

Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

“Đốt đến đâu học lâu cũng biết.”

— Tục ngữ Việt Nam

Không ai sinh ra đã giỏi — sự kiên trì học hỏi có thể giúp bất kỳ ai đạt được tri thức và kỹ năng.

TIN TỨC NỔI BẬT

1

Claude Code đối đầu 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 đặt lên bàn cân trong một loạt các thử nghiệm để xác định hiệu suất và giá cả của chúng. Công cụ SWE-bench, một công cụ đánh giá hiệu suất cho phát triển phần mềm, đã được sử dụng để đánh giá hai trợ lý mã hóa được hỗ trợ bởi AI này. Kết quả cho thấy Claude Code vượt trội so với GitHub Copilot về độ chính xác và tốc độ, đặc biệt là trong các nhiệm vụ mã hóa phức tạp. Khi nói đến giá cả, Claude Code cung cấp một lựa chọn linh hoạt và tiết kiệm chi phí hơn. Mô hình giá của Claude Code dựa trên số lượng dòng mã được tạo, với mức giá cố định là \$0,01 cho mỗi dòng. Ngược lại, GitHub Copilot yêu cầu đăng ký dịch vụ cao cấp của GitHub, bắt đầu từ \$19,99 mỗi tháng. Điều này làm cho Claude Code trở thành một lựa chọn hấp dẫn hơn cho các developer cần tạo ra lượng mã lớn. Tổng thể, kết quả của so sánh này cho thấy Claude Code là một đối thủ mạnh trên thị trường trợ lý mã hóa được hỗ trợ bởi AI, cung cấp hiệu suất tốt hơn và giá cả cạnh tranh hơn.

2

Lớp lệnh AI Multi-Agent cho kho hàng: Đạt hiệu suất vận hành vượt trội và thông minh chuỗi cung ứng

Multi-Agent Warehouse AI Command Layer Enables Operational Excellence and Supply Chain Intelligence

NVIDIA Developer [Đọc bài viết →](#)

NVIDIA đã giới thiệu Lớp lệnh AI Nhà kho đa tác nhân, được thiết kế để nâng cao hiệu quả hoạt động và trí tuệ chuỗi cung ứng. Giải pháp sáng tạo này tận dụng trí tuệ nhân tạo (AI) để tối ưu hóa hoạt động nhà kho, cho phép các doanh nghiệp tối ưu hóa hậu cần và quản lý hàng tồn kho của họ. Lớp lệnh AI Nhà kho đa tác nhân được xây dựng trên nền tảng Omniverse của NVIDIA, cho phép trao đổi dữ liệu thời gian thực và hợp tác giữa các hệ thống khác nhau. Điều này cho phép các tác nhân AI làm việc cùng nhau một cách liền mạch, đưa ra quyết định thông minh để giảm thiểu độ trễ, giảm chi phí và cải thiện hiệu suất chuỗi cung ứng tổng thể. Các tính năng chính của Lớp lệnh AI Nhà kho đa tác nhân bao gồm: - Trao đổi dữ liệu thời gian thực và hợp tác giữa các hệ thống - Ra quyết định dựa trên AI để tối ưu hóa hậu cần và quản lý hàng tồn kho - Tăng cường trí tuệ chuỗi cung ứng và khả năng hiển thị - Cải thiện hiệu quả hoạt động và giảm chi phí Bằng cách triển khai giải pháp này, các doanh nghiệp có thể đạt được lợi thế cạnh tranh trên thị trường bằng cách cải thiện hoạt động chuỗi cung ứng của họ và đưa ra quyết định dựa trên dữ liệu. Lớp lệnh AI Nhà kho đa tác nhân có tiềm năng cách mạng hóa cách thức hoạt động của các nhà kho, cho phép họ phản ứng nhanh chóng với các điều kiện thị trường và nhu cầu khách hàng thay đổi.

3

Tối ưu workflow GitHub bằng generative AI với 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 Model Canvas Platform (MCP) để tối ưu hóa các quy trình làm việc trên GitHub bằng cách sử dụng AI tạo sinh. Tích hợp này cho phép các developer 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 và MCP, các developer hiện có thể sử dụng AI tạo sinh để tự động hóa các nhiệm vụ như tạo mã, tài liệu và kiểm thử. Điều này có thể giúp giảm thời gian và công sức cần thiết để hoàn thành các nhiệm vụ này, cho phép các developer tập trung vào các nhiệm vụ phức tạp và có giá trị cao hơn. Tích hợp này cũng cho phép các developer sử dụng MCP để tạo và quản lý các model có thể được sử dụng để tạo mã và tài liệu được hỗ trợ bởi AI. Điều này có thể giúp cải thiện chất lượng mã, giảm lỗi và tăng năng suất. Quan hệ đối tác giữa Amazon Bedrock và MCP nhằm mục đích giúp các developer để

dàng áp dụng AI tạo sinh vào các quy trình làm việc trên GitHub của họ, cho phép họ xây dựng các quy trình phát triển phần mềm hiệu quả và hiệu quả hơn.

