BIS - MT · 01/06/2026

🤖 Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

✨ *“Dream big, start small, but most of all, start.”*

↪ Mơ lớn, bắt đầu nhỏ, nhưng quan trọng nhất, hãy bắt đầu.

— Simon Sinek

💡 *Ước mơ lớn cần hành động nhỏ đầu tiên — sự hoàn hảo của kế hoạch không quan trọng bằng việc thực sự bắt tay vào làm.*

📰 TIN TỨC NỔI BẬT

1

AWS mở mã nguồn một máy chủ MCP cho Bedrock AgentCore để đơn giản hóa phát triển AI agent

🇬🇧 *AWS Open-Sources an MCP Server for Bedrock AgentCore to Streamline AI Agent Development*

📰 MarkTechPost 🔗 Đọc bài viết →

> Amazon Web Services (AWS) đã mở nguồn một máy chủ MCP (Nền tảng Đa đám mây) cho Bedrock AgentCore, một framework để phát triển các tác nhân AI. Động thái này nhằm mục đích đơn giản hóa việc phát triển tác nhân AI bằng cách cung cấp một nền tảng tiêu chuẩn hóa cho việc thử nghiệm và triển khai trên nhiều môi trường đám mây. Máy chủ MCP được thiết kế để hoạt động liền mạch với Bedrock AgentCore, cho phép các nhà phát triển tạo, thử nghiệm và triển khai các tác nhân AI một cách dễ dàng. Bằng cách mở nguồn máy chủ MCP, AWS đang cho phép cộng đồng nhà phát triển rộng lớn hơn đóng góp và hưởng lợi từ nền tảng này. Framework Bedrock AgentCore là một thành phần quan trọng trong nỗ lực của AWS nhằm đơn giản hóa việc phát triển AI. Bằng cách cung cấp một nền tảng tiêu chuẩn hóa cho các tác nhân AI, AWS đang giúp các nhà phát triển xây dựng, thử nghiệm và triển khai các ứng dụng được hỗ trợ bởi AI trên nhiều môi trường đám mây. Việc mở nguồn máy chủ MCP dự kiến sẽ đẩy nhanh việc phát triển các tác nhân AI và thúc đẩy sự hợp tác giữa các nhà phát triển. Động thái này phù hợp với mục tiêu của AWS là làm cho việc phát triển AI trở nên dễ tiếp cận và hiệu quả hơn cho khách hàng của mình.

2

Xây dựng một agent phân tích tài chính thông minh với LangGraph và Strands Agents | Amazon Web Services

🇬🇧 *Build an intelligent financial analysis agent with LangGraph and Strands Agents | Amazon Web Services*

📰 Amazon Web Services (AWS) 🔗 Đọc bài viết →

> Amazon Web Services (AWS) đã giới thiệu một giải pháp mới để xây dựng một tác nhân phân tích tài chính thông minh sử dụng LangGraph và Strands Agents. Cách tiếp cận sáng tạo này cho phép người dùng tạo ra một hệ thống phân tích tài chính tinh vi có thể xử lý và phân tích lượng lớn dữ liệu tài chính. LangGraph là một thư viện xử lý ngôn ngữ tự nhiên (NLP) cho phép người dùng trích xuất thông tin chi tiết từ các tài liệu tài chính, chẳng hạn như báo cáo và tuyên bố tài chính. Strands Agents, mặt khác, là một nền tảng low-code cho phép người dùng xây dựng, triển khai và quản lý các tác nhân được hỗ trợ bởi AI. Bằng cách kết hợp LangGraph và Strands Agents, người dùng có thể tạo ra một tác nhân phân tích tài chính có thể tự động trích xuất thông tin liên quan từ các tài liệu tài chính, phân tích nó và cung cấp thông tin chi tiết có thể hành động. Tác nhân này có thể được tích hợp với các dịch vụ AWS khác, chẳng hạn như Amazon SageMaker và Amazon Comprehend, để tăng cường khả năng của nó. Giải pháp này được thiết kế để giúp các tổ chức tài chính, kế toán và nhà phân tích tối ưu hóa quy trình phân tích tài chính của họ, cải thiện độ chính xác và giảm công việc thủ công. Với LangGraph và Strands Agents, người dùng có thể xây dựng một hệ thống phân tích tài chính thông minh và hiệu quả hơn có thể theo kịp sự phức tạp của tài chính hiện đại.

