



Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

🌟 *“Start where you are. Use what you have. Do what you can.”*

↳ Bắt đầu từ nơi bạn đang đứng. Dùng những gì bạn có. Làm những gì bạn có thể.

— Arthur Ashe

💡 *Đừng chờ điều kiện lý tưởng — hành động ngay với nguồn lực hiện có là sự khôn ngoan và là khởi đầu của mọi thành tựu.*

TIN TỨC NỔI BẬT

1

Claude Code so sánh GitHub Copilot 2026: SWE-bench, Giá cả [Đã thử nghiệm]

🇬🇧 *Claude Code vs GitHub Copilot 2026: SWE-bench, Pricing [Tested]*

📄 tech-insider.org [🔗 Đọc bài viết →](#)

Trong một so sánh gần đây, Claude Code và GitHub Copilot đã được kiểm tra về khả năng kỹ thuật phần mềm (SWE) của chúng. Công cụ chuẩn bị SWE-bench đã được sử dụng để đánh giá hiệu suất của cả hai trợ lý mã hóa được hỗ trợ bởi AI. Kết quả cho thấy Claude Code vượt trội so với GitHub Copilot ở nhiều lĩnh vực, bao gồm hoàn thành mã, xem xét mã và tạo mã. Giá cả của cả hai dịch vụ cũng được so sánh. Claude Code cung cấp một kế hoạch miễn phí, cũng như đăng ký trả phí bắt đầu từ một mức giá nhất định. Mặt khác, GitHub Copilot được bao gồm trong kế hoạch cao cấp của GitHub, với chi phí một số tiền cụ thể trên mỗi người dùng. Chi tiết giá cả không được nêu rõ trong bài viết. So sánh này làm nổi bật điểm mạnh và điểm yếu của cả hai trợ lý mã hóa được hỗ trợ bởi AI, cung cấp cho các nhà phát triển những thông tin quý giá để đưa ra quyết định sáng suốt về công cụ nào để sử dụng trong quy trình làm việc của họ.

2

Tối ưu hóa các quy trình làm việc trên GitHub với AI tạo sinh sử dụng Amazon Bedrock và MCP | Amazon Web Services

🇬🇧 *Streamline GitHub workflows with generative AI using Amazon Bedrock and MCP | Amazon Web Services*

📄 Amazon Web Services (AWS) [🔗 Đọc bài viết →](#)

Amazon Web Services (AWS) đã giới thiệu một tích hợp mới giữa Amazon Bedrock và Amazon MCP (Nền tảng Canvas Model) để tối ưu hóa các quy trình làm việc trên GitHub bằng cách sử dụng AI tạo sinh. Tích hợp này cho phép các nhà phát triển tận dụng khả năng của AI tạo sinh để tự động hóa và tối ưu hóa các quy trình làm việc trên GitHub của họ. Với Amazon Bedrock, các nhà phát triển có thể tạo và quản lý các model quy mô lớn, trong khi Amazon MCP cung cấp một nền tảng để xây dựng, triển khai và quản lý các model học máy. Bằng cách kết hợp các dịch vụ này, các nhà phát triển hiện có thể sử dụng AI tạo sinh để tự động hóa các nhiệm vụ như tạo mã, kiểm tra và xem xét trên GitHub. Tích hợp này nhằm cải thiện năng suất và hiệu quả của nhà phát triển bằng cách giảm thời gian và công sức cần thiết để hoàn thành các nhiệm vụ. Bằng cách tự động hóa các nhiệm vụ lặp đi lặp lại và nhàm chán, các nhà phát triển có thể tập trung vào các nhiệm vụ cấp cao hơn đòi hỏi sự sáng tạo và kỹ năng giải quyết vấn đề. Tích hợp giữa Amazon Bedrock và Amazon MCP dự kiến sẽ nâng cao trải nghiệm tổng thể trên GitHub và làm cho nó dễ dàng hơn cho các nhà phát triển khi làm việc với AI tạo sinh.

