

DAILY BRIEF · WHAT AI ENGINEERS ARE BUILDING

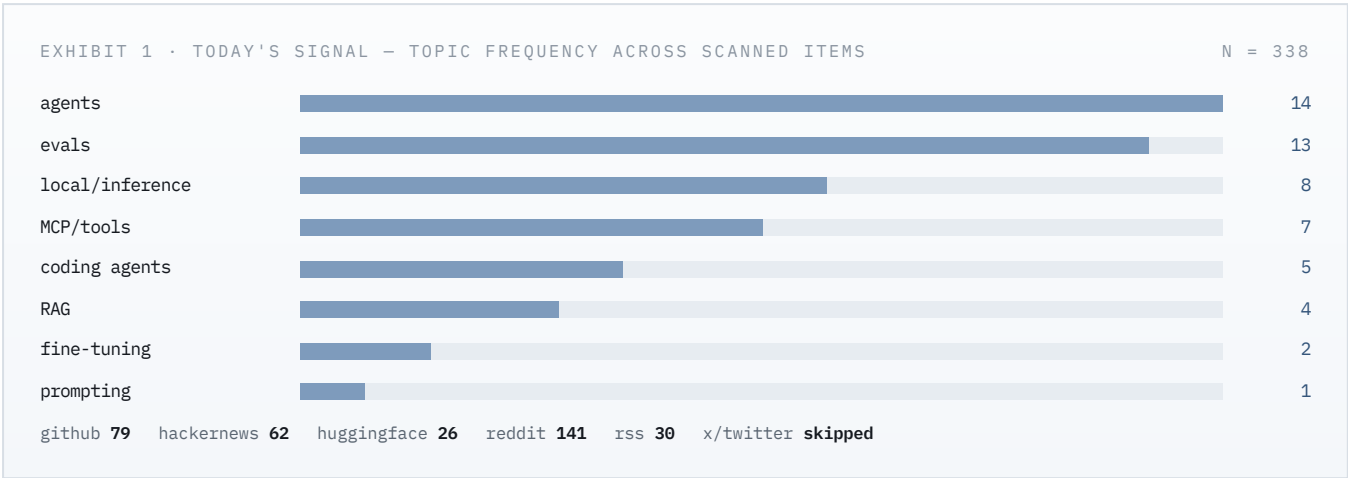
Today in AI engineering

2026-09-04

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



DEEP DIVES

Ryze-AI-Adgent/open-seo-mcp-skills

GitHub

Open-source SEO + GEO skills for Claude — keyword research, rank tracking, site audits, backlinks, competitor gaps, AI visibility. Runs on your real Search Console / GA4 / ads data via MCP, with DataForSEO built in. Free, MIT.

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/Ryze-AI-Adgent/open-seo-mcp-skills>

2akouwu/reverify

[GitHub](#)

Anti-hallucination for AI agents that read binaries. The model proposes, deterministic tools decide: every claim is VERIFIED or REFUTED against the real bytes, with evidence, and grounded facts survive context resets. MCP server + CLI.

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/2akouwu/reverify>

Codex outage

[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/openai/codex/issues/28756>

Show HN: LibPolyCall – a C runtime broker for cross-language function calls

[Show HN](#)

LibPolyCall is an open-source C runtime for connecting programs across language boundaries without requiring each language pair to implement its own integration layer. The architecture is program-first rather than binding-first, with a stable C ABI, FFI bindings, Polycallfile/Polycallrc configuratio

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/obinexus/libpolycall-v1>

Show HN: Wordmate – local speech-to-text for macOS

[Show HN](#)

Hi HN! I built Wordmate because I wanted WisprFlow-style dictation locally on my machine. It's a free macOS app. Audio and transcripts never leave your Mac. Speech recognition uses Parakeet and optional post-processing uses a tuned Qwen3 0.6B 4-bit to improve punctuation and capitalization and clean

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/vdszds/wordmate>

WHAT PEOPLE ACTUALLY BUILT

Porting my 1993 Amiga game to Godot, with an LLM reading the 68000 assembly

[HN](#)

These are my notes from porting my Amiga game, which I originally built in Baghdad in 1993 in MC68000 assembly, to Godot, using Claude Fable 5 during last July holiday. It took an evening! Getting the feel right and shipping it took a few more weekends and evenings. I spent the l

