DAILY BRIEF · WHAT AI ENGINEERS ARE BUILDING

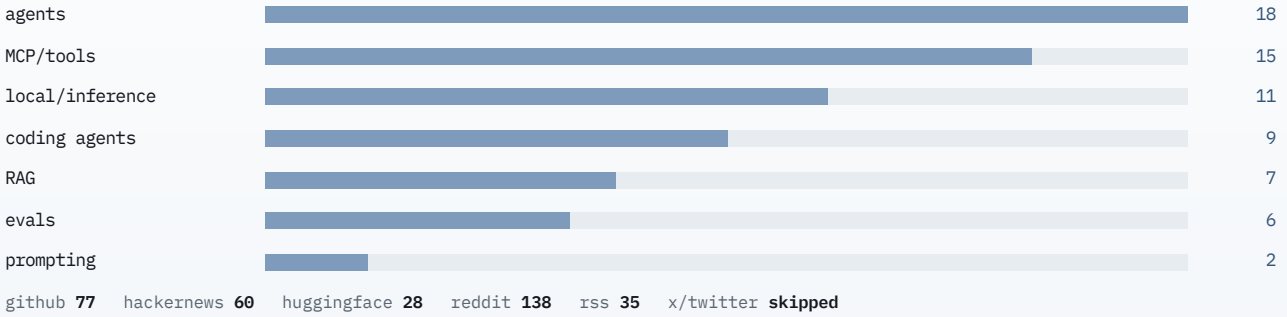
Today in AI engineering
2026-10-17

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 **338** items scanned mode: **mechanical fallback**

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

EXHIBIT 1 · TODAY'S SIGNAL — TOPIC FREQUENCY ACROSS SCANNED ITEMS N = 338



github **77** hackernews **60** huggingface **28** reddit **138** rss **35** x/twitter **skipped**

DEEP DIVES

zhuyansen/new-term-radar

GitHub

Daily radar for new AI terms: names of new models, products and repos, caught on X, Hugging Face, GitHub and sitemaps before keyword tools list them.

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/zhuyansen/new-term-radar

Llama.cpp on a jailbroken PS5: run any GGUF model natively on the console

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/cobanov/PS5LM

anujasid/mcp-for-blender-pro

GitHub

AI assistant and MCP server for Blender: visual chat, scene tools, model selection, multi-agent workflows. Works with OpenAI, Claude, Gemini, Ollama and OpenCode. Blender AI automation, bpy scripting, 3D modeling, geometry nodes and MCP integration.

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/anujasid/mcp-for-blender-pro

sepehr071/awesome-iran-mcp

GitHub

19 open-source MCP servers (308 tools) that give AI agents the Iranian web: Digikala, Snappfood, Jabama, MrBilit, Paziresh24, Taaghche and more. Prices in Toman.

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/sepehr071/awesome-iran-mcp

YuyaoGe/doc2harness

GitHub

Official Code for the paper "Doc2Harness: Benchmarking LLMs on Long-Horizon Harness Engineering of Coding Agents "

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/YuyaoGe/doc2harness

WHAT PEOPLE ACTUALLY BUILT

Show HN: Sill for Android – control your Mac's windows from an Android phone

Show HN

Sill (discussed here: https://news.ycombinator.com/item?id=50112739) streams one Mac window, or the whole Desktop, to an iPhone or iPad and lets you control it by touch. It's open source, so I made an Android client for it. It's unofficial and not affiliated with Sill's author:

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

https://github.com/rahimnathwani/sill-android

I built a tiny LLM inference engine to learn how vLLM works - looking for feedback

r/LLMDevs

I've been learning how LLM inference engines work, so I built a small one from scratch rather than just reading about the internals. GitHub: https://github.com/Aparajith24/nano-vLLM The main things I've implemented so far: Continuous batching Paged KV cache with a shared memory p

> **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/1x7io8v/i_built_a_tiny_llm_inference_engine_to_learn_how/

I built a CLI agent that triages SAST findings using an adversarial LLM pass (Open Source) r/AI_Agents

Traditional SAST tools drown you in false positives, while dumping an entire repo into an LLM is slow, expensive, and hits context limits fast. To fix this in my own workflow, I built PuedoBot a CLI security auditor that uses deterministic scanning to find suspicious code, and a

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

https://www.reddit.com/r/AI_Agents/comments/1x7my7o/i_built_a_cli_agent_that_triages_sast_findings/

QUICK HITS

90% of US working hours are spent on tasks not covered well by AI benchmarks HN
https://www.mercor.com/research/labor-index/benchmark-map/

abenzerps/Qwen-Image-2.1-Uncensored-Quantized HF MODEL
Trending model on Hugging Face — text-to-image, gguf; 4459 likes, 2,571,722 downloads.
https://huggingface.co/abenzerps/Qwen-Image-2.1-Uncensored-Quantized

Show HN: Gemini 4 Argon plays StarCraft HN
I've refreshed StarSkirmish Bench, a benchmark where LLMs play StarCraft BW by writing code, with results from the flagship models of OpenAI
https://starskirmish.com/bench/v0-2/

Jev Decision Index HF SPACE
Trending Space (static, 705 likes). Benchmarks and news on various repros of TypeSafe's Jev
https://huggingface.co/spaces/multimodalart/jev-decision-index

ziplab/ATAS GITHUB
Any Time, Any Space: A World-State-Centric Framework for Agentic Video Generation
https://github.com/ziplab/ATAS

Graphicle-AI/graphicle GITHUB
Code graph and memory for your coding agents, in one Rust program. For Claude Code, Codex, Cursor and any MCP client.
https://github.com/Graphicle-AI/graphicle

cdcailf/cjkfold GITHUB
Fold look-alike CJK glyphs (Kangxi radicals, CJK radical supplements) so text extracted from PDFs matches again.
https://github.com/cdcailf/cjkfold

google/embeddinggemma-2 HF MODEL
Trending model on Hugging Face — feature-extraction, transformers; 2539 likes, 158,490 downloads.
https://huggingface.co/google/embeddinggemma-2

jialinyyzz/humanizer HF MODEL
Trending model on Hugging Face — text-generation, gguf; 1851 likes, 95,870 downloads.
https://huggingface.co/jialinyyzz/humanizer

LSPosed/dexsql-release GITHUB
SQL for Android apk and dex files
https://github.com/LSPosed/dexsql-release

openbmb/MiniCPM-V-4.7-35B-A3B HF MODEL
Trending model on Hugging Face — image-text-to-text, transformers; 864 likes, 36,252 downloads.
https://huggingface.co/openbmb/MiniCPM-V-4.7-35B-A3B

fishaudio/drama-3 HF MODEL

Trending model on Hugging Face — text-to-speech; 574 likes, 8,193 downloads.
https://huggingface.co/fishaudio/drama-3

Claude reveals ECDSA curve origins HN

https://words.filippo.io/seeds-solved/
https://www.anthropic.com/research/ecdsa-curve-origins

Show HN: Stoke Money – Personal finance app built around a cash flow forecast SHOW HN

Hi HN, I've been through many finance tracking apps, starting with Personal Capital, then Monarch, and most recently Copilot. None of them h
https://stoke.money

Lightricks/LTX-2.5 HF MODEL

Trending model on Hugging Face — image-to-video, diffusion-single-file; 8129 likes, 1,993,003 downloads.
https://huggingface.co/Lightricks/LTX-2.5

Generated 2026-10-17 · 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.