BIS - MT · 25/06/2026

Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

> *“Thất bại là mẹ thành công.”*
>
> — Tục ngữ Việt Nam

Mỗi thất bại đều mang theo bài học quý giá — người biết rút kinh nghiệm từ thất bại sẽ tiến gần hơn đến thành công.

TIN TỨC NỔI BẬT

1

Giới thiệu Open Agent Specification (Agent Spec): Một định dạng thống nhất cho AI Agent

Introducing the Open Agent Specification (Agent Spec): A Unified Representation for AI Agents

Oracle Blogs Đọc bài viết →

Oracle đã giới thiệu Open Agent Specification (Agent Spec), một biểu diễn thống nhất cho các tác nhân AI. Agent Spec nhằm cung cấp một cách tiêu chuẩn hóa để mô tả các tác nhân AI, cho phép tích hợp và giao tiếp liền mạch giữa các hệ thống khác nhau. Thông số kỹ thuật này được thiết kế để mở và có thể mở rộng, cho phép các nhà phát triển tạo và chia sẻ các tác nhân AI có thể tương tác với các nền tảng và ứng dụng khác nhau. Agent Spec được xây dựng trên các công nghệ hiện có như JSON và JSON-LD, giúp nó dễ dàng tiếp cận và hiểu. Nó cung cấp một cách rõ ràng và súc tích để định nghĩa các tác nhân AI, bao gồm khả năng, mục tiêu và hành vi của chúng. Biểu diễn tiêu chuẩn hóa này dự kiến sẽ tạo điều kiện cho sự phát triển của các hệ thống AI tinh vi và tương tác hơn. Bằng cách áp dụng Agent Spec, các nhà phát triển có thể tạo ra các tác nhân AI có thể làm việc cùng nhau một cách liền mạch, dẫn đến các giải pháp hiệu quả và hiệu quả hơn. Agent Spec là một bước quan trọng hướng tới việc tạo ra một hệ sinh thái AI thống nhất, nơi các tác nhân khác nhau có thể giao tiếp và cộng tác với nhau.

2

AWS open-source MCP Server cho Bedrock AgentCore để tinh gọn 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 thực hiện một bước tiến quan trọng trong lĩnh vực phát triển trí tuệ nhân tạo (AI) bằng cách mở nguồn mã một máy chủ MCP (Giao thức Đa đám mây) cho Bedrock AgentCore. Động thái này nhằm mục đích đơn giản hóa quá trình phát triển các tác nhân AI, là những thành phần quan trọng trong nhiều ứng dụng như trò chơi, robot và dịch vụ khách hàng. Máy chủ MCP được thiết kế để hoạt động mượt mà với Bedrock AgentCore, một framework cho phép phát triển các tác nhân AI. Bằng cách mở nguồn mã máy chủ MCP, AWS cung cấp cho các nhà phát triển một cách tiêu chuẩn hóa và hiệu quả để tích hợp các tác nhân AI của họ với các dịch vụ đám mây khác nhau. Động thái này dự kiến sẽ đẩy nhanh sự phát triển của các tác nhân AI, giúp các nhà phát triển dễ dàng tạo và triển khai các ứng dụng thông minh. Việc mở nguồn mã máy chủ MCP cũng thúc đẩy sự hợp tác và đổi mới trong cộng đồng AI, cho phép các nhà phát triển đóng góp và hưởng lợi từ công nghệ.

3

Tinh gọn workflow trên 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 giới thiệu một tích hợp mới cho phép người dùng tối ưu hóa các quy trình làm việc trên GitHub với sự giúp đỡ của AI tạo sinh. Tích hợp này kết hợp Amazon Bedrock và MCP (Meta's Codebert), cho phép các nhà phát triển tận dụng sức mạnh của AI để tự động hóa và tối ưu hóa các quy trình phát triển của họ. Với tích hợp này, các nhà phát triển có thể sử dụng Amazon Bedrock để tạo và quản lý các model AI, và sau đó triển khai chúng trên GitHub bằng MCP. Điều này cho phép các nhà phát triển tự động hóa các nhiệm vụ như tạo mã, kiểm tra và gỡ lỗi, giúp các quy trình làm việc của họ trở nên hiệu quả và năng suất hơn. Bằng cách kết hợp điểm mạnh của Amazon Bedrock và MCP, các nhà phát triển có thể mở khóa toàn bộ tiềm năng của AI tạo sinh và đưa các quy trình làm việc trên GitHub của họ lên tầm cao mới. Tích hợp này dự kiến sẽ cách mạng hóa cách các nhà phát triển làm việc, giúp họ dễ dàng xây dựng, kiểm tra và triển khai các ứng dụng phần mềm chất lượng cao.

