



TreScout

RELATÓRIO DIÁRIO DE TECNOLOGIA

As tendências de tecnologia de hoje

15 de agosto de 2026 (sábado)

EM POUCOS GOLES

A agenda atual é baseada em novas técnicas de busca que aumentam a capacidade de raciocínio em grandes modelos de linguagem e na rápida expansão de arquiteturas de modelos de código aberto. Em particular, 30 bilhões de modelos de nível de parâmetro, como Qwen e Muse, continuam a ultrapassar os limites em termos de eficiência de hardware e capacidades de codificação.

35 EM DESTAQUE · ~10 MIN DE LEITURA

17 GitHub
Trending

5 Hacker
News

5 HuggingFace ·
modelos do
dia

5 HuggingFace ·
artigos do dia

3 Lobsters

01 cathrynlavery/diagram-design

O design de diagramas desenvolvido para Claude Code oferece 29 tipos diferentes de diagramas editoriais preparados diretamente em formato HTML e SVG sem usar software complexo. Esta ferramenta facilita a criação de gráficos fora do padrão em documentação técnica com seu design simples que reduz a confusão visual.

HTML · ★ 17.449 · +3.646 hoje <https://github.com/cathrynlavery/diagram-design>

02 cactus-compute/needle

Desenvolvido pela Cactus Compute, o Needle é um pequeno modelo básico de 14 megabytes que pode ser executado em sistemas com hardware limitado, como smartphones e dispositivos vestíveis. Este projeto oferece uma alternativa leve e otimizada para executar recursos de IA em dispositivos locais (computação de ponta).

Python · ★ 5.678 · +662 hoje <https://github.com/cactus-compute/needle>

03 megadose/holehe

Holehe é uma ferramenta Python que consulta se um endereço de e-mail é usado em plataformas como Twitter ou Instagram. Ele detecta a existência de contas associadas ao email relevante usando funções de redefinição de senha.

Python · ★ 12.887 · +427 hoje <https://github.com/megadose/holehe>

04 macro-inc/macro

Macro combina ferramentas de negócios como e-mail, chat, documentos, gerenciamento de tarefas e gerenciamento de relacionamento com o cliente (CRM) em um único espaço de trabalho. Graças à memória de inteligência artificial compartilhada, centraliza a colaboração da equipe conectando os dados em todas essas plataformas.

Rust · ★ 3.068 · +436 hoje <https://github.com/macro-inc/macro>

05 smicallef/spiderfoot

SpiderFoot é uma ferramenta Python que automatiza a coleta de inteligência de código aberto (OSINT) para inteligência de ameaças e processos de mapeamento de superfície de ataque. Ele agrega automaticamente informações sobre seus ativos digitais, verificando fontes de dados na Internet.

Python · ★ 20.983 · +293 hoje <https://github.com/smicallef/spiderfoot>

06 citrolabs/ego-lite

Ego-lite é um navegador rápido projetado para agentes de inteligência artificial realizarem automação do navegador. Ele permite que os usuários tomem medidas sem exigir intervenção manual, compartilhando o estado do navegador conectado com assistentes de codificação, como Codex ou Claude Code.

JavaScript · ★ 10.442 · +165 hoje <https://github.com/citrolabs/ego-lite>

07 **holaboss-ai/holaOS**

holaOS é uma plataforma de código aberto que permite aos usuários gerenciar agentes de IA como Claude Code e Codex em um único espaço de trabalho com navegador, arquivos e mais de cem integrações. Oferecendo uma estrutura de memória compartilhada, esta ferramenta permite utilizar seus próprios modelos ou escolher modelos prontos.

TypeScript · ★ 7.355 · +769 hoje <https://github.com/holaboss-ai/holaOS>

08 **github/spec-kit**

Desenvolvido pelo GitHub, Spec-kit é um kit de ferramentas Python que acelera processos de desenvolvimento orientados por especificações. Tem como objetivo aumentar a qualidade do código e a disciplina de desenvolvimento, estabelecendo padrões em projetos de software.

Python · ★ 128.591 · +1.160 hoje <https://github.com/github/spec-kit>

09 **lightningpixel/modly**

Modly é um aplicativo de desktop que gera modelos tridimensionais a partir de imagens ou comandos de texto. Funciona sem a necessidade de serviços baseados em nuvem, realizando todas as operações na unidade de processamento gráfico (GPU) local.

TypeScript · ★ 5.998 · +579 hoje <https://github.com/lightningpixel/modly>

10 **infiniflow/ragflow**

RAGFlow é um mecanismo de geração assistida por busca (RAG) de código aberto que cria uma camada de contexto avançada para grandes modelos de linguagem (LLM). O sistema combina a tecnologia RAG com recursos de agente para automatizar processos complexos de processamento de dados.

