



Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

✨ *“Change your thoughts and you change your world.”*

↳ Thay đổi suy nghĩ của bạn và bạn thay đổi thế giới của mình.

— Norman Vincent Peale

💡 *Tư duy quyết định thực tại — khi thay đổi cách nhìn nhận và suy nghĩ, bạn sẽ nhận ra những cơ hội và khả năng mà trước đây không thấy.*

TIN TỨC NỔI BẬT

Lớp lệnh AI nhà kho đa tác nhân (Multi-Agent) cho phép đạt được sự xuất sắc trong hoạt động và thông minh chuỗi cung ứng

1


🇬🇧 [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 thông tin chuỗi cung ứng. Giải pháp sáng tạo này tận dụng AI để tối ưu hóa hoạt động nhà kho, tự động hóa quy trình và cung cấp khả năng hiển thị thời gian thực vào các hoạt động chuỗi cung ứng. Lớp lệnh AI Nhà kho Đa tác nhân sử dụng kiến trúc đa tác nhân, nơi nhiều tác nhân AI làm việc cùng nhau để quản lý các khía cạnh khác nhau của hoạt động nhà kho. Các tác nhân này có thể được đào tạo để thực hiện các nhiệm vụ như quản lý hàng tồn kho, hoàn tất đơn hàng và bảo trì dự đoán, cho phép hoạt động nhà kho hiệu quả và tự động hơn. Bằng cách triển khai giải pháp này, các doanh nghiệp có thể thu được thông tin chi tiết quý giá về chuỗi cung ứng của mình, cho phép họ đưa ra quyết định dựa trên dữ liệu và cải thiện tính xuất sắc hoạt động tổng thể. Lớp lệnh AI Nhà kho Đa tác nhân được xây dựng trên nền tảng của nhà phát triển NVIDIA, cung cấp một khuôn khổ có thể mở rộng và linh hoạt để tích hợp AI vào hoạt động nhà kho. Giải pháp này có tiềm năng thay đổi cách thức hoạt động của nhà kho, cải thiện hiệu quả, giảm chi phí và nâng cao sự hài lòng của khách hàng.

2

Nhà máy tác nhân (Agent Factory): Kết nối các tác nhân, ứng dụng và dữ liệu với các tiêu chuẩn mở mới như MCP và A2A

 *Agent Factory: Connecting agents, apps, and data with new open standards like MCP and A2A*

 Microsoft Azure [Đọc bài viết →](#)

Microsoft đã giới thiệu Agent Factory, một nền tảng được thiết kế để kết nối các tác nhân, ứng dụng và dữ liệu bằng cách sử dụng các tiêu chuẩn mở mới. Nền tảng này tận dụng công nghệ Cloud PC (MCP) và Ứng dụng đến Ứng dụng (A2A) của Microsoft để tạo điều kiện cho các tương tác không gián đoạn giữa các thành phần khác nhau. Agent Factory cho phép các nhà phát triển tạo ra các tác nhân tùy chỉnh có thể tương tác với các ứng dụng và nguồn dữ liệu khác nhau, thúc đẩy sự linh hoạt và khả năng tương tác cao hơn. Bằng cách sử dụng các tiêu chuẩn MCP và A2A, Agent Factory cho phép trao đổi dữ liệu an toàn và hiệu quả giữa các tác nhân, ứng dụng và nguồn dữ liệu. Nền tảng này dự kiến sẽ mang lại lợi ích cho các ngành công nghiệp khác nhau, bao gồm chăm sóc sức khỏe, tài chính và bán lẻ, nơi tích hợp nhiều hệ thống và nguồn dữ liệu là rất quan trọng để hoạt động hiệu quả. Mục tiêu của Agent Factory của Microsoft là đơn giản hóa quá trình kết nối các thành phần khác nhau, cho phép các nhà phát triển tập trung vào việc xây dựng các ứng dụng và dịch vụ sáng tạo mang lại giá trị kinh doanh.

