BIS - MT · 21/06/2026

Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

"How wonderful it is that nobody need wait a single moment before starting to improve the world."

↪ Thật kỳ diệu khi không ai cần phải chờ một giây nào trước khi bắt đầu cải thiện thế giới.

— Anne Frank

Sự thay đổi tích cực không cần điều kiện hoàn hảo hay nguồn lực lớn — mỗi người đều có thể bắt đầu tạo ra tác động ngay hôm nay.

TIN TỨC NỔI BẬT

1

Tăng tốc phát triển với server Amazon Bedrock AgentCore MCP | Trí tuệ nhân tạo

Accelerate development with the Amazon Bedrock AgentCore MCP server | Artificial Intelligence

Amazon Web Services (AWS) Đọc bài viết →

> Amazon Web Services (AWS) đã giới thiệu máy chủ Amazon Bedrock AgentCore MCP, được thiết kế để tăng tốc phát triển trong lĩnh vực Trí tuệ Nhân tạo (AI). Máy chủ Bedrock AgentCore MCP là một phần của nền tảng Bedrock, cung cấp nền tảng cho việc xây dựng, triển khai và quản lý các model AI. Máy chủ AgentCore MCP là một dịch vụ được quản lý cho phép các nhà phát triển xây dựng, đào tạo và triển khai các model AI với hiệu suất và khả năng mở rộng tăng cao. Nó cung cấp một môi trường an toàn và đáng tin cậy cho việc phát triển model, cho phép các đội ngũ cộng tác và lặp lại các dự án AI hiệu quả hơn. Lợi ích chính của máy chủ Amazon Bedrock AgentCore MCP bao gồm tốc độ phát triển model được cải thiện, độ chính xác của model tăng cao và khả năng cộng tác được nâng cao. Bằng cách tận dụng máy chủ Bedrock AgentCore MCP, các nhà phát triển có thể tối ưu hóa các quy trình phát triển AI, giảm chi phí và tăng tốc thời gian đưa ứng dụng được hỗ trợ bởi AI ra thị trường. Dịch vụ này là một phần của nỗ lực liên tục của AWS để hỗ trợ phát triển và triển khai các model AI trong các ngành công nghiệp khác nhau.

2 Agent Factory: Từ prototype đến production—công cụ developer và phát triển agent nhanh chóng

Agent Factory: From prototype to production—developer tools and rapid agent development

Microsoft Azure Đọc bài viết →

Microsoft đã giới thiệu Agent Factory, một nền tảng được thiết kế để tối ưu hóa quá trình phát triển các tác nhân thông minh. Agent Factory nhằm mục đích bắc cầu khoảng cách giữa nguyên mẫu và sản xuất, cho phép các nhà phát triển tạo và triển khai các tác nhân một cách hiệu quả hơn. Nền tảng này cung cấp một loạt các công cụ cho nhà phát triển, bao gồm giao diện trực quan và thư viện các thành phần đã được xây dựng sẵn, để thúc đẩy sự phát triển tác nhân nhanh chóng. Với Agent Factory, các nhà phát triển có thể thiết kế, xây dựng và triển khai các tác nhân có thể tương tác với các hệ thống và ứng dụng khác nhau. Các công cụ và thành phần của nền tảng được thiết kế để có tính mô-đun và có thể mở rộng, cho phép các nhà phát triển tùy chỉnh và tích hợp các tác nhân với cơ sở hạ tầng hiện có của họ. Agent Factory được xây dựng trên Microsoft Azure, tận dụng khả năng mở rộng và độ tin cậy của nền tảng dựa trên đám mây này. Sự tích hợp này cho phép các nhà phát triển triển khai các tác nhân với quy mô lớn, đồng thời cũng được hưởng lợi từ các tính năng bảo mật và giám sát của Azure. Bằng cách cung cấp một bộ công cụ toàn diện và cơ sở hạ tầng có thể mở rộng, Agent Factory nhằm mục đích đẩy nhanh quá trình phát triển và triển khai các tác nhân thông minh, cho phép các tổ chức mở khóa các cơ hội kinh doanh mới và cải thiện hiệu quả hoạt động.

3 Tối ưu hóa workflow GitHub với generative AI 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 công bố một tích hợp mới giữa Amazon Bedrock và 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 và MCP, các nhà phát triển có thể tạo và quản lý các mô hình AI quy mô lớn, sau đó có thể được sử dụng để tự

động hóa các nhiệm vụ và cải thiện hiệu suất trong các quy trình làm việc trên GitHub. Điều này bao gồm các nhiệm vụ như tạo mã, xem xét và kiểm tra, cho phép các nhà phát triển tập trung vào các nhiệm vụ cấp cao hơn và cải thiện năng suất tổng thể. Tích hợp giữa Amazon Bedrock và MCP cung cấp một trải nghiệm liền mạch cho các nhà phát triển để tạo, đào tạo và triển khai các mô hình AI trực tiếp từ GitHub. Điều này cho phép các nhà phát triển tận dụng toàn bộ tiềm năng của AI tạo sinh để tối ưu hóa các quy trình làm việc của họ và cải thiện quá trình phát triển tổng thể. Bằng cách tận dụng tích hợp này, các nhà phát triển có thể đẩy nhanh chu kỳ phát triển của họ và cung cấp phần mềm chất lượng cao một cách hiệu quả hơn.