Go · ★ 88.424 · +473 hoje <https://github.com/infiniflow/ragflow>

11 **cursor/plugins**

A ferramenta de edição de código Cursor publicou uma especificação oficial de plugins para que os desenvolvedores possam criar seus próprios plugins. Este padrão permite estender a funcionalidade do ambiente de desenvolvimento de software com plugins baseados em TypeScript.

TypeScript · ★ 2.831 · +41 hoje <https://github.com/cursor/plugins>

12 **deepseek-ai/awesome-deepseek-agent**

compartilhado no GitHub, awesome-deepseek-agent é uma coleção de recursos que lista agentes de inteligência artificial (agentes de IA) desenvolvidos com modelos e ferramentas DeepSeek neste campo. Reúne projetos e integrações que ajudam desenvolvedores a construir sistemas autônomos utilizando a infraestrutura DeepSeek.

★ 5.755 · +222 hoje <https://github.com/deepseek-ai/awesome-deepseek-agent>

13 **semantica-agi/semantica**

A Semantica oferece uma infraestrutura nativa de gráficos que fornece gerenciamento contextual de dados e responsabilidade em sistemas de inteligência artificial. Desenvolvida com Python, esta ferramenta ajuda modelos de inteligência artificial a produzir resultados mais confiáveis e rastreáveis, gerenciando relacionamentos de dados complexos.

Python · ★ 7.580 · +1.181 hoje <https://github.com/semantica-agi/semantica>

14 **rustdesk/rustdesk**

Desenvolvido com Rust, RustDesk é um aplicativo de desktop remoto de código aberto que você pode hospedar em seu próprio servidor. Oferecendo acesso e controle remotos, este software cria uma solução alternativa aos serviços pagos como o TeamViewer.

Rust · ★ 120.685 · +143 hoje <https://github.com/rustdesk/rustdesk>

15 **OpenCut-app/OpenCut**

OpenCut é um projeto de software desenvolvido como uma alternativa de código aberto à ferramenta de edição de vídeo CapCut. Este aplicativo, criado com a linguagem TypeScript, permite aos usuários gerenciar seus processos de edição de vídeo em seus próprios sistemas locais.

TypeScript · ★ 83.233 · +255 hoje <https://github.com/OpenCut-app/OpenCut>

16 **unslothai/unsloth**

Unsloth fornece um ambiente nativo de trabalho e ajuste fino para modelos de linguagem grandes (LLM) e modelos de renderização. Esta ferramenta baseada em Python permite treinar modelos populares de inteligência artificial com mais rapidez e menor uso de memória.

Python · ★ 71.559 · +501 hoje <https://github.com/unslothai/unsloth>

17 **ToolJet/ToolJet**

ToolJet é uma plataforma de criação de aplicativos de código aberto para o desenvolvimento de ferramentas internas, painéis e agentes de inteligência artificial. Essa infraestrutura baseada em JavaScript permite que as empresas digitalizem fluxos de trabalho complexos, reduzindo a necessidade de escrever código.

JavaScript · ★ 39.148 · +132 hoje <https://github.com/ToolJet/ToolJet>

01 GLM-5.3: Frontier coding with emergent cyber capabilities

GLM-5.3, publicado pela Z.ai, é um novo modelo de linguagem grande usado em processos de desenvolvimento de software e inclui recursos de segurança cibernética. O modelo é otimizado para alto desempenho em tarefas complexas, como codificação e análise de sistema.

1.054 pontos · 522 comentários <https://z.ai/blog/glm-5.3>

Discussão: news.ycombinator.com/item?id=49294997 (1.054 pontos · 522 comentários)

02 Qwen 3.8 27B

Qwen 3.8 27B, o grande modelo de linguagem desenvolvido pela Alibaba, oferece inferência de alto desempenho com formato de dados de baixa precisão (FP8). O modelo visa otimizar o equilíbrio entre eficiência e velocidade em tarefas complexas com sua estrutura de 27 bilhões de parâmetros.

978 pontos · 629 comentários <https://huggingface.co/Qwen/Qwen3.8-27B-FP8>

Discussão: news.ycombinator.com/item?id=49299605 (978 pontos · 629 comentários)

03 Why does Opus 5 feel worse to work with?

Os usuários afirmam que as versões atuais do Claude 3 Opus, o grande modelo de linguagem desenvolvido pela Anthropic, apresentam desempenho inferior às versões anteriores e cometem mais erros em tarefas complexas. Isso questiona os efeitos das mudanças na qualidade da resposta do modelo na experiência do desenvolvedor.

