BIS - MT · 26/06/2026

Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

> *"Đi một ngày đàng, học một sàng khôn."*
>
> — Tục ngữ Việt Nam

Trải nghiệm thực tế luôn mang lại tri thức và sự hiểu biết sâu sắc hơn bất kỳ sách vở nào.

TIN TỨC NỔI BẬT

1

Tăng tốc phát triển với server Amazon Bedrock AgentCore MCP | Artificial Intelligence

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 server Amazon Bedrock AgentCore MCP, được thiết kế để tăng tốc phát triển trong lĩnh vực Artificial Intelligence (AI). Server Bedrock AgentCore MCP là một thành phần chủ chốt của nền tảng Amazon Bedrock, với mục tiêu đơn giản hóa việc phát triển và triển khai các AI model. Server này được tối ưu hóa cho high-performance computing và cung cấp một môi trường scalable, secure cho việc phát triển AI. Nó cho phép các developer tập trung vào việc xây dựng và training các AI model mà không cần lo lắng về hạ tầng bên dưới. Server Bedrock AgentCore MCP hỗ trợ nhiều AI framework và tool khác nhau, giúp các developer lựa chọn môi trường phát triển ưa thích của mình. Bằng cách tận dụng server Bedrock AgentCore MCP, các developer có thể tăng tốc quá trình phát triển AI, giảm chi phí và cải thiện hiệu quả tổng thể. Các tính năng scalability và security của server cũng cho phép các developer triển khai AI model của họ trong môi trường production-ready, biến nó thành một giải pháp hấp dẫn cho các doanh nghiệp muốn tích hợp AI vào hoạt động của mình.

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 Azure đã giới thiệu Agent Factory, một developer tool được thiết kế để tạo điều kiện thuận lợi cho việc phát triển và triển khai agent nhanh chóng. Tool này cho phép các developer tạo và quản lý các agent, vốn là các software component có thể tương tác với nhiều hệ thống và service khác nhau. Agent Factory cho phép tạo ra các prototype và chuyển đổi liền mạch sang các agent production-ready. Với Agent Factory, các developer có thể xây dựng, test và deploy agent một cách nhanh chóng và hiệu quả. Tool này cung cấp một loạt các tính năng và khả năng, bao gồm phát triển, triển khai và quản lý agent. Điều này giúp các developer tập trung vào việc xây dựng và tích hợp agent vào ứng dụng của họ, thay vì lo lắng về hạ tầng bên dưới. Agent Factory là một phần trong nỗ lực rộng lớn hơn của Microsoft Azure nhằm hỗ trợ phát triển các intelligent agent và các giải pháp AI-powered khác. Tool này được thiết kế để có tính scalable và flexible cao, phù hợp với nhiều loại ứng dụng và use case. Bằng cách cung cấp trải nghiệm phát triển tinh gọn, Agent Factory hướng tới việc tăng tốc tạo và triển khai các intelligent agent cũng như các giải pháp AI-powered khác.

3

Xây dựng 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 để xây dựng một intelligent financial analysis agent sử dụng LangGraph và Strands Agents. LangGraph là một natural language processing (NLP) library cho phép người dùng phân tích và hiểu dữ liệu tài chính phức tạp. Strands Agents, mặt khác, là một low-code, no-code platform để xây dựng các conversational interface. Bằng cách kết hợp LangGraph và Strands Agents, người dùng có thể tạo ra một financial analysis agent có khả năng trích xuất insight từ dữ liệu tài chính, cung cấp các recommendation được cá nhân hóa và tương tác trò chuyện với người dùng. Agent này có thể được tích hợp với nhiều hệ thống tài chính khác nhau, cho phép nó truy cập và phân tích các dataset lớn. Giải pháp này cho phép người dùng xây dựng một conversational interface có thể hiểu các user query, cung cấp thông tin tài chính liên quan và đưa ra lời khuyên hữu ích. Intelligent financial analysis agent này có thể giúp người dùng đưa ra các quyết định đầu tư sáng suốt, theo dõi

hiệu suất tài chính và tối ưu hóa kế hoạch tài chính của họ. Giải pháp được thiết kế để có tính scalable và flexible, phù hợp với nhiều ứng dụng tài chính khác nhau.