3

Các model trọng lượng mở 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 thảo luận về các mô hình trọng lượng mở hàng đầu của Trung Quốc và các đối thủ mạnh nhất của Mỹ. Các mô hình trọng lượng mở đề cập đến xe điện không có giới hạn trọng lượng, cho phép nhiều tùy chọn tùy chỉnh rộng rãi. Các nhà sản xuất Trung Quốc đã đạt được những bước tiến đáng kể trong lĩnh vực này, với một số mô hình nổi bật về hiệu suất và tính năng. Một số mô hình trọng lượng mở hàng đầu của Trung Quốc được đề cập trong bài viết bao gồm XPeng P7, BYD Song và Geely Geometry A. Những xe này cung cấp phạm vi ấn tượng, tốc độ và các tính năng công nghệ tiên tiến, khiến chúng trở

thành những đối thủ mạnh trên thị trường. Bài viết cũng nhấn mạnh các đối thủ mạnh nhất của Mỹ đối với các mô hình Trung Quốc này. Tuy nhiên, các mô hình cụ thể của Mỹ không được cung cấp trong nội dung cho trước. Bài viết gợi ý rằng các nhà sản xuất Trung Quốc đang tạo ra sự cạnh tranh gay gắt cho các công ty Mỹ trong thị trường xe điện trọng lượng mở, với các thiết kế sáng tạo và giá cả cạnh tranh.

4

datasette 1.0a32

🇬🇧 *datasette 1.0a32*

📰 Simon Willison 🔗 Đọc bài viết →

Datasette đã phát hành phiên bản 1.0a32, một bản cập nhật nhỏ tập trung vào việc sửa lỗi. Phiên bản này giải quyết hai vấn đề cụ thể: một vấn đề với các truy vấn INSERT ... RETURNING qua điểm cuối /db/-/execute-write và các vấn đề base_url khác phát sinh trong quá trình thử nghiệm với Service Workers. Những sửa lỗi này nhằm cải thiện sự ổn định và chức năng của Datasette.

5

Các trang web có một cách mới để theo dõi khách truy cập: Phân tích hoạt động SSD của họ

🇬🇧 *Websites have a new way to spy on visitors: Analyzing their SSD activity*

📰 Ars Technica 🔗 Đọc bài viết →

Các trang web đã phát hiện ra một cách mới để theo dõi bí mật thói quen duyệt web và sử dụng thiết bị của khách truy cập. Một kỹ thuật gọi là FROST (lấy dấu vân tay từ xa sử dụng thời gian SSD dựa trên OPFS) cho phép các trang web theo dõi các trang web khác mà khách truy cập đang xem và các ứng dụng nào đang mở trên thiết bị của họ. Phương pháp này khai thác một kênh phụ, đo lường sự tương tác giữa các quá trình cạnh tranh cho một tài nguyên, chẳng hạn như ổ đĩa trạng thái rắn (SSD). Bằng cách phân tích thời gian của các hoạt động SSD nhất định, kẻ tấn công có thể suy luận dữ liệu bí mật, bao gồm cả trang web và ứng dụng đang mở. FROST không cần sự tương tác từ khách truy cập và có thể được thực hiện chỉ trong trình duyệt sử dụng JavaScript. Các nhà nghiên cứu đã chứng minh tính hiệu quả của FROST trong việc theo dõi hoạt động của người dùng, nhưng kỹ thuật này có những hạn chế, bao gồm cả nhu cầu về một tệp OPFS lớn và yêu cầu tệp đó phải được lưu trữ trên cùng một SSD với thiết bị của khách truy cập. Để ngăn chặn các cuộc tấn công FROST, người dùng có thể đóng các tab không cần thiết và theo dõi việc tạo và kích thước của các tệp OPFS được phân bổ bởi các trang web không xác định. Các nhà sản xuất trình duyệt cũng đang tìm cách giảm thiểu lỗ hổng này.

