DAILY BRIEF · WHAT AI ENGINEERS ARE BUILDING

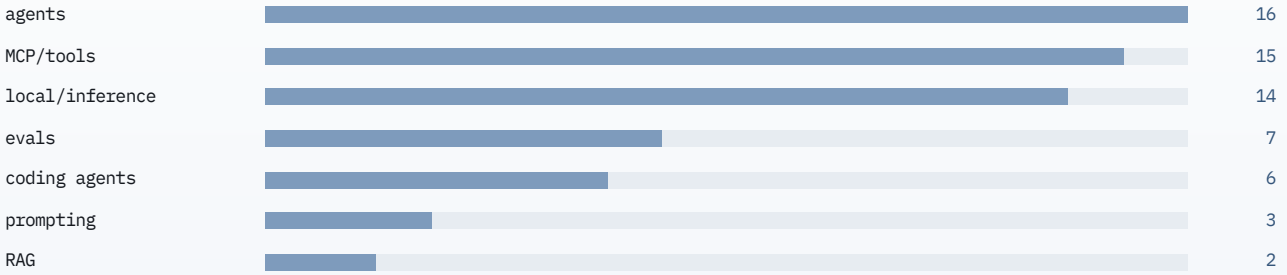
Today in AI engineering 2026-11-03

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 **330** items scanned mode: **mechanical fallback**

agents mcp/tools local/inference evals coding agents prompting rag

EXHIBIT 1 · TODAY'S SIGNAL — TOPIC FREQUENCY ACROSS SCANNED ITEMS N = 330



github **79** hackernews **61** huggingface **29** reddit **143** rss **18** x/twitter **skipped**

DEEP DIVES

glincker/awesome-selfhosted-ai

GitHub

Curated list of self-hosted open-source AI: local LLMs, chat UIs, agents, MCP servers, RAG, speech and image generation. Health-checked weekly, dead projects removed automatically.

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/glincker/awesome-selfhosted-ai

muaaz9329/cheap-ag

GitHub

50x your usage with cheap-ag. Opus/Astra plans, a dirt-cheap model writes every line, and tests decide what ships. Your $20 plan stops hitting limits. One MCP server for Claude Code & Codex.

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/muaaz9329/cheap-ag

abenzerps/Qwen-Image-2.1-Uncensored-Quantized

HF model

Trending model on Hugging Face — text-to-image, gguf; 5588 likes, 2,508,590 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/abenzerps/Qwen-Image-2.1-Uncensored-Quantized

LiquidAI/LFM3-1.7B

HF model

Trending model on Hugging Face — text-generation, transformers; 1321 likes, 47,817 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/LiquidAI/LFM3-1.7B

Show HN: PyWolf3D, a wolfenstein3D clone written from scratch in Python

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/atifcodesalot/pyWolf3D

WHAT PEOPLE ACTUALLY BUILT

Show HN: I built a little browser game to practice Vim commands

Show HN

I made hjkl to help myself remember Vim commands. It’s a little browser game you can play when you want to unwind and get a bit of Vim practice in. There are still a lot of commands I want to add, but I hope someone finds it useful in the meantime :)

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

https://hjkl.duy.run

Launch HN: Momentic (YC W24) – Agents that catch bugs by using your app

HN

Hey HN, we’re Wei-Wei and Jeff. We’re building Momentic, QA agents that autonomously plan, explore, and bug-bash controls and user flows in your web, iOS, Android, and (soon) desktop apps to catch bugs. Here’s a demo: https://youtu.be/BFsM-Se1Fh8 We’re both software engineers tha

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

https://momentic.ai/

Show HN: HN Atlas – Discover tools to better engage with Hacker News

Show HN

HackerTube and HackerBooks remind me of the awesome HN products I use daily: - hckr news: a collection of top stories. - hnrss: custom RSS feeds filtered by keywords and points. - HN Popularity: explore popular HN bloggers. Hacker News features a massive amount of high-quality co

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

https://hn-atlas.bensontech.dev/

QUICK HITS

Show HN: Subvocal – Talk to your computer without making noise SHOW HN

Hi HN, we're Aarush and Arnav from The Subvocal Company (YC F26) (https://subvocal.company/). We're building smart glasses that let you ta
https://news.ycombinator.com/item?id=50340088

Show HN: I've updated my Linux System Call website and API SHOW HN

I first released those in back in 2020 and now I finally added more architectures, Kernel selector, syscall descriptions and a few other nic
https://syscall.sh/

Net worth for people aged 30–33 in the U.S. HN

https://www.keepingup.com/benchmarks/net-worth/age-30-33

exzade/EXZA-Remote GITHUB

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

deepseek-ai/DeepSeek-V4.1-Flash-1017 HF MODEL

Trending model on Hugging Face — image-text-to-text, transformers; 4793 likes, 187,260 downloads.
https://huggingface.co/deepseek-ai/DeepSeek-V4.1-Flash-1017

nvidia/NVIDIA-Nemotron-3.5-Transcribe-30B-A4B HF MODEL

Trending model on Hugging Face — automatic-speech-recognition, nemo; 706 likes, 38,942 downloads.
https://huggingface.co/nvidia/NVIDIA-Nemotron-3.5-Transcribe-30B-A4B

Show HN: Hacker Times, a New York Times-inspired version of Hacker News SHOW HN

I thought it would be fun to browse HN as if it were a newspaper. The site updates once an hour. Enjoy!
https://hackertimes.ca/

peterpoal134/memory-brain GITHUB

给 AstrBot 的可落地 AI 记忆与生活层：五类记忆、混合检索、遗忘治理、情绪边界、自主生活循环。不覆盖人格。
https://github.com/peterpoal134/memory-brain

Show HN: Warsaw in voxels – An interactive 3D map of the city centre SHOW HN

Explore central Warsaw in voxels - slightly inspired by Minecraft, and MagicaVoxel. You can change the date, time of day and weather, explor
https://p.migdal.pl/warsaw-vibe-voxel/

wendersoon/radar-cnpj-mcp GITHUB

Servidor MCP e CLI para localizar empresas brasileiras por CNAE e região, utilizando dados abertos de CNPJ da Receita Federal.
https://github.com/wendersoon/radar-cnpj-mcp

thefahien/elpis-on-7900xtx GITHUB

Qwen3.8-27B EXL3 + DFlash2 speculative decoding on an AMD RX 7900 XTX (exllamav3-rocm): recipe, patches, benchmarks
https://github.com/thefahien/elpis-on-7900xtx

mistralai/Mistral-Large-4-1T-A52B HF MODEL

Trending model on Hugging Face — image-text-to-text, transformers; 1523 likes, 6,556 downloads.

https://huggingface.co/mistralai/Mistral-Large-4-1T-A52B

Show HN: Open a 300k-row XLSX in the browser in ~300ms, no server SHOW HN

https://github.com/p10node/office-engine

speridlabs/idol HF MODEL

Trending model on Hugging Face — image-to-3d, idol; 467 likes, 10,742 downloads.

https://huggingface.co/speridlabs/idol

beipo-ai/NorthCUA GITHUB

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

https://github.com/beipo-ai/NorthCUA

Generated 2026-11-03 · 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.