4

Microsoft phát hiện backdoor nhẹ mới chuyên đánh cắp cryptocurrency

Microsoft discovers new lightweight backdoor that steals cryptocurrency

Ars Technica Đọc bài viết →

Microsoft đã phát hiện ra một loại malware mới được gọi là Crypto Clipper, lan truyền qua các ổ USB và nhắm vào thông tin đăng nhập tiền điện tử. Malware này theo dõi nội dung của bộ nhớ tạm thiết bị để tìm kiếm các mẫu phù hợp với địa chỉ ví hoặc cụm từ hạt giống, và khi tìm thấy, nó gửi thông tin đăng nhập đến các máy chủ được kiểm soát bởi kẻ tấn công thông qua mạng Tor. Để che giấu sự hiện diện của mình, Crypto Clipper thiết lập một kết nối Tor bằng cách sử dụng SOCKS5 proxy và trộn việc đánh cắp dữ liệu với thực thi mã từ xa. Malware này cũng chụp năm ảnh chụp màn hình trong khoảng thời gian 10 giây và gửi chúng đến kẻ tấn công. Microsoft quan sát thấy Crypto Clipper lan truyền qua các tệp .lnk trên ổ USB, lưu trữ mã thực thi. Công ty tin rằng mục đích của các ảnh chụp màn hình là cung cấp ngữ cảnh có thể hữu ích cho kẻ tấn công. Microsoft Defender for Endpoint và Antivirus phát hiện các thành phần của Crypto Clipper, và dấu hiệu của nhiễm malware bao gồm trình thông dịch kịch bản tạo ra các quá trình con đáng ngờ, sử dụng proxy trên localhost:9050, và dấu hiệu của việc kiểm tra bộ nhớ tạm hoặc thay thế địa chỉ tiền điện tử.

5

Máy in 3D Flashforge Adventurer 5M đạt mức giá thấp kỷ lục $133

Flashforge Adventurer 5M 3D Printer Hits All-Time Low Price of $133

Dev.to AI Đọc bài viết →

Máy in 3D Flashforge Adventurer 5M đã đạt mức giá thấp nhất từ trước đến nay là 133 đô la. Máy in 3D giá cả phải chăng này được biết đến với độ tin cậy, khả năng tùy chỉnh và dễ sử dụng, khiến nó trở thành một lựa chọn tuyệt vời cho cả người mới bắt đầu và người dùng có kinh nghiệm. Thiết bị này có màn hình LCD đơn sắc 5 inch lớn, cho phép người dùng dễ dàng điều hướng và điều chỉnh cài đặt. Thiết kế khung mở và cấu trúc mô-đun cung cấp sự tự do sáng tạo, vì nó tương thích với nhiều loại filament khác nhau. Máy in 3D Flashforge Adventurer 5M phù hợp với những người muốn bắt đầu với in 3D, cũng như các chuyên gia đang tìm kiếm một máy in đáng tin cậy và chất lượng cao. Với các tính năng tiên tiến và giao diện thân thiện với người dùng, máy in này là một lựa chọn tuyệt vời cho bất kỳ ai muốn khám phá thế giới của in 3D.

6

MCP trên Code Mode

MCP on Code Mode

Changelog Đọc bài viết →

Matt Carey của Cloudflare thảo luận về Code Mode và mối quan hệ của nó với MCP (Model-Code-Protocol) trong một tập gần đây. Carey tiết lộ rằng nhiều người đã hiểu lầm MCP, và ông giải thích cách Code Mode phía máy chủ cho phép một máy chủ MCP duy nhất lộ hơn 2.500 điểm cuối API của Cloudflare bằng cách sử dụng khoảng 1.000 token ngữ cảnh. Ông cũng nói về trình tải Worker động, điều này an toàn chạy mã được viết bởi model trong một V8 isolate. Ngoài ra, Carey chia sẻ quy trình làm việc cá nhân của mình với Claude, một model, và thảo luận về vai trò của bộ nhớ trong tương lai của các agent. Ông cũng đề cập đến việc sử dụng một công cụ bao git gọi là Zaggy để ngăn chặn force-push đến các kho lưu trữ của mình. Tập này được tài trợ bởi một số công ty, bao gồm Coder, Tailscale, RWX và Fly.io.