4

Từ các dự án open-source thành công 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 các dự án mã nguồn mở khác nhau và tác động của chúng. Ông thảo luận về các dự án mã nguồn mở ít được biết đến, chẳng hạn như react-boilerplate và styled-components, và cách chúng so sánh với các dự án nổi bật hơn. Stoiber cũng nói về sự phát triển của GitHub, đặc biệt là vai trò của Spectrum trong việc định hình GitHub Discussions. Ngoài ra, ông chia sẻ kinh nghiệm xây dựng Stellate, một bộ nhớ đệm GraphQL đã được Shopify và The Guild mua lại. Cuộc trò chuyện cũng đề cập đến việc phát triển các ứng dụng ChatGPT, mà Stoiber mô tả là một bề mặt mới cho phần mềm. Tập này được tài trợ bởi Coder, WorkOS, Notion, Fly.io và Changelog++.

5

Hình dạng, đối xứng và cấu trúc: Vai trò thay đổi của toán học trong nghiên cứu Machine Learning

Shape, Symmetries, and Structure: The Changing Role of Mathematics in Machine Learning Research

The Gradient Đọc bài viết →

> Trong những năm gần đây, nghiên cứu học máy đã chứng kiến một sự thay đổi đáng kể, với tiến bộ thực nghiệm vượt qua sự hiểu biết lý thuyết. Mặc dù vậy, toán học vẫn là một thành phần quan trọng của nghiên cứu học máy, mặc dù vai trò của nó đang thay đổi. Sự phụ thuộc truyền thống vào các nguyên tắc toán học để hướng dẫn nghiên cứu đã nhường chỗ cho các phương pháp tính toán và kỹ thuật mạnh mẽ hơn. Tuy nhiên, toán học không bị thay thế, mà ứng dụng của nó đang mở rộng sang các lĩnh vực và mục đích mới. Hiện nay, nó được sử dụng để giải thích hậu hoc các hiện tượng thực nghiệm và cho các lựa chọn thiết kế cấp cao hơn, chẳng hạn như phù hợp kiến trúc với cấu trúc nhiệm vụ hoặc đối xứng dữ liệu. Sự chuyển dịch sang quy mô cũng đã mở rộng phạm vi các lĩnh vực toán học có thể áp dụng cho học máy, bao gồm cả các lĩnh vực toán học "tinh khiết" như tô pô, đại số và hình học. Những lĩnh vực này đang được tận dụng để giải quyết các thách thức lớn nhất trong học sâu hiện đại. Các nhà nghiên cứu đang khám phá các cách mới để áp dụng toán học nhằm hướng dẫn

khám phá và hiểu biết trong học máy, nhấn mạnh sự liên quan bền vững của nó trong lĩnh vực này.

6 AI Agent của doanh nghiệp bạn nên tự động ghi nhớ model phù hợp cho từng task. Mindstone đã xây dựng tính năng này với Rebel

Your enterprise AI agents should automatically remember which model is right for which task. Mindstone built the capability with Rebel

VentureBeat Đọc bài viết →

Công ty khởi nghiệp chuyển đổi AI có trụ sở tại London, Mindstone, đã ra mắt một hệ điều hành AI cục bộ đầu tiên, được gọi là Rebel. Hệ thống này được thiết kế để đơn giản, có thể tùy chỉnh và mở rộng, cho phép các đội áp dụng và sửa đổi nó để phù hợp với nhu cầu của họ. Các tính năng chính của Rebel bao gồm một lớp bộ nhớ chung đảm bảo các tác nhân sử dụng các model AI được doanh nghiệp ưa chuộng cho từng nhiệm vụ hoặc nhiệm vụ con, chuyển đổi động giữa các model cục bộ và đám mây khi cần. Cách tiếp cận này nhằm tiết kiệm chi phí, duy trì quyền riêng tư và bảo mật dữ liệu, và cung cấp một cách làm việc có thể dự đoán và rõ ràng. Rebel lưu trữ trạng thái, lời nhắc, hướng dẫn nhiệm vụ và thứ bậc bộ nhớ trong các tệp markdown, giúp dễ dàng kiểm tra, di chuyển hoặc sửa đổi chúng khi cần. Hệ thống cũng cho phép người dùng tạo các quy trình làm việc AI có thể lặp lại, bao gồm kỹ năng, toán tử và tự động hóa. Điều khiển đa model là một tính năng quan trọng khác, cho phép Rebel chia nhiệm vụ thành các phần và định tuyến các bước khác nhau đến các model khác nhau, bao gồm cả model cục bộ và dựa trên đám mây. CEO của Mindstone, Joshua Wöhle, cho rằng Rebel giảm thiểu rủi ro bị mắc kẹt trong giao diện hoặc cơ sở dữ liệu của một nhà cung cấp SaaS duy nhất. Công ty đang phát hành Rebel dưới giấy phép Fair Source, cho phép các cá nhân và tổ chức có tối đa 100 người dùng đồng thời chạy nó miễn phí. Rebel đã được triển khai trên lực lượng lao động 250 người của khách hàng Epignosis, thu hồi lại khả năng tương đương với tám vai trò toàn thời gian trong thời gian triển khai 12 tuần.

