

DAILY BRIEF · WHAT AI ENGINEERS ARE BUILDING

Today in AI engineering

2026-07-18

AI agents and LLM tooling advancements drive innovation in automation and evaluation

4 deep dives 2 builder stories 16 quick hits 351 items scanned

llm

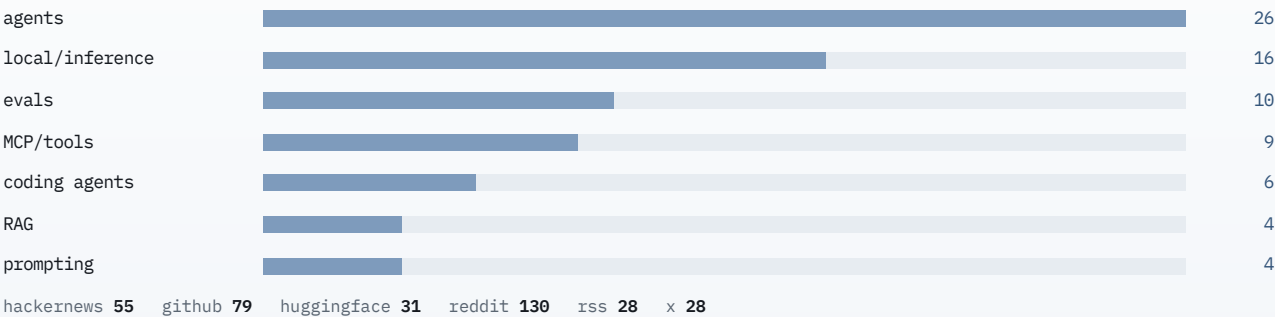
ai agents

automation

evaluation

EXHIBIT 1 · TODAY'S SIGNAL – TOPIC FREQUENCY ACROSS SCANNED ITEMS

N = 351



DEEP DIVES

mereyabdenbekuly-ctrl/clodex-ide

github

Mereyabdenbekuly-ctrl has released Clodex-ide, a local-first, zero-trust agentic IDE for verifiable autonomous software development. This project aims to provide a secure and transparent development environment. The IDE is designed to work with various LLMs and AI agents.

Why it matters. Clodex-ide is an innovative approach to software development, focusing on security, transparency, and autonomy. It has the potential to revolutionize the way developers work with AI agents and LLMs.

TRY THIS

Explore the Clodex-ide repository and experiment with its features to understand its capabilities and limitations.

<https://github.com/mereyabdenbekuly-ctrl/clodex-ide>

llm

ai agents

ide

PengZhang64/circuit-framework

github

PengZhang64 has developed the Circuit Framework, a multi-agent LLM trading research system. This framework is designed to facilitate research and development of AI-powered trading systems. It provides a comprehensive platform for building and testing trading strategies.

Why it matters. The Circuit Framework is a significant contribution to the field of AI-powered trading research. It has the potential to accelerate the development of more sophisticated trading systems.

TRY THIS

Investigate the Circuit Framework's architecture and explore its potential applications in trading research.

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

llm

ai agents

trading

thinkingmachines/Inkling

huggingface

ThinkingMachines has released Inkling, a trending model on Hugging Face. Inkling is an image-text-to-text model that has gained significant attention in the AI community. It has been downloaded over 7,870 times and has 969 likes.

Why it matters. Inkling is a notable example of the rapid progress being made in AI model development. Its popularity demonstrates the growing interest in image-text-to-text models.

TRY THIS

Experiment with Inkling and explore its capabilities in image-text-to-text tasks.

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

llm

image-text

hugging face

vshulcz/deja-vu

github

Vshulcz has developed Deja Vu, a memory layer for coding agents. Deja Vu provides features such as search, MCP recall, auto-context, secret redaction, stats, share, and sync over session logs. It is designed to work with various coding agents, including Claude Code, Codex, and Gemini CLI.

Why it matters. Deja Vu is an innovative solution for coding agents, providing a comprehensive memory layer that enhances their capabilities. Its features have the potential to improve the efficiency and effectiveness of coding agents.

TRY THIS

Investigate Deja Vu's features and explore its potential applications in coding agent development.

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

llm

coding agents

memory layer

WHAT PEOPLE ACTUALLY BUILT

I made LLMs debate each other about the question you ask until the consensus

reddit

The author built a system that enables LLMs to debate each other on a given question until they reach a consensus. This system is open-source and browser-only, allowing users to explore the debate process and gain insights into the LLMs' decision-making.

How it works. The system uses a consensus-driven approach, where the LLMs engage in a debate and iteratively refine their arguments until they reach a consensus. The author provides an insightful result of the LLMs reaching consensus, demonstrating the potential of this approach.