813 pontos · 745 comentários <https://mun-logadan.github.io/why-does-opus-5-feel-worse/>

Discussão: news.ycombinator.com/item?id=49296740 (813 pontos · 745 comentários)

04 Firefox is now the last major browser that still supports uBlock Origin

O Google Chrome concluiu sua migração do Manifest V3, que restringe o funcionamento de extensões como o bloqueador de conteúdo uBlock Origin. Apesar dessa mudança tecnológica, o Firefox é o único navegador importante que continua a oferecer suporte ao popular bloqueador de anúncios uBlock Origin.

594 pontos · 216 comentários <https://www.pcworld.com/article/3212428/firefox-is-now-the-last-major-browser-that-still-supports-ublock-origin.html>

Discussão: news.ycombinator.com/item?id=49303202 (594 pontos · 216 comentários)

05 Seven books I keep close because I love them

Mark Dominus, conhecido por seu trabalho em desenvolvimento de software e linguística, compartilha sete livros básicos que ele constantemente consulta em sua biblioteca pessoal. A lista concentra-se em obras que moldam a mentalidade dos profissionais do mundo do software com sua profundidade técnica e qualidade literária.

317 pontos · 141 comentários <https://blog.plover.com/2026/08/02/>

Discussão: news.ycombinator.com/item?id=49299675 (317 pontos · 141 comentários)

01 Qwen/Qwen3.8-27B

Desenvolvido pela Alibaba, Qwen2.5-27B é um modelo multimodal de linguagem grande que pode processar dados de texto e imagem. O modelo se destaca pela capacidade de gerar respostas baseadas em texto por meio da análise de conteúdo visual complexo.

♥ 9.189 · 2 downloads <https://huggingface.co/Qwen/Qwen3.8-27B>

02 meta-models/Muse-Glimmer-30B

Muse-Glimmer-30B, desenvolvido pela Meta, é um modelo de linguagem multimodal (modelo multimodal) que pode processar dados visuais e textuais. Este modelo facilita a compreensão do conteúdo visual, analisando imagens complexas e gerando respostas baseadas em texto.

♥ 1.523 · 165.300 downloads <https://huggingface.co/meta-models/Muse-Glimmer-30B>

03 Qwen/Qwen3.8-2.4T-A95B

Desenvolvido pela Alibaba, Qwen3.8-2.4T-A95B é um modelo de linguagem abrangente que se concentra nos recursos de geração de texto. Treinado em grandes conjuntos de dados, este modelo visa entregar alto desempenho em tarefas complexas de processamento de linguagem.

♥ 927 · 3.832 downloads <https://huggingface.co/Qwen/Qwen3.8-2.4T-A95B>

04 unsloth/Qwen3.8-27B-GGUF

A Unsloth introduziu uma versão otimizada do modelo Qwen2.5-32B (formato GGUF) que reduz o uso de memória e aumenta a velocidade de treinamento. Esta versão permite que grandes modelos de linguagem sejam executados com mais eficiência, mesmo em sistemas de baixo custo.

♥ 887 · 0 downloads <https://huggingface.co/unsloth/Qwen3.8-27B-GGUF>

05 MiniMaxAI/MiniMax-H3

MiniMax-H3, desenvolvido pela MiniMaxAI, é um modelo generativo de IA que cria vídeo usando entradas visuais e de texto. Ele permite aos usuários converter conteúdo estático em imagens em movimento.

♥ 3.925 · 1.997.541 downloads <https://huggingface.co/MiniMaxAI/MiniMax-H3>

01 **Maglev: Sliding Recurrent Memory**

Maglev é uma arquitetura Transformer recorrente que generaliza o mecanismo de atenção de janela deslizante usando memória de tamanho fixo. Graças à sua estrutura paralelizável durante o treinamento, ele processa contextos longos de forma eficiente, combinando um pré-preenchimento e um modelo de decodificador com acesso ao histórico completo.

3 curtidas · 2 comentários <https://huggingface.co/papers/2608.02870>

02 **Thought-Level Beam Search for Reasoning**

A pesquisa de feixe em nível de pensamento é um método que se concentra no uso mais eficiente do poder de processamento em grandes modelos de raciocínio. Esta abordagem visa otimizar os processos de raciocínio direcionando recursos de hardware para as soluções mais promissoras.

4 curtidas · 2 comentários <https://huggingface.co/papers/2608.08020>

03 **RibAssist 3D: Biplanar Rib-Fracture Detection, Addressing, and Selective 3D Localization from CT-Derived Projections**

RibAssist 3D é um sistema de diagnóstico desenvolvido para detectar e posicionar fraturas de costelas em três dimensões (3D) em imagens de tomografia computadorizada (TC). O objetivo é determinar pontos de fratura com alta precisão, combinando imagens tiradas de dois ângulos diferentes.