7 Nắng nóng cực đoan ở châu Âu đang khiến các nhà máy điện phải ngừng hoạt động

Europe's extreme heat is shutting down power plants

MIT Tech Review Đọc bài viết →

> Châu Âu đang trải qua một đợt sóng nhiệt kỷ lục, đẩy mạng lưới điện đến giới hạn. Khi mọi người sử dụng quạt và điều hòa nhiệt độ để giữ mát, các nhà máy điện đang gặp khó khăn để đáp ứng nhu cầu. Tại Pháp, ngày nóng nhất kể từ năm 1947 đã thấy nhiệt độ vượt quá 44°C (111°F), buộc các nhà máy điện hạt nhân phải đóng cửa hoặc giảm công suất. Nhà máy điện hạt nhân Golfech ở miền nam Pháp đã buộc phải đóng cửa một lò phản ứng do nhiệt độ nước cao trong sông Garonne, được sử dụng để làm mát. Các lò phản ứng khác đang bị hạn chế hoặc sẽ bị đóng cửa vào cuối tuần. Đây không phải là lần đầu tiên nhiệt độ cực đoan ảnh hưởng đến ngành công nghiệp hạt nhân của Pháp, với ít nhất bảy gigawatt năng lượng hạt nhân bị đóng cửa trong một đợt sóng nhiệt vào tháng 7 năm 2025. Các hình thức sản xuất điện khác, bao gồm thủy điện và nhà máy khí đốt tự nhiên, cũng đang gặp khó khăn do nhiệt độ cao. Nhu cầu tăng cao cho việc làm mát đang gây áp lực lên mạng lưới điện, với các chuyên gia cảnh báo rằng những thách thức mà nhiệt độ gây ra cho mạng lưới điện sẽ chỉ trở nên tồi tệ hơn khi biến đổi khí hậu mang lại những đợt sóng nhiệt thường xuyên và mạnh mẽ hơn.

8

Các ưu đãi Prime Day tốt nhất cho mặt nạ LED và dụng cụ mọc tóc thực sự hiệu quả

Best Prime Day Deals on LED Masks and Hair Growth Tools That Actually Work

Wired Đọc bài viết →

TIPS & TRICKS CHO DEV

Tự động hóa test case

Vấn đề: Viết test case thủ công tốn thời gian và dễ xảy ra lỗi.

Cách làm: Sử dụng AI tools như Testim.io để tự động hóa test case, nhập lệnh `testim.io record` để bắt đầu.

Đánh giá: Hiệu quả cao, giảm thời gian viết test case, nhưng cần config chính xác.

Code review tự động

Vấn đề: Code review thủ công tốn thời gian và dễ bỏ qua lỗi.

Cách làm: Sử dụng GitHub Code Review với prompt "review this code" để tự động kiểm tra code.

Đánh giá: Hiệu quả cao, giảm thời gian review, nhưng cần setup đúng.

QA automation

Vấn đề: Kiểm tra chất lượng sản phẩm thủ công tốn thời gian và dễ xảy ra lỗi.

Cách làm: Sử dụng Selenium với lệnh `selenium webdriver` để tự động hóa kiểm tra.

Đánh giá: Hiệu quả cao, giảm thời gian kiểm tra, nhưng cần viết script chính xác.

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 nhà phát triển 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, đồng thời đảm bảo hiệu suất của ứng dụng. Việc tối ưu hóa LLM cũng giúp các nhà phát triển có thể triển khai ứng dụng trên nhiều nền tảng khác nhau.
3. Ví dụ, có thể sử dụng kỹ thuật fine-tuning và LoRA (Low-Rank Adaptation) để tối ưu hóa LLM cho một use case cụ thể. Điều này giúp giảm thiểu số lượng tham số cần đào tạo và tăng tốc độ xử lý.
4. Tip hoặc bước tiếp theo: Các nhà phát triển nên thử nghiệm với các kỹ thuật tối ưu hóa khác nhau, như quantization và pruning, để tìm ra phương pháp phù hợp nhất cho ứng dụng của mình.