STEAL THIS

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

<https://babyloniantwins.com/blog/porting-a-1993-amiga-game-to-godot/>

Show HN: Litelink – local-first, embedded stream capture into Iceberg tables

[Show HN](#)

Hi HN! I just wanted to share litelink a local-first, embedded capture library I built in python (code is heavily AI generated but designed and reviewed by yours truly). I've been using this for point-and-shoot WebSocket capture but I imagine it could also be useful for observabi

STEAL THIS

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

<https://github.com/nhobin219/litelink>

Show HN: I made text selection on macOS programmable

[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/ganeshmshetty/openclip>

QUICK HITS

OpenAI begins rolling out GPT-6 Astra [HN](#)

<https://thenewstack.io/openai-gpt6-astra-benchmarks/> , image: <https://cdn.thenewstack.io/media/2026/09/358eb84a-screenshot...>
<https://ventur>
<https://www.cnbc.com/2026/09/03/open-ai-astra-gpt-6-cyber.html>

GPT-6 Astra [HN](#)

System Card: <https://deploymentsafety.openai.com/gpt-6-astra> Related ongoing threads: OpenAI's GPT-6 Astra on ARC-AGI-3 - <https://news.ycomb>
<https://openai.com/index/gpt-6-astra/>

nahidalam/world_model_from_scratch [GITHUB](#)

Code repository for World Model From Scratch book
https://github.com/nahidalam/world_model_from_scratch

zai-org/GLM-5.3 [HF](#) [MODEL](#)

Trending model on Hugging Face — text-generation, transformers; 1636 likes, 303,534 downloads.
<https://huggingface.co/zai-org/GLM-5.3>

Vuk97/forward-implementation-first [GITHUB](#)

Stop your coding agent from stalling real work on self-invented bookkeeping - receipts, hashes, locks, certification rituals. Ship first, th
<https://github.com/Vuk97/forward-implementation-first>

Qwen/Qwen3.8-Flash-Next HF MODEL

Trending model on Hugging Face — image-text-to-text, transformers; 4832 likes, 351,374 downloads.
<https://huggingface.co/Qwen/Qwen3.8-Flash-Next>

adtexerry-lgtm/unigit-ecosystem GITHUB

UNIGIT public brand and ecosystem hub — AI should work for everyone.
<https://github.com/adtexerry-lgtm/unigit-ecosystem>

Qwen/Qwen3.8-27B HF MODEL

Trending model on Hugging Face — image-text-to-text, transformers; 13870 likes, 5,739,341 downloads.
<https://huggingface.co/Qwen/Qwen3.8-27B>

zai-org/GLM-5.3-Flash HF MODEL

Trending model on Hugging Face — image-text-to-text, transformers; 2030 likes, 654,957 downloads.
<https://huggingface.co/zai-org/GLM-5.3-Flash>

deepseek-ai/DeepSeek-V4-Flash-Vision-Exp HF MODEL

Trending model on Hugging Face — image-text-to-text, transformers; 564 likes, 133,024 downloads.
<https://huggingface.co/deepseek-ai/DeepSeek-V4-Flash-Vision-Exp>

Show HN: ForkForensics – recovers a deleted GitHub repo’s history from forks SHOW HN

<https://github.com/siris9476/forkforensics>

vcruz305/vllm-exl3 GITHUB

Serve EXL3 (ExLlamaV3 trellis) quantized models on vLLM fork runtimes — any architecture, mixed per-layer bitrates, composable with source-f
<https://github.com/vcruz305/vllm-exl3>

ashutoshsinghpr7/wikiskill GITHUB

WikiSkill (arXiv:2608.27454) for Hermes Agent — self-evolving agent skills via a persistent knowledge wiki. Faithful Algorithm 1 implementat
<https://github.com/ashutoshsinghpr7/wikiskill>

Show HN: LibreDB Studio, a self-hosted MIT browser IDE for 10 database engines SHOW HN

<https://github.com/libredb/libredb-studio>

Show HN: We built client-side AST SEO audits for a Next.js CMS to avoid LLM lag SHOW HN

<https://github.com/nextblock-cms/nextblock>