4

Định vị lại ngành bán lẻ trong kỷ nguyên AI

Repositioning retail for the AI era

MIT Tech Review Đọc bài viết →

Macy's đang định vị lại chiến lược bán lẻ của mình để ưu tiên artificial intelligence (AI) như một triết lý vận hành. Theo Murali Murugan, senior director of engineering tại Macy's, cách tiếp cận "AI-first" này bao gồm việc nhúng intelligence trực tiếp vào các hệ thống, thay vì đặt nó lên trên các workflow hiện có. Cách tiếp cận này nhằm mục đích đưa ra quyết định nhanh hơn và tạo ra những trải nghiệm phù hợp hơn cho khách hàng một cách mặc định. Macy's đang tích hợp AI vào nhiều lĩnh vực khác nhau, bao gồm personalization, search, operational planning và software development. Chiến lược của công ty là một phần của sự thay đổi lớn hơn trong ngành bán lẻ, chuyển từ các AI pilot riêng lẻ sang các hệ thống tích hợp có thể rút ngắn khoảng cách giữa tín hiệu và hành động. Cách tiếp cận này đã mang lại những thành công nhanh chóng trong các lĩnh vực như search recommendation và customer engagement, và hiện đang được mở rộng sang conversational commerce thông qua các tool như Ask Macy's, một AI-powered shopping assistant. Tầm nhìn dài hạn là một ngành bán lẻ liền mạch, thích ứng và được cá nhân hóa, được hỗ trợ bởi các hệ thống mà khách hàng thậm chí có thể không nhận ra sự hiện diện của chúng.

5

Chi phí ẩn của những dòng code không ai đụng đến

The hidden cost of code that nobody touches

Sourcegraph Blog Đọc bài viết →

Chi phí ẩn của những dòng code không ai đụng đến. Mọi engineering organization đều có một tập hợp các file code bị bỏ quên và không được quản lý, thường được gọi là "dead code". Những file này có thể là một gánh nặng đáng kể, nhưng chi phí thực sự của chúng thường bị bỏ qua. Sự tồn tại của dead code có thể dẫn đến một loạt các vấn đề, bao gồm tăng chi phí maintenance, rủi ro security và giảm productivity. Khi các developer phải đối mặt với một codebase lớn, họ có thể gặp

khó khăn trong việc xác định và ưu tiên những file nào là cần thiết và những file nào có thể được loại bỏ một cách an toàn. Kết quả là, dead code có thể tiếp tục tích lũy, khiến việc quản lý và maintain toàn bộ codebase ngày càng khó khăn. Điều này cuối cùng

6

Notion khai tử ứng dụng email lấy cảm hứng từ Skiff vì hầu hết người dùng chuyển sang dùng AI agent

Notion killing Skiff-influenced email app since most users use AI agents instead

Ars Technica Đọc bài viết →

Notion đã bố ngừng hoạt động của ứng dụng email, Notion Mail, được ra mắt vào tháng 4 năm 2025. Quyết định này được đưa ra sau một năm phát triển, chủ yếu bởi các nhân viên cũ của Skiff được Notion mua lại vào tháng 2 năm 2024. Notion cho biết rằng hầu hết người dùng dựa vào các tác nhân AI để xử lý thư từ của họ, với hơn một nửa người dùng Notion Mail quản lý email mà không cần mở hộp thư đến. Kết quả là, Notion đang chuyển trọng tâm sang quản lý email dựa trên AI. Hộp thư Notion Mail sẽ bị tắt vào ngày 22 tháng 9 và người dùng sẽ có thể xuất dữ liệu của mình, bao gồm cả bản nháp và email đã lên lịch, vào ngày 21 tháng 9. Người dùng Notion Mail phụ thuộc vào bảo vệ HIPAA được tư vấn chuyển sang dịch vụ email mới trước ngày 30 tháng 6 năm 2026. Động thái này hiệu quả đã chấm dứt những tàn dư của dịch vụ email Skiff, được Notion mua lại vào năm ngoái.

7

Amazon Prime Day: Giá preorder Project Hail Mary 4K UHD giảm xuống mức thấp kỷ lục

Amazon Prime Day: Project Hail Mary 4K UHD Preorder Price Drops to a New Low

Dev.to AI Đọc bài viết →

Ngày Prime Day của Amazon đang đến gần và gã khổng lồ công nghệ này đã cung cấp một cái nhìn sơ bộ về các ưu đãi và giảm giá đang chờ đợi. Một trong những trò chơi được mong đợi nhất trong năm, Project Hail Mary, đã giảm xuống mức giá đặt hàng trước thấp nhất từ trước đến nay trên Amazon. Được phát triển bởi Night School Studio, trò chơi khoa học viễn tưởng này được đặt trong một tương lai xa nơi loài người đang đứng trên bờ vực tuyệt chủng, và người chơi phải khám phá ra những bí mật của vũ trụ. Giá đặt hàng trước của trò chơi đã được giảm xuống mức thấp mới, khiến nó trở thành một ưu đãi hấp

dẫn cho các game thủ. Thông thường, đặt hàng trước cho các trò chơi có ngày phát hành không chắc chắn đi kèm với mức giá cao hơn, nhưng Amazon đang cung cấp đặt hàng trước Project Hail Mary 4K UHD với mức giá không thể đánh bại. Việc giảm giá này rất đáng kể, đặc biệt là khi xem xét sự không chắc chắn xung quanh ngày phát hành, và đây là một cơ hội tuyệt vời cho các game thủ để tham gia sớm. Với đồ họa tuyệt đẹp, cốt truyện hấp dẫn và gameplay hấp dẫn, Project Hail Mary đang trở thành một trong những trò chơi được mong đợi nhất trong năm.