7

Trích lời Sean Lynch

Quoting Sean Lynch

Simon Willison Đọc bài viết →

Theo Sean Lynch, một lợi thế chính của MCP nằm ở khả năng cách ly luồng xác thực ngoài cửa sổ ngữ cảnh của tác nhân. Điều này cũng có

thể loại bỏ quá trình xác thực khỏi phần khung đỡ hoàn toàn. Lynch đề xuất rằng hình thức lý tưởng của MCP có thể là một cổng xác thực đơn giản cho API, điều này vẫn sẽ là một cải tiến đáng kể.

8

Chiến tranh đã dạy nữ doanh nhân Ukraine này giá trị của sự kiên cường

War Taught this Ukrainian Entrepreneur the Value of Resilience

IEEE Spectrum Đọc bài viết →

Doanh nhân người Ukraine Salome Mikadze-Struk đã dựa trên kinh nghiệm của mình để tư vấn cho các công ty khởi nghiệp về tầm quan trọng của sự kiên cường trong lĩnh vực kinh doanh. Sinh ra trong một gia đình tị nạn, Mikadze-Struk đã xây dựng một doanh nghiệp phát triển phần mềm, Movadex, trong suốt đại dịch COVID-19 và cuộc xâm lược của Nga vào Ukraine. Mặc dù đối mặt với nhiều thách thức, cô đã quản lý để giữ cho công ty của mình hoạt động và tốt nghiệp từ Đại học Georgetown vào năm 2022. Sau khi hoàn thành chương trình MBA tại Đại học Stanford vào năm 2023, Mikadze-Struk đã bắt đầu tư vấn cho các công ty khởi nghiệp và phát biểu về nhu cầu kiên cường trong bối cảnh gián đoạn và bất ổn. Quan điểm độc đáo của cô đặc biệt phù hợp trong bối cảnh các công cụ mã hóa AI đang làm thay đổi các mô hình kinh doanh truyền thống trong ngành công nghiệp phần mềm. Mikadze-Struk nhấn mạnh rằng các doanh nhân phải sẵn sàng chấp nhận rủi ro, thích nghi với thay đổi và điều hướng sự bất ổn để thành công. Kinh nghiệm của chính cô, bao gồm cả việc lớn lên trong một gia đình đã chạy trốn khỏi cuộc xung đột ở Georgia và xây dựng một doanh nghiệp ở Ukraine đang bị chiến tranh tàn phá, đã định hình quan điểm của cô về tầm quan trọng của sự kiên cường trong lĩnh vực kinh doanh.

TIPS & TRICKS CHO DEV

Tích hợp MCP với Git

Vấn đề: Cần kết nối AI với Git để tự động hóa quản lý mã nguồn.

Cách làm: Sử dụng MCP với Claude Code, nhập lệnh `git clone` và quản lý repository.

Đánh giá: Hiệu quả trong quản lý mã nguồn, nên dùng khi cần tự động hóa.

Sử dụng MCP với Database

Vấn đề: Cần kết nối AI với database để truy xuất dữ liệu.

Cách làm: Sử dụng MCP với Claude Desktop, nhập lệnh `SELECT` để truy xuất dữ liệu.

Đánh giá: Hiệu quả trong truy xuất dữ liệu, nên dùng khi cần phân tích dữ liệu.

Tối ưu hóa MCP với API

Vấn đề: Cần kết nối AI với API để mở rộng chức năng.

Cách làm: Sử dụng MCP với Claude Code, nhập lệnh `curl` để gọi API.

Đánh giá: Hiệu quả trong mở rộng chức năng, nên dùng khi cần tích hợp với dịch vụ khác.

BÀI HỌC AI HÔM NAY CHO DEV

1. **Tối ưu chi phí & hiệu năng LLM**
2. Để phát triển ứng dụng trí tuệ nhân tạo (AI) hiệu quả, các nhà phát triển cần biết cách tối ưu hóa 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, đồng thời đảm bảo hiệu suất của ứng dụng. Việc tối ưu hóa LLM cũng giúp giảm thiểu rủi ro về bảo mật và tăng cường sự ổn định của hệ thống.
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 xử lý. Điều này có thể được thực hiện thông qua việc điều chỉnh các tham số của mô hình để phù hợp với nhiệm vụ cụ thể.
4. Tip hoặc bước tiếp theo: Để bắt đầu tối ưu hóa LLM, hãy bắt đầu bằng việc phân tích yêu cầu của ứng dụng và xác định các điểm cần tối ưu hóa, sau đó áp dụng các kỹ thuật như **fine-tuning** và **LoRA** để đạt được hiệu suất cao hơn.