4

Từ các dự án open source đình đám đến OpenAI

From open source hits to OpenAI

Changelog [Đọc bài viết →](#)

Trong tập này, Max Stoiber, một developer làm việc trên thư mục plugin và nền tảng ứng dụng của ChatGPT tại OpenAI, chia sẻ những hiểu biết của mình về ngành công nghệ. Ông thảo luận về các dự án mã nguồn mở ít được biết đến đã có tác động đáng kể, chẳng hạn như react-boilerplate và styled-components. Stoiber cũng đề cập đến sự phát triển của GitHub, đặc biệt là vai trò của Spectrum trong việc định hình GitHub Discussions. Ngoài ra, ông nói về sự thành công của Stellate, một bộ nhớ đệm GraphQL được Shopify và The Guild mua lại. Hơn nữa, Stoiber giải thích tại sao các ứng dụng ChatGPT cảm giác như một bề mặt mới cho phát triển phần mềm. Tập này được tài trợ bởi Coder.com, WorkOS, Notion, Fly.io và Changelog++.

5

sqlite-utils 4.0rc2, chủ yếu do Claude Fable viết (với chi phí khoảng 149.25 USD)

sqlite-utils 4.0rc2, mostly written by Claude Fable (for about \$149.25)

Simon Willison [Đọc bài viết →](#)

Tôi là biên tập viên tin tức công nghệ. Một bản cập nhật gần đây cho sqlite-utils, một thư viện quản lý cơ sở dữ liệu phổ biến, đã được phát hành dưới dạng phiên bản 4.0rc2. Bản cập nhật này được thực hiện nhờ Claude Fable, một tác nhân lập trình, người đã làm việc trên dự án với khoảng \$149,25. Phiên bản mới này giải quyết một số vấn đề quan trọng, bao gồm một lỗi nghiêm trọng có thể dẫn đến mất dữ liệu. Lỗi này được Fable xác định và sửa thông qua một loạt 37 yêu cầu và 34 commit. Bản cập nhật cũng bao gồm tài liệu toàn diện về xử lý giao dịch, một tính năng mới trong thư viện. Tài liệu này giải thích cách giao dịch hoạt động trong sqlite-utils và cung cấp hướng dẫn về khi nào sử dụng db.atomic() và các phương thức liên quan đến giao dịch khác. Ngoài ra, bản cập nhật đảm bảo rằng thư viện hoạt động chính xác với các kết nối Python 3.12+, có các cài đặt autocommit khác

nhau. Việc xem xét cuối cùng được thực hiện thông qua giao diện PR của GitHub và bản cập nhật hiện đã có sẵn để thử nghiệm.

6

Thiên tài ẩn mình của Jim Henson: Khám phá kiệt tác The Cube, tiền thân của Black Mirror

The Unseen Genius of Jim Henson: Unpacking the Proto-Black Mirror Masterpiece, The Cube

Dev.to AI [Đọc bài viết →](#)

Jim Henson, người tạo ra nổi tiếng của The Muppets và The Dark Crystal, có một viên ngọc ít được biết đến trong tác phẩm của mình - The Cube, một vở kịch truyền hình năm 1969 xuất hiện trước những tác phẩm như Black Mirror hàng thập kỷ. Sản xuất cho Experiment in Television của NBC, The Cube là một cuộc khám phá 30 phút về tình trạng con người, đi sâu vào các chủ đề về sự cô lập, hoảng loạn và sự□□ của thực tại. Câu chuyện theo dõi một nhóm người bị mắc kẹt trong một khối lập phương bí ẩn, luôn thay đổi với không có lối thoát hoặc giải thích rõ ràng. The Cube là một lớp học thạc sĩ về sự tinh tế, sử dụng khối lập phương như một ẩn dụ cho sự phức tạp của sự tồn tại con người. Vở kịch truyền hình này□□ khả năng của Henson trong việc tạo ra những câu chuyện phức tạp, sâu sắc và phù hợp với khán giả, và các chủ đề của nó hiện diện trong các tác phẩm sau này của ông, mặc dù ở những hình thức tinh tế hơn. The Cube là một viên ngọc ẩn trong tác phẩm của Henson, cung cấp một cái nhìn về phía tối, sâu sắc hơn của ông, và là một tác phẩm phải xem cho người hâm mộ công việc của ông và những người quan tâm đến sự tiến hóa của khoa học viễn tưởng và cách kể chuyện thử nghiệm.