0 curtidas · 2 comentários <https://huggingface.co/papers/2608.06914>

04 **Context-Matched Distillation: Teacher Causality for Autoregressive Video Distillation**

A destilação contextual é um método desenvolvido para fornecer baixa latência e controle on-line preciso na produção de vídeo interativo. Esta técnica treina modelos de vídeo autorregressivos, visando garantir que os modelos dos alunos se adaptem às restrições causais baseadas apenas em dados históricos.

4 curtidas · 2 comentários <https://huggingface.co/papers/2608.13391>

05 **From Inaudible Inputs to Model Failures: Low-Frequency Safety Risks in LALMs**

Grandes modelos de linguagem de áudio (LALMs) podem processar sinais de baixa frequência que os humanos não conseguem ouvir, afetando sua saída. Os pesquisadores revelam a vulnerabilidade dos modelos a esses sinais com o método denominado Bloqueio Intermitente de Baixa Frequência, que desenvolveram para testar essa vulnerabilidade.

2 curtidas · 2 comentários <https://huggingface.co/papers/2608.09158>

01 RISC-V: They Should Have Known Better

A arquitetura de processador de código aberto RISC-V foi criticada pelas complexidades no design de hardware e na arquitetura do conjunto de instruções (ISA). Os desenvolvedores de software argumentam que esta arquitetura contém alguns erros de design em termos de escalabilidade e segurança em comparação com os padrões existentes.

34 pontos · 23 comentários <https://dmitry.gr/?r=06.%20Thoughts&proj=12.%20RV>

Discussão: lobste.rs/s/pu3yzg/risc_v_they_should_have_known_better (34 pontos · 23 comentários)

02 ActivityPub Won by Being Boring

O protocolo descentralizado de rede social ActivityPub ganhou ampla aceitação graças à sua estrutura simples e sustentável que evita recursos complexos. Esta abordagem demonstra a importância da simplicidade no sucesso a longo prazo dos padrões abertos que permitem que as plataformas comuniquem entre si.

17 pontos · 8 comentários <https://o.ee/blog/activitypub-won-by-being-boring/>

Discussão: lobste.rs/s/ngz2wn/activitypub_won_by_being_boring (17 pontos · 8 comentários)

03 Curl Performance

A ferramenta de transferência de dados curl realiza otimizações no protocolo HTTP/3 e nos processos de gerenciamento de conexão para melhorar o desempenho da rede. Essas melhorias visam proporcionar uma transferência de dados mais eficiente, reduzindo a latência em sistemas de alto tráfego.

35 pontos · 6 comentários <https://daniel.haxx.se/blog/2026/08/14/curl-performance-2/>

Discussão: lobste.rs/s/avjfxw/curl_performance (35 pontos · 6 comentários)

foundation model

Estrutura básica de IA pré-treinada com grandes quantidades de dados para realizar muitas tarefas diferentes.

edge computing

Ganhar velocidade processando dados no ponto mais próximo do próprio dispositivo, em vez de servidores remotos no centro.

CRM

Sistema de gestão que permite às empresas acompanhar em um só lugar suas interações e informações históricas com seus clientes.

OSINT

Um método de coleta e análise de informações de fontes disponíveis publicamente na Internet.

browser state

Todas as informações armazenadas do seu navegador de internet, como abas abertas no momento, cookies e as contas nas quais você está conectado.

AI agents

Software autônomo de inteligência artificial que toma decisões por conta própria e utiliza ferramentas para atingir determinados objetivos.

spec-driven development

O processo de determinação detalhada de todos os requisitos e regras técnicas antes de começar a codificar o software.

GPU

Processador especial usado para processar imagens em um computador e acelerar cálculos matemáticos pesados, como inteligência artificial.

LLM

Modelos de inteligência artificial treinados com enormes dados que podem compreender a linguagem humana e produzir texto.

RAG

Um método que permite que a inteligência artificial se beneficie de documentos atuais e privados, além de seu próprio conhecimento, na produção de respostas.

plugin specification

Um conjunto de regras técnicas que devem ser seguidas para fazer um complemento externo a um software.

graph-native

Uma estrutura de banco de dados especial que armazena e processa dados com base em seus relacionamentos entre si.

fine-tuning

O processo de reciclagem de uma inteligência artificial treinada para uso geral com um pequeno conjunto de dados para torná-la mais especialista em um assunto específico.

cyber capabilities

Capacidades técnicas e ferramentas utilizadas para proteger ou interferir em sistemas digitais.

large language model

Modelos de inteligência artificial treinados com enormes dados que podem compreender a linguagem humana e produzir texto.

FP8

Um formato de dados que representa números em um espaço menor para agilizar os cálculos de IA.
