

DAILY BRIEF · WHAT AI ENGINEERS ARE BUILDING

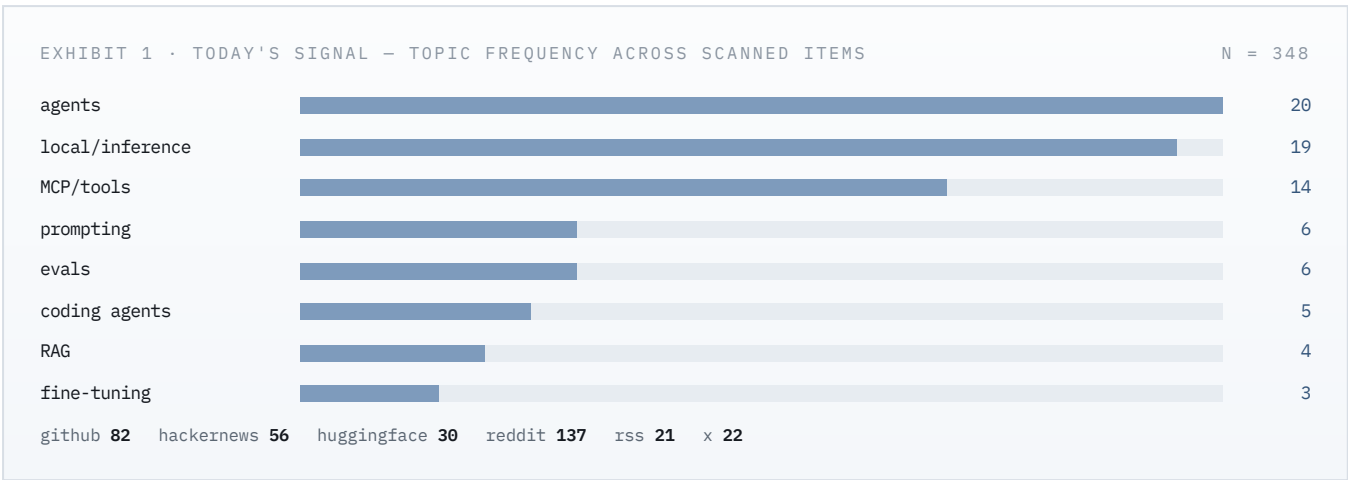
Today in AI engineering

2026-07-21

Developers are building and sharing AI-powered tools for automation, evaluation, and inference, with a focus on agents, LLMs, and RAGs.

3 deep dives 2 builder stories 10 quick hits 348 items scanned

- llm
- ai agents
- rag
- inference
- automation



DEEP DIVES

EthanXiang777/circuit-framework

github

Ethan Xiang built Circuit Framework, a multi-agent LLM trading research system. The system allows for the creation and testing of trading strategies using LLMs. The framework is designed to be flexible and adaptable to different market conditions.

Why it matters. Circuit Framework is significant because it demonstrates the potential of LLMs in trading research. The system's ability to adapt to different market conditions makes it a valuable tool for traders and researchers.

TRY THIS

Explore the Circuit Framework repository and experiment with creating your own trading strategies using LLMs.

<https://github.com/EthanXiang777/circuit-framework>

- llm
- trading
- research

KlaaTAl/klaatcode

[github](#)

The KlaatAI team built KlaatCode, an open-source AI coding agent for the terminal. KlaatCode uses smart model routing to cut costs and supports multiple AI models, including Claude, GPT, and Gemini.

Why it matters. KlaatCode is important because it provides a cost-effective and flexible solution for coding tasks. The use of smart model routing makes it an attractive option for developers and researchers.

TRY THIS

Install KlaatCode and experiment with using it for coding tasks, exploring the different AI models and routing options.

<https://github.com/KlaaTAl/klaatcode>

[ai](#)[coding](#)[terminal](#)

PromptPartner/agentsmith

[github](#)

The PromptPartner team built AgentSmith, a universal, model-agnostic operating harness for AI agents. AgentSmith provides a lean core and work-type profiles assembled by a single setup script.

Why it matters. AgentSmith is significant because it provides a flexible and adaptable solution for working with AI agents. The model-agnostic design makes it a valuable tool for researchers and developers.

TRY THIS

Explore the AgentSmith repository and experiment with using it to manage and deploy AI agents.

<https://github.com/PromptPartner/agentsmith>

[ai](#)[agents](#)[model-agnostic](#)

WHAT PEOPLE ACTUALLY BUILT

Show HN: A Pipeline for Making 10-minute AI Movies with Claude Code and Seedance

[hackernews](#)

A pipeline for making 10-minute AI movies using Claude Code and Seedance. The pipeline glues together Seedance, Nano Banana, and ElevenLabs to create a movie.

How it works. The pipeline uses Claude as the director and combines the outputs of Seedance, Nano Banana, and ElevenLabs to create a movie.

STEAL THIS

The use of a pipeline to combine multiple AI tools can lead to the creation of complex and interesting outputs, such as movies.

<https://github.com/dawndrain/movie-gen>

Show HN: The0 – self-hosted runtime for trading bots, bring your own language

[hackernews](#)

A self-hosted runtime for trading bots that allows users to bring their own programming language. The runtime provides a flexible and adaptable solution for trading bot development.

How it works. The runtime uses a modular design to allow users to plug in their own programming language and trading strategy.

STEAL THIS

The use of a self-hosted runtime can provide a high degree of flexibility and customizability for trading bot development.

<https://github.com/alexanderwanyoike/the0>

QUICK HITS

thinkingmachines/Inkling HUGGINGFACE

A trending model on Hugging Face for image-text-to-text tasks.
<https://huggingface.co/thinkingmachines/Inkling>

prism-ml/Ternary-Bonsai-27B-gguf HUGGINGFACE

A trending model on Hugging Face for text generation tasks.
<https://huggingface.co/prism-ml/Ternary-Bonsai-27B-gguf>

LoRA Speedrun – a public wall-clock leaderboard for fine-tuning techniques HACKERNEWS

A public leaderboard for comparing fine-tuning techniques for LLMs.
<https://github.com/Saivineeth147/lora-speedrun>

CyberSunil/LLMVault GITHUB

An intentionally vulnerable platform for testing AI security and prompt injection attacks.
<https://github.com/CyberSunil/LLMVault>

drpwchen/textbook-to-note GITHUB

A tool for converting PDF textbooks into AI-searchable knowledge bases.
<https://github.com/drpwchen/textbook-to-note>

EndymionLee/PilotBrowseMCP GITHUB

A browser runtime that allows AI agents to control a real Chrome browser via MCP.
<https://github.com/EndymionLee/PilotBrowseMCP>

yashneil75/gitlord GITHUB

An agent orchestration framework with Git-backed storage.
<https://github.com/yashneil75/gitlord>

aliakrami1375/Li-Translate GITHUB

An open-source AI-powered subtitle generation and translation platform.
<https://github.com/AliAkrami1375/Li-Translate>

MemTensor/memmy-agent GITHUB

A shared memory agent for AI applications.
<https://github.com/MemTensor/memmy-agent>

FareedKhan-dev/glm-5.2-in-c GITHUB

A C inference engine for the GLM-5.2 model.
<https://github.com/FareedKhan-dev/glm-5.2-in-c>
