimodal-image-retrieval

xtools-at/DeepSeek-V4-Pro-Qwen3.5-9B-UNCENSORED-heretic HF MODEL

Trending model on Hugging Face — text-generation, transformers; 682 likes, 7,088 downloads.

https://huggingface.co/xtools-at/DeepSeek-V4-Pro-Qwen3.5-9B-UNCENSORED-heretic

moonshotai/Kimi-K3.1-Pro HF MODEL

Trending model on Hugging Face — image-text-to-text, transformers; 1684 likes, 25,569 downloads.
https://huggingface.co/moonshotai/Kimi-K3.1-Pro

YT-Proteus-bot/MineAgent GITHUB

agent定制自由化，agent流程可视化。在浏览器里用画布搭出工业级 Agent 流程，选框架自动生成架构图， 拖节点、连线、写代码、配模型，点「保存」就把整张画布编译成一个可独立运行的 Python Agent 项目， 点「运行」直接在你的机器上真实执行并流式回传每一步。欢迎
https://github.com/YT-Proteus-bot/MineAgent

sepehr071/iran-jobs-mcp GITHUB

Unofficial read-only MCP server for Iranian job sites (استخدام) IranTalent, Jobinja and Jobvision: job search, companies, reviews, interview
https://github.com/sepehr071/iran-jobs-mcp

ccore/Qwen3.5-4B-3D-Coder HF MODEL

Trending model on Hugging Face — model; 894 likes, 254,053 downloads.
https://huggingface.co/ccore/Qwen3.5-4B-3D-Coder

Lightricks/LTX-2.5 HF MODEL

Trending model on Hugging Face — image-to-video, diffusion-single-file; 11056 likes, 2,661,026 downloads.
https://huggingface.co/Lightricks/LTX-2.5

exzade/EXZA-Remote GITHUB

Secure AI-powered remote computer access for ChatGPT, Claude and compatible MCP clients.
https://github.com/exzade/EXZA-Remote

Show HN: Maple – Sub-2-BPW Moe Pruning to Run Gemma 4 26B at 45 TPS on a 6GB GPU SHOW HN

https://github.com/AmSach/maple

Generated 2026-11-08 · Sources: Hacker News, Reddit, GitHub, Hugging Face, arXiv, engineering RSS. Filtered for real, implementable work; marketing and cash-grabs removed. Built by the ai-daily-brief automation · LUCA CONARROE — AI Automation.