STEAL THIS

The ability to create a consensus-driven debate system using LLMs can provide valuable insights into their decision-making processes and has potential applications in various domains.

https://www.reddit.com/r/LLMDevs/comments/1uz1qv/i_made_llms_debate_each_other_about_the_question/

I built an open-source profiler for voice agents

reddit

The author built an open-source profiler for voice agents, called VoiceGateway, which allows developers to add a single line of code to their voice agent application and gain insights into the cost, latency, and quality of each call.

How it works. VoiceGateway works by integrating with LiveKit and Pipecat, providing a comprehensive view of the voice agent's performance. The author demonstrates the profiler's capabilities and provides a clear explanation of its inner workings.

STEAL THIS

The creation of an open-source profiler for voice agents can significantly improve the development and optimization of voice agent applications, enabling developers to better understand and optimize their systems.

https://www.reddit.com/r/AI_Agents/comments/1uz737x/i_built_an_opensource_profiler_for_voice_agents/

QUICK HITS

Show HN: Sandboxd – Self-Hosted Lovable HACKERNEWS

Sandboxd is a self-hosted infrastructure for building and managing AI-powered applications, providing a comprehensive platform for developers.

<https://github.com/tastyeffectco/sandboxd>

KlaaAI/klaatcode GITHUB

Klaatcode is an open-source AI coding agent that provides Claude Code-grade accuracy with smart model routing, supporting various AI models and cutting costs.

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

CyberSunil/LLMVault GITHUB

LLMVault is an intentionally vulnerable OWASP LLM Top 10 training platform for AI security, prompt injection, RAG security, agent security, and GenAI penetration testing.

<https://github.com/CyberSunil/LLMVault>

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

Ternary-Bonsai-27B-gguf is a trending model on HuggingFace, providing a text-generation capability with llama.cpp and gaining significant attention in the AI community.

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

agentic-commerce-lab/pawl GITHUB

Pawl is a registry-as-ground-truth verification and autonomous research framework for agent-driven development, providing a comprehensive platform for building and testing AI-powered systems.

<https://github.com/agentic-commerce-lab/pawl>

zeraix/zeraix GITHUB

Zeraix is an open-source local AI workspace, advancing on-device inference and providing a platform for developers to build and deploy AI-powered applications.

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

MemTensor/memmy-agent GITHUB

Memmy-agent is an open-source agent that provides a shared memory for every AI, enabling developers to build and deploy AI-powered applications with a unified memory architecture.

<https://github.com/MemTensor/memmy-agent>

try1004/FlowLens-AgentOps GITHUB

FlowLens-AgentOps is an observable diagnostics, failure attribution, metrics, and no-key replay for multi-agent runtimes, providing a comprehensive platform for building and testing AI-powered systems.

<https://github.com/try1004/FlowLens-AgentOps>

pur4v/p2ptokens GITHUB

P2ptokens is a distributed, peer-to-peer, decentralized network for LLM inference, providing a platform for developers to build and deploy AI-powered applications with a decentralized architecture.

<https://github.com/pur4v/p2ptokens>

empero-ai/Qwythos-9B-Claude-Mythos-5-1M-GGUF HUGGINGFACE

Qwythos-9B-Claude-Mythos-5-1M-GGUF is a trending model on Hugging Face, providing an image-text-to-text capability with gguf and gaining significant attention in the AI community.

<https://huggingface.co/empero-ai/Qwythos-9B-Claude-Mythos-5-1M-GGUF>

bottlecapai/ThinkingCap-Qwen3.6-27B HUGGINGFACE

ThinkingCap-Qwen3.6-27B is a trending model on Hugging Face, providing an image-text-to-text capability with transformers and gaining significant attention in the AI community.

<https://huggingface.co/bottlecapai/ThinkingCap-Qwen3.6-27B>

conradlocke/krea2-identity-edit HUGGINGFACE

Krea2-identity-edit is a trending model on Hugging Face, providing a model capability and gaining significant attention in the AI community.

<https://huggingface.co/conradlocke/krea2-identity-edit>

Bonsai 27B WebGPU Kernels HUGGINGFACE

Bonsai 27B WebGPU Kernels is a trending Space on Hugging Face, providing a run-time environment for a 1-bit 27B LLM locally in the browser on WebGPU.

<https://huggingface.co/spaces/webml-community/bonsai-webgpu-kernels>

Show HN: ReasonGate- An explainable gate that blocks LLM prompt injection HACKERNEWS

ReasonGate is an explainable gate that blocks LLM prompt injection, providing a security solution for AI-powered applications.

<https://github.com/cgrtml/reasongate>

Show HN: Forall – Spec-driven AI coding with formal verification HACKERNEWS

Forall is a spec-driven AI coding framework with formal verification, providing a comprehensive platform for building and testing AI-powered applications.

<https://github.com/astrio-labs/forall>

aglio-lab/ai-evaluation-tools GITHUB

AI-evaluation-tools is a comprehensive list of AI evaluation tools, providing a platform for developers to build and test AI-powered applications.

<https://github.com/aglio-lab/ai-evaluation-tools>