3

Giới thiệu Thông số kỹ thuật tác nhân mở (Open Agent Specification - Agent Spec): Một biểu diễn thống nhất cho các tác nhân AI

 *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. Đặc tả này nhằm cung cấp một khuôn khổ chung cho các nhà phát triển để tạo, tích hợp và tương tác với các tác nhân AI trên nhiều nền tảng và ứng dụng khác nhau. Agent Spec được thiết kế để mở, có thể mở rộng và không phụ thuộc vào ngôn ngữ, cho phép các nhà phát triển xây dựng các tác nhân AI có thể dễ dàng tích hợp với các hệ thống và dịch vụ khác nhau. Đặc tả này phác thảo một tập hợp các giao diện và cấu trúc dữ liệu tiêu chuẩn cho các tác nhân AI, cho phép giao tiếp và cộng tác liền mạch giữa các tác nhân và hệ thống. Biểu diễn thống nhất này dự kiến sẽ

thúc đẩy sự phát triển của các hệ thống AI tinh vi và liên kết hơn, thúc đẩy đổi mới trong các lĩnh vực như AI đối thoại, robot và nhà thông minh. Bằng cách cung cấp một ngôn ngữ và khuôn khổ chung cho các tác nhân AI, Agent Spec có tiềm năng tăng tốc việc áp dụng các công nghệ AI và thúc đẩy một hệ sinh thái AI gắn kết và tương tác hơn.

4

Tại sao nợ lời nhắc (prompt debt), nợ thu hồi (retrieval debt) và nợ đánh giá (evaluation debt) đang âm thầm thay đổi lại rủi ro AI trong doanh nghiệp

 *Why prompt debt, retrieval debt, and evaluation debt are quietly reshaping enterprise AI risk*

 VentureBeat [Đọc bài viết →](#)

Hệ thống AI doanh nghiệp đang giới thiệu các lớp nợ kỹ thuật mới, thường được gọi là "nợ AI", có thể tinh vi và phi tuyến hơn so với nợ kỹ thuật truyền thống. Nợ này biểu hiện trên các lời nhắc, model, và phụ thuộc dữ liệu, khiến nó khó đo lường và nguy hiểm hơn. Một nghiên cứu của MIT năm 2025 cho thấy 95% dự án AI không đạt được sản xuất hoặc mang lại giá trị, trong khi một nghiên cứu của S&P Global Market Intelligence cho thấy 42% doanh nghiệp đã hủy bỏ nhiều sáng kiến AI vào năm 2025. Nợ AI thường đến dưới bốn hình thức: nợ lời nhắc, nợ phụ thuộc model, nợ truy xuất, và nợ đánh giá. Nợ lời nhắc đề cập đến các lời nhắc không được ghi chép hoặc quản lý kém, trong khi nợ phụ thuộc model phát sinh từ việc phụ thuộc vào các model bên ngoài có thể thay đổi và mất khả năng tái tạo. Nợ truy xuất xảy ra khi hệ thống AI trả về thông tin lỗi thời hoặc không liên quan do các kho dữ liệu lộn xộn. Nợ đánh giá được gây ra bởi sự thiếu tiêu chuẩn hóa trong kiểm tra và giám sát các model và ứng dụng AI. Những hình thức nợ AI này kết hợp với nợ kỹ thuật truyền thống để tạo ra các rủi ro quy mô lớn có thể gây ra sự thất bại thảm hại của toàn bộ triển khai doanh nghiệp. Để ngăn chặn nợ AI, các doanh nghiệp cần đối xử với các lời nhắc như mã code, thiết lập các đường ống đánh giá liên tục, và bao gồm khả năng giải thích trong kết quả AI. Điều này yêu cầu các chương trình giảm nợ AI rõ ràng và các ngân sách liên quan, được thúc đẩy bởi các nhà lãnh đạo chủ chốt ở cấp CXO.