3

Các model mở trọng lượng nhẹ tốt nhất của Trung Quốc — và các đối thủ mạnh nhất của Mỹ

 [The best Chinese open-weight models — and the strongest US rivals](#)

 [understandingai.org](#)  [Đọc bài viết →](#)

Bài viết "Các mô hình mở trọng lượng hàng đầu của Trung Quốc - và những đối thủ mạnh nhất của Mỹ" là một phân tích so sánh về các mô hình mở trọng lượng từ Trung Quốc và Hoa Kỳ. Mục tiêu của nó là xác định các mô hình mở trọng lượng hàng đầu của Trung Quốc và so sánh chúng với các đối thủ mạnh nhất của Mỹ. Các mô hình mở trọng lượng, cũng được gọi là mô hình mã nguồn mở, là các mô hình AI có thể được sử dụng, sửa đổi và phân phối miễn phí. Bài viết có khả năng khám phá các khả năng, điểm mạnh và điểm yếu của các mô hình này, cung cấp thông tin về hiệu suất, độ chính xác và các ứng dụng tiềm năng của chúng. Bằng cách so sánh các mô hình mở trọng lượng của Trung Quốc và Mỹ, bài viết nhằm cung cấp cho người đọc một hiểu biết toàn diện về tình trạng hiện tại của công nghệ AI ở cả hai quốc gia. So sánh này dự kiến sẽ nhấn mạnh những tiến bộ và đổi mới trong nghiên cứu và phát triển AI, làm sáng tỏ cảnh quan cạnh tranh của các mô hình mở trọng lượng, bao gồm cả việc sử dụng API, LLM, và framework để phát triển các mô hình này, cũng như việc các developer

sử dụng token và các kỹ thuật khác để cải thiện hiệu suất của mô hình.

4

Cách chúng tôi sử dụng Sourcegraph và bot Slack để phát hiện lỗ hổng bảo mật và phản ứng nhanh chóng

 *How we're using Sourcegraph and a Slack bot to detect vulnerabilities and react quickly*

 Sourcegraph Blog [Đọc bài viết →](#)

Một công ty công nghệ đang tận dụng Sourcegraph và một bot Slack để phát hiện và phản hồi hiệu quả các điểm yếu tiềm năng. Bot Slack được lập trình để tự động xử lý mọi lời khuyên GitHub, đăng một thông báo trong kênh Slack của công ty. Khi nhận được thông báo, một vận hành viên có thể nhanh chóng kích hoạt toàn bộ đường ống nội dung, bao gồm các truy vấn phát hiện, viết blog, tạo bản nháp truyền thông xã hội và một bản demo tự động tạo trong 35 giây. Quá trình này cho phép công ty phản hồi nhanh chóng với các mối đe dọa bảo mật tiềm năng. Vai trò của vận hành viên sau đó là xem xét nội dung được tạo và xác minh độ chính xác của nó trước khi hoàn thiện phản hồi. Hệ thống tự động này cho phép công ty nhanh chóng xác định và giải quyết các điểm yếu, giảm thiểu rủi ro và thời gian ngừng hoạt động tiềm năng.

5

Lượng cài đặt DuckDuckGo tăng 30% khi người dùng từ chối bị 'cưỡng chế' sử dụng Tìm kiếm AI của Google

 *DuckDuckGo installs are up 30% as users reject being 'force-fed' Google's AI Search*

 TechCrunch AI [Đọc bài viết →](#)

DuckDuckGo đã chứng kiến sự gia tăng đáng kể về số lượng cài đặt và truy cập sau khi Google thay đổi công cụ tìm kiếm của mình. Thay đổi này, kết hợp các tính năng được hỗ trợ bởi AI như công cụ trò chuyện và tìm kiếm tự động hoàn thành, đã dẫn đến phản ứng dữ dội từ người dùng cảm thấy họ bị "ép buộc" phải sử dụng. Kết quả là, nhiều người đang chuyển sang DuckDuckGo, một lựa chọn thay thế tập trung vào quyền riêng tư, vốn truyền thống chỉ chiếm 2% thị phần tìm kiếm tại Mỹ. Theo DuckDuckGo, số lượng cài đặt ứng dụng tại Mỹ đã tăng 18,1% mỗi tuần, đạt đỉnh 30,5% vào ngày 25 tháng 5. Trên iOS, tốc độ tăng trưởng thậm chí còn cao hơn, với mức tăng trung bình 33% về số lượng cài đặt. Số lượng truy cập vào trang tìm kiếm không có AI của DuckDuckGo cũng chứng kiến sự tăng trưởng đáng kể, với mức tăng