7

Thiết bị hồi sinh nhãn cầu từ người hiến đã mất có thể giúp cấy ghép mắt thành hiện thực

A device that revives eyeballs from dead donors could make eye transplants possible

MIT Tech Review [Đọc bài viết →](#)

Các nhà khoa học đã có một bước tiến đột phá trong cấy ghép mắt với sự phát triển của một thiết bị gọi là Eye-in-a-Care-Box (ECaBox). Thiết bị này sử dụng kỹ thuật perfusion, một phương pháp cung cấp oxy và chất dinh dưỡng cho các cơ quan, để duy trì và hồi sinh các quả cầu mắt vừa được loại bỏ. Các nhà nghiên cứu do Pia Cosma dẫn đầu tại Trung tâm Quản lý Genomic ở Tây Ban Nha đã thử nghiệm thiết bị trên mắt lợn và mắt người từ các nhà tài trợ đã qua đời. Kết quả cho

thấy mắt được điều trị bằng ECaBox vẫn giữ khả năng truyền tín hiệu điện và có khả năng nhìn, thậm chí sau 24 giờ. Ngược lại, mắt không được điều trị nhanh chóng suy giảm và mất cấu trúc. Thiết bị có một cửa sổ trong để các nhà nghiên cứu nghiên cứu và hình ảnh hóa mắt, và nó duy trì nhiệt độ và áp suất cụ thể để bảo tồn cơ quan. Đội ngũ hy vọng rằng ECaBox có thể một ngày nào đó biến cấy ghép toàn bộ mắt thành một khả năng có thể, cho phép mọi người nhìn lại sau khi nhận được một mắt hiến tặng. Mặc dù thiết bị đã cho thấy kết quả đầy hứa hẹn, nhưng vẫn cần thêm thử nghiệm và cải tiến trước khi nó có thể được sử dụng trong cấy ghép cho người.

8

Mức tiêu thụ điện năng biến động của AI âm thầm thử thách giới hạn lưới điện

AI's Volatile Power Use Quietly Tests Grid Limits

IEEE Spectrum [Đọc bài viết →](#)

Sự phát triển nhanh chóng của cơ sở hạ tầng trí tuệ nhân tạo (AI) đang gây áp lực lên lưới điện, không chỉ do nhu cầu năng lượng tăng cao, mà còn do tính chất khó dự đoán và đồng bộ của các khối lượng công việc tính toán AI. Các trung tâm dữ liệu, nơi đặt cơ sở hạ tầng AI, dự kiến sẽ tiêu thụ 3-4% nhu cầu điện toàn cầu trong thập kỷ này. Vấn đề không chỉ là về lượng điện tiêu thụ, mà còn là về tốc độ và vị trí nhu cầu thay đổi, tạo ra những thách thức vận hành mới cho các nhà điều hành lưới điện. Không giống như nhu cầu công nghiệp truyền thống, các khối lượng công việc AI có thể tăng và giảm nhanh chóng, tạo ra những thay đổi đáng kể về tiêu thụ điện trong khoảng thời gian ngắn. Sự biến động này có thể gây áp lực lên các nguồn dự phòng, cơ chế kiểm soát tần số và cơ sở hạ tầng truyền tải địa phương. Do đó, các nhà điều hành trung tâm dữ liệu đang triển khai các công nghệ giảm thiểu, và các tổ chức nghiên cứu đang nhấn mạnh sự phức tạp ngày càng tăng của việc tích hợp các nguồn lực động vào hoạt động lưới điện hiện đại.

TIPS & TRICKS CHO DEV

Sử dụng Ollama

Vấn đề: Cần chạy LLM trên máy cá nhân mà không cần kết nối internet.

Cách làm: Cài đặt Ollama, chọn mô hình LLM phù hợp, và nhập prompt như "Tôi cần thông tin về chủ đề AI".

Đánh giá: Hiệu quả cho các dự án cần bảo mật dữ liệu, không nên dùng cho mô hình quá lớn.

Tối ưu hóa LM Studio

Vấn đề: LM Studio chạy chậm trên máy cá nhân.

Cách làm: Điều chỉnh cấu hình LM Studio, giảm kích thước mô hình, và nhập lệnh `lmstudio --optimize` để tối ưu hóa.

Đánh giá: Hiệu quả cho máy cá nhân cấu hình thấp, không nên dùng cho dự án cần độ chính xác cao.

Chạy LLM Offline

Vấn đề: Cần chạy LLM trên máy cá nhân mà không có kết nối internet.

Cách làm: Sử dụng LM Studio, nhập lệnh `lmstudio --offline` và nhập prompt như "Tôi cần thông tin về chủ đề machine learning".

Đánh giá: Hiệu quả cho dự án cần bảo mật và không cần cập nhật dữ liệu mới.

BÀI HỌC AI HÔM NAY CHO DEV

1. Tối ưu chi phí & hiệu năng LLM

2. Để xây dựng ứng dụng AI hiệu quả, các developer cần biết cách tối ưu chi phí và hiệu năng của mô hình ngôn ngữ lớn (LLM). Điều này giúp giảm thiểu chi phí tính toán và tăng tốc độ xử lý, từ đó cải thiện trải nghiệm người dùng. Việc tối ưu hóa LLM cũng giúp giảm thiểu tác động môi trường của các ứng dụng AI.

3. Ví dụ, bằng cách sử dụng kỹ thuật fine-tuning và LoRA, các developer có thể giảm kích thước mô hình và tăng tốc độ đào tạo, từ đó giảm chi phí và tăng hiệu năng.

4. Tip hoặc bước tiếp theo: Các developer nên khám phá các kỹ thuật tối ưu hóa LLM như quantization, pruning và knowledge distillation để giảm chi phí và tăng hiệu năng cho ứng dụng AI của mình.

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

Nguồn: Google News · Groq AI