6

Erin Brockovich nhắm vào bí mật của trung tâm dữ liệu

🇬🇧 *Erin Brockovich takes aim at data center secrecy*

📰 TechCrunch AI 🔗 Đọc bài viết →

Nhà hoạt động môi trường Erin Brockovich đã khởi xướng một nhiệm vụ mới để mang lại minh bạch cho việc xây dựng trung tâm dữ liệu và tác động của nó đối với các cộng đồng lân cận. Brockovich, được biết đến với vụ kiện cao cấp chống lại Pacific Gas & Electric, đã tạo một trang web có tính năng bản đồ các trung tâm dữ liệu trên toàn nước Mỹ, được đóng góp từ các thành viên cộng đồng. Sau khi nhận được gần 4.000 bản gửi trong tháng đầu tiên, Brockovich đã nhấn mạnh mối quan ngại chung về sự thiếu minh bạch trong các dự án trung tâm dữ liệu. Cô ấy nhấn mạnh rằng mục tiêu của mình không phải là phản đối trung tâm dữ liệu hoặc AI, mà là giải quyết mô hình bí mật bao quanh sự phát triển trung tâm dữ liệu, bao gồm cả các dự án được công bố sau khi giấy phép được bảo đảm và các quan chức địa phương ký thỏa thuận không tiết lộ mà không có kiến thức của cộng đồng. Sáng kiến của Brockovich nhằm mục đích làm sáng tỏ các vấn đề ảnh hưởng đến các cộng đồng lân cận và thúc đẩy giao tiếp cởi mở hơn trong xây dựng trung tâm dữ liệu.

7

Mọi người đều đặt mục tiêu vào MacBook Neo

🇬🇧 *Everyone Has Their Targets Set on the MacBook Neo*

📰 Wired 🔗 Đọc bài viết →

Tác động của MacBook Neo đối với thị trường laptop vẫn còn được cảm nhận, với các nhà sản xuất Windows phản ứng lại mức giá 599 đô la của nó. Dell XPS 13 mới và Microsoft Surface Laptop 8 là những sản phẩm mới nhất, cả hai đều nhằm mục đích khớp với vẻ ngoài và cảm giác cao cấp của Neo trong khi cung cấp giá cả cạnh tranh. Dell XPS 13 được làm hoàn toàn bằng nhôm, có màn hình IPS cao cấp và có tốc độ làm mới nhanh hơn Neo. Tuy nhiên, nó bắt đầu với 8 GB RAM và bộ xử lý Intel Core 5 chậm hơn, tương tự như Neo. XPS 13 có thể được cấu hình lên đến 32 GB RAM và 1 TB lưu trữ, cung cấp nhiều tính linh hoạt hơn so với Neo. Mặt khác, Microsoft Surface Laptop 8 bắt đầu với 16 GB RAM trong mẫu cao cấp hơn nhưng chỉ 8 GB trong phiên bản 13 inch giá rẻ hơn, dự kiến ra mắt vào cuối năm nay. Dell XPS 13 mới được định vị là một đối thủ cạnh tranh trực tiếp hơn với MacBook Neo, cung cấp trải nghiệm cao cấp với mức giá thấp hơn.

8

Cách Composer 2.5 của Cursor sử dụng tự chưng cất để vượt qua các LLM tiền phong trong mã hóa

🇬🇧 *How Cursor's Composer 2.5 uses self-distillation to beat the frontier LLMs at coding*