8

Tại sao một ngân hàng cần có Chief Scientist?

Why Does a Bank Need a Chief Scientist?

IEEE Spectrum Đọc bài viết →

Capital One, một tổ chức tài chính phục vụ hơn 100 triệu khách hàng, đang chuyển đổi lĩnh vực tài chính thông qua đổi mới được thúc đẩy bởi AI. Để đạt được điều này, ngân hàng đang xây dựng một cộng đồng khoa học và tổ chức nghiên cứu để thúc đẩy ranh giới của AI và khám phá khoa học trong lĩnh vực tài chính. Prem Natarajan, một cựu nhân viên nghiên cứu được tài trợ bởi DARPA và học thuật, đang dẫn đầu nỗ lực này với tư cách là Chief Scientist của ngân hàng. Natarajan tin rằng những vấn đề AI phức tạp nhất được tìm thấy trong các lĩnh vực dọc như tài chính, nơi các vấn đề của khách hàng trong thế giới thực, kiến thức kinh doanh theo ngữ cảnh và học tập liên tục là rất quan trọng. Không giống như các công ty công nghệ lớn, tập trung vào các nền tảng ngang, Capital One đang giải quyết các thách thức cụ thể cho lĩnh vực, chẳng hạn như phát hiện gian lận trong thời gian thực và cung cấp các công cụ trò chuyện tiên tiến. Những thách thức này đòi hỏi nghiên cứu ban đầu và đổi mới khoa học, sau đó được chuyển trở lại vào kinh doanh để tạo ra các ứng dụng trong thế giới thực. Với cơ sở hạ tầng hiện đại, cách tiếp cận kỷ luật đối với quản trị và đội ngũ tài năng sâu, Capital One được đặt ở vị trí dẫn đầu trong lĩnh vực AI doanh nghiệp. Natarajan nhấn mạnh rằng việc giải quyết các vấn đề AI quan trọng và thấy chúng trở thành hiện thực là có thể tại Capital One, khiến nó trở thành một nơi hấp dẫn cho các nhà nghiên cứu và nhà khoa học.

TIPS & TRICKS CHO DEV

Cài Đặt Ollama

Vấn đề: Chạy LLM trên máy cá nhân mà không cần internet.

Cách làm: Sử dụng Ollama, chạy lệnh `ollama --model tiny` để tải mô hình. Ví dụ prompt: "Tóm tắt nội dung của văn bản này".

Đánh giá: Hiệu quả cho các mô hình nhỏ, phù hợp với máy tính cấu hình thấp.

Tối Ưu LM Studio

Vấn đề: Tối ưu hóa hiệu suất của LM Studio trên máy cá nhân.

Cách làm: Sử dụng lệnh `lm-studio --optimize` để tối ưu hóa mô hình. Ví dụ: `lm-studio --optimize --model medium`.

Đánh giá: Phù hợp với máy tính cấu hình trung bình, giúp tăng tốc độ xử lý.

Sử Dụng CLI

Vấn đề: Chạy LLM trên máy cá nhân mà không cần giao diện đồ họa.

Cách làm: Sử dụng lệnh `ollama --cli` để chạy Ollama trên CLI. Ví dụ: `ollama --cli --prompt "Tóm tắt nội dung"`.

Đánh giá: Phù hợp với người dùng quen với giao diện lệnh, giúp tăng tốc độ làm việc.

BÀI HỌC AI HÔM NAY CHO DEV

1. **Tối ưu chi phí & hiệu năng LLM**
2. Để phát triển ứng dụng hiệu quả, các nhà phát triển cần tối ưu hóa chi phí và hiệu năng của ngôn ngữ mô hình lớn (LLM). Điều này giúp giảm thiểu chi phí vận hành và cải thiện tốc độ xử lý.
3. Ví dụ, bằng cách sử dụng kỹ thuật fine-tuning và LoRA, chúng ta có thể giảm kích thước mô hình và tăng tốc độ đào tạo.
4. Tip: Sử dụng kỹ thuật pruning và quantization để giảm kích thước mô hình và tăng tốc độ xử lý, từ đó tối ưu hóa chi phí và hiệu năng của LLM.