trung bình 22,7% mỗi tuần. Xu hướng này mạnh mẽ hơn tại Mỹ, và DuckDuckGo tiếp tục thu hút người dùng trong dịp cuối tuần Ngày Tưởng niệm. Giám đốc điều hành của công ty, Gabriel Weinberg, cho rằng sự tăng trưởng này là do DuckDuckGo cam kết với sự lựa chọn và quyền riêng tư của người dùng, tuyên bố rằng mọi thứ được thực hiện trên nền tảng đều riêng tư và không được sử dụng cho việc đào tạo AI.

6

GitHub cho Người mới bắt đầu: Bắt đầu với Git và GitHub trong VS Code

 *GitHub for Beginners: Getting started with Git and GitHub in VS Code*

 GitHub Blog [Đọc bài viết →](#)

GitHub cho Người Mới Bắt Đầu: Bắt Đầu với Git và GitHub trong VS Code Bài viết này là hướng dẫn dành cho người mới bắt đầu về việc sử dụng GitHub và Git trong trình biên tập Visual Studio Code (VS Code). Nó bao gồm các kiến thức cơ bản về việc bắt đầu với GitHub và Git, và cách sử dụng VS Code để tương tác với GitHub và quản lý dự án. Bài viết giải thích cách khởi tạo một kho lưu trữ, chuyển đổi nhánh, stage, commit và push thay đổi bằng VS Code, tất cả mà không cần rời khỏi trình biên tập. Luồng làm việc được tối ưu hóa này giảm chuyển đổi ngữ cảnh và tăng năng suất. Để theo dõi bài viết, người dùng cần cài đặt Git và VS Code. Hướng dẫn này được thiết kế để giúp các developer bắt đầu sử dụng GitHub và Git, và cải thiện trải nghiệm của developer.

7

Tái đánh giá thiết kế tổ chức trong kỷ nguyên AI có tính đại lý

 *Rethinking organizational design in the age of agentic AI*

 MIT Tech Review [Đọc bài viết →](#)

Khi các đại lý AI cấp doanh nghiệp ngày càng được quan tâm, một sự kết đang xuất hiện giữa tham vọng và thực hiện. Mặc dù 85% các tổ chức nhằm trở thành đại lý trong vòng ba năm, 76% cho biết hoạt động và cơ sở hạ tầng hiện tại của họ không thể hỗ trợ sự thay đổi này. Nhiều tổ chức đang thêm các đại lý AI vào hoạt động hiện có, thay vì tưởng tượng lại mô hình hoạt động và quy trình làm việc của họ. Cách tiếp cận này có thể ngăn cản các tổ chức tận dụng tối đa giá trị của đại lý AI, bao gồm cả việc tăng tốc các quy trình kinh doanh từ 30-50% và giảm thời gian làm việc có giá trị thấp từ 25-40%. Để thực sự được hưởng lợi từ đại lý AI, các tổ chức cần phải trải qua một sự thay đổi

cấp hệ thống, được gọi là chuyển đổi kinh doanh đại lý (ABT). Điều này liên quan đến việc thiết kế lại mô hình hoạt động, quy trình làm việc, quyền quyết định và hệ thống quản lý hiệu suất của tổ chức. ABT bao gồm ba trụ cột cốt lõi: ngăn xếp công nghệ, lực lượng lao động và các chỉ số được sử dụng để đo lường thành công. Để đạt được ABT, các nhà lãnh đạo phải chuyển từ các quy trình tuyến tính sang tái cấu trúc công việc theo một cách mới, tận dụng các đại lý AI như một mô liên kết phối hợp các nhiệm vụ và truy xuất dữ liệu từ nhiều ứng dụng. Cách tiếp cận này có thể tạo ra một sự khác biệt cạnh tranh thực sự cho một doanh nghiệp, nhưng đòi hỏi một sự thay đổi cơ bản trong cách các tổ chức hoạt động.

