



TreScout

INFORME DIARIO DE TECNOLOGÍA

Las tendencias tecnológicas de hoy

15 de agosto de 2026 (sábado)

EN POCOS SORBOS

La agenda de hoy se basa en nuevas técnicas de búsqueda que aumentan la capacidad de razonamiento en grandes modelos de lenguaje y la rápida expansión de las arquitecturas de modelos de código abierto. En particular, 30 mil millones de modelos a nivel de parámetros, como Qwen y Muse, continúan superando los límites en términos de eficiencia de hardware y capacidades de codificación.

35 DESTACADOS · ~10 MIN DE LECTURA

17 GitHub
Trending

5 Hacker
News

5 HuggingFace ·
modelos del
día

5 HuggingFace ·
artículos del
día

3 Lobsters

01 cathrynlavery/diagram-design

El diseño de diagramas desarrollado para Claude Code ofrece 29 tipos de diagramas editoriales diferentes preparados directamente en formato HTML y SVG sin utilizar software complejo. Esta herramienta facilita la creación de gráficos no estándar en documentación técnica con su diseño simple que reduce el desorden visual.

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

02 cactus-compute/needle

Desarrollado por Cactus Compute, Needle es un pequeño modelo básico de 14 megabytes que puede ejecutarse en sistemas con restricciones de hardware, como teléfonos inteligentes y dispositivos portátiles. Este proyecto ofrece una alternativa ligera optimizada para ejecutar capacidades de IA en dispositivos locales (computación de borde).

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

03 megadose/holehe

Holehe es una herramienta de Python que consulta si una dirección de correo electrónico se utiliza en plataformas como Twitter o Instagram. Detecta la existencia de cuentas asociadas al correo electrónico correspondiente mediante funciones de restablecimiento de contraseña.

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

04 macro-inc/macro

Macro combina herramientas comerciales como correo electrónico, chat, documentos, gestión de tareas y gestión de relaciones con el cliente (CRM) en un único espacio de trabajo. Gracias a la memoria de inteligencia artificial compartida, centraliza la colaboración en equipo conectando los datos en todas estas plataformas.

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

05 smicallef/spiderfoot

SpiderFoot es una herramienta de Python que automatiza la recopilación de inteligencia de código abierto (OSINT) para procesos de mapeo de superficies de ataque y inteligencia de amenazas. Agrega automáticamente información sobre sus activos digitales escaneando fuentes de datos en Internet.

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

06 citrolabs/ego-lite

Ego-lite es un navegador web rápido diseñado para que agentes de inteligencia artificial realicen la automatización del navegador. Permite a los usuarios tomar medidas sin necesidad de intervención manual al compartir el estado de su navegador conectado con asistentes de codificación como Codex o Claude Code.

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

07 holaboss-ai/holaOS

holaOS es una plataforma de código abierto que permite a los usuarios gestionar agentes de IA como Claude Code y Codex en un único espacio de trabajo con el navegador, los archivos y más de cien integraciones. Al ofrecer una estructura de memoria compartida, esta herramienta le permite utilizar sus propios modelos o elegir modelos ya preparados.

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

08 github/spec-kit

Desarrollado por GitHub, Spec-kit es un conjunto de herramientas de Python que acelera los procesos de desarrollo basados en especificaciones. Su objetivo es aumentar la calidad del código y la disciplina de desarrollo estableciendo estándares en proyectos de software.

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

09 lightningpixel/modly

Modly es una aplicación de escritorio que genera modelos tridimensionales a partir de imágenes o comandos de texto. Funciona sin la necesidad de servicios basados en la nube al realizar todas las operaciones en la unidad de procesamiento de gráficos (GPU) local.

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

10 infiniflow/ragflow

RAGFlow es un motor de generación asistida por recuperación (RAG) de código abierto que crea una capa de contexto avanzada para modelos de lenguaje grandes (LLM). El sistema combina la tecnología RAG con capacidades de agente para automatizar procesos complejos de procesamiento de datos.

