

DAILY BRIEF · WHAT AI ENGINEERS ARE BUILDING

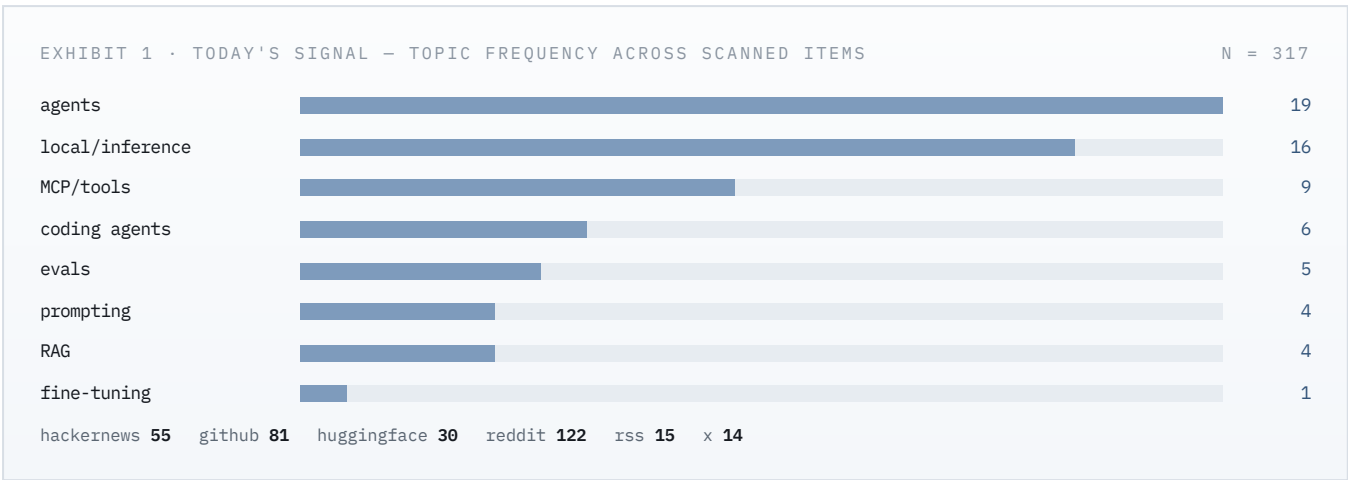
Today in AI engineering

2026-07-20

Developers are building and sharing open-source AI tools and models for various applications, including coding, translation, and data analysis.

3 deep dives 2 builder stories 5 quick hits 317 items scanned

- llm
- ai agents
- mcp
- transformers
- inference engines



DEEP DIVES

PengZhang64/circuit-framework

github

PengZhang64 has created a multi-agent LLM trading research system called Circuit Framework. This system allows for the creation and testing of various trading strategies using LLMs. The framework is open-source and available on GitHub. PengZhang64 is actively maintaining and updating the project.

Why it matters. Circuit Framework is significant because it provides a platform for researchers and developers to experiment with LLMs in trading applications. This can lead to the development of more sophisticated trading strategies and improved decision-making tools.

TRY THIS

Explore the Circuit Framework repository and try implementing a simple trading strategy using LLMs.

<https://github.com/PengZhang64/circuit-framework>

- llm
- trading
- research

vshulcz/deja-vu

github

Vshulcz has developed a tool called Deja Vu, which enables retroactive search and recall of session logs for coding agents. This tool can help improve the efficiency and accuracy of coding tasks. Vshulcz has made Deja Vu open-source and available on GitHub.

Why it matters. Deja Vu is important because it addresses a common problem in coding agent development, which is the inability to recall previous interactions. By providing a solution to this problem, Deja Vu can enhance the overall performance of coding agents.

TRY THIS

Download and install Deja Vu, then experiment with using it to improve the recall of your coding agents.

<https://github.com/vshulcz/deja-vu>

coding agents

recall

efficiency

thinkingmachines/Inkling

huggingface

Thinking Machines has released a model called Inkling, which is an image-text-to-text transformer model. Inkling has gained significant attention on Hugging Face, with over 1,000 likes and 10,000 downloads. Thinking Machines is a prominent developer of AI models and tools.

Why it matters. Inkling is notable because it demonstrates the growing interest in multimodal models that can process both images and text. This can lead to the development of more sophisticated applications, such as image captioning and visual question answering.

TRY THIS

Explore the Inkling model on Hugging Face and try using it for an image-text-to-text task, such as image captioning.

<https://huggingface.co/thinkingmachines/Inkling>

transformers

image-text-to-text

multimodal

WHAT PEOPLE ACTUALLY BUILT

Show HN: I replaced a \$120k bowling center system with \$1,600 in ESP32s

hackernews

The author built a bowling center system using ESP32 microcontrollers, replacing a \$120,000 system with a much cheaper alternative.

How it works. The system uses ESP32s to control the bowling center's machinery and provide a user interface for customers.

STEAL THIS

The author's project demonstrates the potential for using low-cost microcontrollers to replace expensive proprietary systems in various industries.

<https://news.ycombinator.com/item?id=48968606>

Show HN: A self-hosted AI that turns Hacker News into a daily briefing

hackernews

The author built a self-hosted AI system that summarizes Hacker News articles into a daily briefing.

How it works. The system uses natural language processing and machine learning to analyze Hacker News articles and generate a summary.

STEAL THIS

The author's project shows how AI can be used to automate tasks, such as summarizing news articles, and provide personalized briefings.

<https://github.com/RecNes/hn-ai-summarizer>

OpenAI reduces Codex Model Context Size from 372k to 272k HACKERNEWS

OpenAI has reduced the context size of its Codex model, which can improve performance and reduce memory usage.

<https://github.com/openai/codex/pull/33972/files>

KlaaTAl/klaaTcode GITHUB

KlaaTAl has released an open-source AI coding agent called KlaatCode, which supports multiple models and provides smart model routing.

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

PromptPartner/agentsmith GITHUB

PromptPartner has developed a universal operating harness for AI agents, called AgentSmith, which supports multiple models and provides a lean core and work-type profiles.

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

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

Prism-ml has released a trending model on Hugging Face called Ternary-Bonsai-27B-gguf, which is a text-generation model with over 800 likes and 300,000 downloads.

<https://huggingface.co/prism-ml/Ternary-Bonsai-27B-gguf>

zeraix/zeraix GITHUB

Zeraix has developed an open-source local AI workspace, which provides on-device inference and is advancing the field of edge AI.

<https://github.com/zeraix/zeraix>