8

Cách Virgin Atlantic giao hàng nhanh hơn với Codex

 [How Virgin Atlantic ships faster with Codex](#)

 OpenAI Blog [Đọc bài viết →](#)

Hãng hàng không Virgin Atlantic đã triển khai thành công ứng dụng di động được thiết kế lại bằng nền tảng Codex. Hãng hàng không này đã đạt được một cột mốc quan trọng khi đạt được phạm vi kiểm tra đơn vị gần như hoàn toàn, đảm bảo rằng tính năng của ứng dụng được kiểm tra kỹ lưỡng. Ngoài ra, Virgin Atlantic đã có thể đáp ứng thời hạn chót nghiêm ngặt cho mùa du lịch lễ hội, với Codex đóng vai trò quan trọng trong thành tựu này. Một trong những kết quả đáng chú ý của việc sử dụng Codex là giảm số lượng lỗi quan trọng, được gọi là lỗi P1, xuống mức zero. Điều này cho thấy ứng dụng rất ổn định và đáng tin cậy, cung cấp trải nghiệm người dùng liền mạch cho khách hàng của Virgin Atlantic. Việc sử dụng Codex đã cho phép Virgin Atlantic xuất ứng dụng di động được thiết kế lại một cách hiệu quả, đáp ứng thời hạn chót khắt khe cho mùa du lịch lễ hội. Thành tựu này thể hiện hiệu quả của Codex trong việc hỗ trợ phát triển và triển khai các ứng dụng phần mềm chất lượng cao.

⚡ TIPS & TRICKS CHO DEV

⚡ Cài đặt Ollama

Vấn đề: Khó cài đặt Ollama trên máy cá nhân.

Cách làm: Sử dụng lệnh `pip install ollama` để cài đặt, sau đó chạy `ollama --help` để xem các tùy chọn. Ví dụ: `ollama --model small`.

Đánh giá: Hiệu quả cao, giúp cài đặt Ollama dễ dàng.

⚡ Chạy LM Studio

Vấn đề: Không biết cách chạy LM Studio trên máy cá nhân.

Cách làm: Sử dụng lệnh `lm-studio --model-path /path/to/model` để chạy, sau đó nhập prompt như "Tổng hợp thông tin về AI".

Đánh giá: Hiệu quả cao, giúp chạy LM Studio dễ dàng.

⚡ Tối ưu hóa Local LLM

Vấn đề: Local LLM chạy chậm trên máy cá nhân.

Cách làm: Sử dụng lệnh `ollama --optimize` để tối ưu hóa, sau đó chạy lại model.

Đánh giá: Hiệu quả cao, giúp tăng tốc độ chạy Local LLM.

📖 BÀI HỌC AI HÔM NAY CHO DEV

1. Tối ưu chi phí & hiệu năng LLM

2. Để tối ưu hóa chi phí và hiệu năng của Mô hình Ngôn ngữ Lớn (LLM), các nhà phát triển cần biết cách giảm thiểu tài nguyên tính toán và lưu trữ. Điều này giúp giảm chi phí vận hành và tăng tốc độ xử lý của ứng dụng.

3. Ví dụ, sử dụng kỹ thuật fine-tuning và LoRA (Low-Rank Adaptation) có thể giúp giảm kích thước mô hình và tăng tốc độ huấn luyện.

4. 💡 Tip: Sử dụng thư viện như Hugging Face Transformers để thực hiện fine-tuning và LoRA, và thử nghiệm với các kích thước mô hình khác nhau để tìm ra sự cân bằng giữa chi phí và hiệu năng.

💡 Luôn đi đầu trong thế giới AI! · Stay ahead in AI!

Nguồn: Google News · Groq AI