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

11 cursor/plugins

La herramienta de edición de código Cursor ha publicado una especificación de complemento oficial para que los desarrolladores puedan crear sus propios complementos. Este estándar permite ampliar la funcionalidad del entorno de desarrollo de software con complementos basados en TypeScript.

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

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

compartido en GitHub, awesome-deepseek-agent es una colección de recursos que enumera agentes de inteligencia artificial (agentes de IA) desarrollados con modelos y herramientas de DeepSeek en este campo. Reúne proyectos e integraciones que ayudan a los desarrolladores a construir sistemas autónomos utilizando la infraestructura de DeepSeek.

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

13 **semantica-agi/semantica**

Semantica ofrece una infraestructura nativa de gráficos que proporciona gestión de datos contextuales y responsabilidad en sistemas de inteligencia artificial. Desarrollada con Python, esta herramienta ayuda a los modelos de inteligencia artificial a producir resultados más confiables y rastreables mediante la gestión de relaciones de datos complejas.

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

14 **rustdesk/rustdesk**

Desarrollada con Rust, RustDesk es una aplicación de escritorio remoto de código abierto que puedes alojar en tu propio servidor. Al ofrecer acceso y control remotos, este software crea una solución alternativa a los servicios pagos como TeamViewer.

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

15 **OpenCut-app/OpenCut**

OpenCut es un proyecto de software desarrollado como una alternativa de código abierto a la herramienta de edición de vídeo CapCut. Esta aplicación, creada con el lenguaje TypeScript, permite a los usuarios gestionar sus procesos de edición de vídeo en sus propios sistemas locales.

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

16 **unslothai/unsloth**

Unsloth proporciona un entorno nativo de trabajo y ajuste para modelos de lenguaje grandes (LLM) y modelos de renderizado. Esta herramienta basada en Python le permite entrenar modelos populares de inteligencia artificial más rápido y con menor uso de memoria.

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

17 **ToolJet/ToolJet**

ToolJet es una plataforma de creación de aplicaciones de código abierto para desarrollar herramientas internas, paneles y agentes de inteligencia artificial. Esta infraestructura basada en JavaScript permite a las empresas digitalizar flujos de trabajo complejos al reducir la necesidad de escribir código.

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

01 GLM-5.3: Frontier coding with emergent cyber capabilities

GLM-5.3, publicado por Z.ai, es un nuevo modelo de lenguaje grande que se utiliza en procesos de desarrollo de software e incluye capacidades de seguridad cibernética. El modelo está optimizado para un alto rendimiento en tareas complejas como codificación y análisis de sistemas.

1.054 puntos · 522 comentarios <https://z.ai/blog/glm-5.3>

Discusión: news.ycombinator.com/item?id=49294997 (1054 puntos · 522 comentarios)

02 Qwen 3.8 27B

Qwen 3.8 27B, el modelo de lenguaje grande desarrollado por Alibaba, ofrece inferencia de alto rendimiento con formato de datos de baja precisión (FP8). El modelo tiene como objetivo optimizar el equilibrio entre eficiencia y velocidad en tareas complejas con su estructura de 27 mil millones de parámetros.

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

Discusión: news.ycombinator.com/item?id=49299605 (978 puntos · 629 comentarios)

03 Why does Opus 5 feel worse to work with?

Los usuarios afirman que las versiones actuales de Claude 3 Opus, el modelo de lenguaje grande desarrollado por Anthropic, tienen un rendimiento inferior que las versiones anteriores y cometen más errores en tareas complejas. Esto pone en duda los efectos de los cambios en la calidad de respuesta del modelo en la experiencia del desarrollador.

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

Discusión: news.ycombinator.com/item?id=49296740 (813 puntos · 745 comentarios)

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

Google Chrome ha completado su migración Manifest V3, que restringe el funcionamiento de extensiones como el bloqueador de contenido uBlock Origin. A pesar de este cambio tecnológico, Firefox es el único navegador importante que sigue siendo compatible con el popular bloqueador de anuncios uBlock Origin.