📰 BD Tech Talks 🔗 Đọc bài viết →

> Composer 2.5 của Cursor đã nổi lên như một mô hình mã hóa hàng đầu, vượt trội so với các mô hình lớn hơn, đắt tiền hơn như GPT-5.5 của OpenAI và Opus 4.7 của Anthropic trong các nhiệm vụ kỹ thuật hàng ngày. Điều này được thực hiện thông qua một kỹ thuật gọi là tự chưng cất, cho phép một mô hình nhỏ hơn bắt chước đầu ra của một mô hình lớn hơn. Composer 2.5 sử dụng một quá trình gọi là "học tăng cường có mục tiêu với phản hồi văn bản" để cải thiện hiệu suất của nó. Điều này liên quan đến việc tiêm các gợi ý văn bản cục bộ vào ngữ cảnh của mô hình khi nó mắc lỗi, cho phép nó điều chỉnh hành vi và sửa lỗi của mình mà không làm giảm khả năng tổng thể của nó. Hiệu quả thuật toán của mô hình làm cho nó có thể tài chính cho các nhóm nhỏ, với chi phí 0,50 đô la cho mỗi triệu token đầu vào và 2,50 đô la cho mỗi triệu token đầu ra. Mặc dù Composer 2.5 không hoàn hảo và có thể gặp khó khăn với các nhiệm vụ phức tạp, nhưng nó chứng minh rằng các mô hình chuyên dụng có thể cạnh tranh với các mô hình lớn hơn mà không cần nhiều tham số hơn. Cách tiếp cận này có khả năng dân chủ hóa mã hóa agentic mạnh mẽ và làm cho nó dễ tiếp cận hơn với các nhà phát triển.

⚡ TIPS & TRICKS CHO DEV

⚡ Tích hợp File System

Vấn đề: Việc kết nối AI với file system giúp tăng cường khả năng đọc và ghi dữ liệu.

Cách làm: Sử dụng MCP để kết nối Claude Desktop với file system, cho phép đọc và ghi file trực tiếp. Ví dụ, câu prompt "Tạo file mới tên là example.txt" sẽ tạo file mới trong thư mục hiện tại.

Đánh giá: Hiệu quả trong việc tự động hóa các tác vụ liên quan đến file, nhưng cần đảm bảo an toàn dữ liệu.

⚡ Kết nối Database

Vấn đề: Việc kết nối AI với database giúp tăng cường khả năng truy cập và xử lý dữ liệu.
Cách làm: Sử dụng MCP để kết nối Claude Code với database, cho phép thực hiện các truy vấn SQL trực tiếp. Ví dụ, lệnh "SELECT * FROM users" sẽ trả về tất cả dữ liệu từ bảng users.
Đánh giá: Hiệu quả trong việc tăng cường khả năng truy cập dữ liệu, nhưng cần đảm bảo an toàn và bảo mật dữ liệu.

⚡ Tích hợp API

Vấn đề: Việc kết nối AI với API giúp tăng cường khả năng tương tác với các dịch vụ ngoài.
Cách làm: Sử dụng MCP để kết nối Cursor với API, cho phép thực hiện các yêu cầu API trực tiếp. Ví dụ, câu prompt "Gửi yêu cầu GET đến https://api.example.com/data" sẽ trả về dữ liệu từ API.
Đánh giá: Hiệu quả trong việc tăng cường khả năng tương tác với các dịch vụ ngoài, nhưng cần đảm bảo an toàn và bảo mật dữ liệu.

📚 BÀI HỌC AI HÔM NAY CHO DEV

1. **Tích hợp AI API vào ứng dụng**
2. Trong thời đại hiện nay, tích hợp AI vào ứng dụng là một bước quan trọng để tăng cường khả năng và hiệu suất của ứng dụng. Các dev cần biết cách tích hợp AI API để tạo ra các ứng dụng thông minh hơn. Việc này giúp cải thiện trải nghiệm người dùng và tăng cường cạnh tranh trên thị trường.
3. Ví dụ, chúng ta có thể sử dụng API như Google Cloud Vision để phân tích hình ảnh trong ứng dụng di động, hoặc sử dụng API như Dialogflow để tạo ra các chatbot thông minh.
4. 💡 Tip hoặc bước tiếp theo: Hãy bắt đầu bằng cách lựa chọn một AI API phù hợp với nhu cầu của ứng dụng, sau đó tích hợp nó vào mã nguồn của bạn bằng cách sử dụng các thư viện và framework hỗ trợ như RESTful API hoặc gRPC.