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

Discusión: news.ycombinator.com/item?id=49303202 (594 puntos · 216 comentarios)

05 Seven books I keep close because I love them

Mark Dominus, conocido por su trabajo en desarrollo de software y lingüística, comparte siete libros básicos a los que consulta constantemente en su biblioteca personal. La lista se centra en obras que moldean la mentalidad de los profesionales del mundo del software con su profundidad técnica y calidad literaria.

317 puntos · 141 comentarios <https://blog.plover.com/2026/08/02/>

Discusión: news.ycombinator.com/item?id=49299675 (317 puntos · 141 comentarios)

01 Qwen/Qwen3.8-27B

Desarrollado por Alibaba, Qwen2.5-27B es un modelo de lenguaje grande multimodal que puede procesar datos de texto e imágenes. El modelo se destaca por su capacidad para generar respuestas basadas en texto mediante el análisis de contenido visual complejo.

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

02 meta-models/Muse-Glimmer-30B

Muse-Glimmer-30B, desarrollado por Meta, es un modelo de lenguaje multimodal (modelo multimodal) que puede procesar datos tanto visuales como de texto. Este modelo hace que el contenido visual sea más fácil de entender al analizar imágenes complejas y generar respuestas basadas en texto.

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

03 Qwen/Qwen3.8-2.4T-A95B

Desarrollado por Alibaba, Qwen3.8-2.4T-A95B es un modelo de lenguaje integral que se centra en capacidades de generación de texto. Este modelo, entrenado en grandes conjuntos de datos, tiene como objetivo ofrecer un alto rendimiento en tareas complejas de procesamiento de lenguaje.

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

04 unsloth/Qwen3.8-27B-GGUF

Unsloth ha introducido una versión optimizada del modelo Qwen2.5-32B (formato GGUF) que reduce el uso de memoria y aumenta la velocidad de entrenamiento. Esta versión permite que los modelos de lenguaje grandes se ejecuten de manera más eficiente, incluso en sistemas de gama baja.

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

05 MiniMaxAI/MiniMax-H3

MiniMax-H3, desarrollado por MiniMaxAI, es un modelo de IA generativa que crea videos utilizando entradas visuales y de texto. Permite a los usuarios convertir contenido estático en imágenes en movimiento.

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

01 Maglev: Sliding Recurrent Memory

Maglev es una arquitectura Transformer recurrente que generaliza el mecanismo de atención de la ventana deslizante mediante el uso de memoria de tamaño fijo. Gracias a su estructura paralelizable durante el entrenamiento, procesa contextos largos de manera eficiente combinando un modelo de precarga y decodificador con acceso al historial completo.

3 me gusta · 2 comentarios <https://huggingface.co/papers/2608.02870>

02 Thought-Level Beam Search for Reasoning

La búsqueda de haces a nivel de pensamiento es un método que se centra en el uso más eficiente de la potencia de procesamiento en grandes modelos de razonamiento. Este enfoque tiene como objetivo optimizar los procesos de razonamiento dirigiendo los recursos de hardware a las soluciones más prometedoras.

4 me gusta · 2 comentarios <https://huggingface.co/papers/2608.08020>

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

RibAssist 3D es un sistema de diagnóstico desarrollado para detectar y posicionar fracturas costales en tres dimensiones (3D) en imágenes de tomografía computarizada (TC). Su objetivo es determinar puntos de fractura con alta precisión haciendo coincidir imágenes tomadas desde dos ángulos diferentes.

0 me gusta · 2 comentarios <https://huggingface.co/papers/2608.06914>

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

La destilación adaptada al contexto es un método desarrollado para proporcionar baja latencia y control en línea preciso en la producción de videos interactivos. Esta técnica entrena modelos de video autorregresivos, con el objetivo de garantizar que los modelos de los estudiantes se adapten a restricciones causales basadas únicamente en datos históricos.

4 me gusta · 2 comentarios <https://huggingface.co/papers/2608.13391>

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

Los grandes modelos de audiolenguaje (LALM) pueden procesar señales de baja frecuencia que los humanos no pueden oír, lo que afecta su salida. Los investigadores revelan la vulnerabilidad de los modelos a estas señales con el método llamado Bloqueo intermitente de baja frecuencia, que desarrollaron para probar esta vulnerabilidad.

2 me gusta · 2 comentarios <https://huggingface.co/papers/2608.09158>

01 RISC-V: They Should Have Known Better

La arquitectura de procesador de código abierto RISC-V ha sido criticada por las complejidades del diseño de hardware y la arquitectura del conjunto de instrucciones (ISA). Los desarrolladores de software argumentan que esta arquitectura contiene algunos errores de diseño en cuanto a escalabilidad y seguridad en comparación con los estándares existentes.

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

Discusión: lobste.rs/s/pu3yzg/risc_v_they_should_have_known_better (34 puntos · 23 comentarios)

02 ActivityPub Won by Being Boring

El protocolo descentralizado de redes sociales ActivityPub ha ganado una amplia aceptación gracias a su estructura simple y sostenible que evita funciones complejas. Este enfoque demuestra la importancia de la simplicidad en el éxito a largo plazo de los estándares abiertos que permiten que las plataformas se comuniquen entre sí.

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

Discusión: lobste.rs/s/ngz2wn/activitypub_won_by_being_boring (17 puntos · 8 comentarios)

03 Curl Performance

La herramienta de transferencia de datos curl realiza optimizaciones en el protocolo HTTP/3 y los procesos de administración de conexiones para mejorar el rendimiento de la red. Estas mejoras tienen como objetivo proporcionar una transferencia de datos más eficiente al reducir la latencia en sistemas de alto tráfico.

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

Discusión: lobste.rs/s/avjfxw/curl_performance (35 puntos · 6 comentarios)

foundation model

Estructura básica de IA que está previamente entrenada con cantidades masivas de datos para realizar muchas tareas diferentes.

edge computing

Ganar velocidad procesando datos en el punto más cercano al propio dispositivo, en lugar de servidores remotos en el centro.

CRM

Sistema de gestión que permite a las empresas realizar un seguimiento de sus interacciones e información histórica con sus clientes en un solo lugar.

OSINT

Un método para recopilar y analizar información de fuentes disponibles públicamente en Internet.

browser state

Toda la información almacenada en su navegador de Internet, como las pestañas abiertas actualmente, las cookies y las cuentas en las que ha iniciado sesión.

AI agents

Software de inteligencia artificial autónomo que toma decisiones por sí solo y utiliza herramientas para lograr determinados objetivos.

spec-driven development

El proceso de determinar todos los requisitos y reglas técnicas en detalle antes de comenzar a codificar el software.

GPU

Procesador especial utilizado para procesar imágenes en una computadora y acelerar cálculos matemáticos pesados, como la inteligencia artificial.

LLM

Modelos de inteligencia artificial entrenados con enormes datos que pueden comprender el lenguaje humano y producir texto.

RAG

Un método que permite a la inteligencia artificial beneficiarse de documentos actuales y privados además de su propio conocimiento a la hora de producir respuestas.

plugin specification

Un conjunto de reglas técnicas que se deben seguir para crear un complemento externo a un software.

graph-native

Una estructura de base de datos especial que almacena y procesa datos en función de sus relaciones entre sí.

fine-tuning

El proceso de reentrenar una inteligencia artificial entrenada de propósito general con un pequeño conjunto de datos para hacerla más experta en un tema específico.

cyber capabilities

Capacidades y herramientas técnicas utilizadas para proteger o interferir con los sistemas digitales.

large language model

Modelos de inteligencia artificial entrenados con enormes datos que pueden comprender el lenguaje humano y producir texto.

FP8

Un formato de datos que representa números en un espacio más pequeño para agilizar los cálculos de IA.