5

OpenAI, Grupo Folha và Grupo UOL công bố quan hệ đối tác nội dung chiến lược

 *OpenAI, Grupo Folha and Grupo UOL announce strategic content partnership*

 OpenAI Blog [Đọc bài viết →](#)

OpenAI đã hình thành một quan hệ đối tác nội dung chiến lược với Grupo Folha và Grupo UOL, hai tập đoàn truyền thông hàng đầu của Brazil. Sự hợp tác này nhằm mục đích mang báo chí Brazil đáng tin cậy đến ChatGPT, một AI trò chuyện được phát triển bởi OpenAI. Quan hệ đối tác này sẽ cung cấp cho người dùng quyền truy cập vào nội dung tin tức từ Grupo Folha và Grupo UOL, tập trung vào tính minh bạch và thuộc tính. Điều này có nghĩa là người dùng sẽ có thể tương tác với các bài báo tin tức và nhận được thông tin chính xác trong khi nhận thức được nguồn gốc. Quan hệ đối tác này dự kiến sẽ mở rộng sự sẵn có của nội dung tin tức chất lượng cao đến khán giả rộng lớn hơn, tận dụng khả năng của ChatGPT để cung cấp thông tin này theo cách thân thiện với người dùng và dễ tiếp cận. Bằng cách tích hợp báo chí Brazil đáng tin cậy vào ChatGPT, quan hệ đối tác này nhằm mục đích thúc đẩy tính minh bạch, chính xác và uy tín trong việc phổ biến tin tức.

6

Đặt cược lớn của Mỹ vào tính toán lượng tử có thể không hoàn toàn hợp pháp

 *US's big bet on quantum computing may not be entirely legal*

 Ars Technica [Đọc bài viết →](#)

Chính phủ Mỹ đã phân bổ 2 tỷ đô la để đầu tư vào các công ty máy tính lượng tử, với 100 triệu đô la mỗi công ty được trao cho các công ty khởi nghiệp khác nhau để đổi lấy cổ phần. Tuy nhiên, một thành viên của Quốc hội Mỹ, Zoe Lofgren, đang đặt câu hỏi về tính hợp pháp của những thỏa thuận này. Theo Lofgren, nguồn tài trợ này được dành để hỗ trợ nghiên cứu công về bán dẫn thông qua Đạo luật CHIPS và Khoa học, không phải máy tính lượng tử. Phần lớn tiền sẽ đi đến Anderton, một công ty được thành lập bởi IBM và chính phủ, sẽ hoạt động như một xưởng đúc để chế tạo đơn vị xử lý lượng tử. Lofgren cho rằng thỏa thuận này là bất hợp pháp vì nó không phù hợp với mục đích ban đầu của nguồn tài trợ. Cô cũng gợi ý rằng một cựu giám đốc điều hành của IBM có thể đã tham gia vào các cuộc đàm phán, điều này có thể là xung đột lợi ích. Thỏa thuận này đã gây ra tranh cãi, với một số người đặt câu hỏi về sự cần thiết của một công ty xưởng đúc riêng biệt và phân bổ nguồn vốn.

7

Việc sa thải hàng loạt của ClickUp cho chúng ta biết gì về tương lai của công việc

 [What ClickUp's mass layoff tells us about the future of work](#)

 TechCrunch AI [Đọc bài viết →](#)

ClickUp, một công ty khởi nghiệp về phần mềm hợp tác được định giá 4 tỷ đô la vào năm 2021, đã \square bố sa thải hàng loạt 22% lực lượng lao động của mình. Tuy nhiên, CEO Zeb Evans khẳng định việc cắt giảm này không phải là biện pháp cắt giảm chi phí, mà là một động thái chiến lược để tận dụng công nghệ AI. ClickUp đã giới thiệu 3.000 tác nhân AI nội bộ để xử lý các nhiệm vụ phức tạp, cho phép nhân viên tập trung vào chỉ đạo và xem xét đầu ra. Evans nhằm sử dụng AI để thúc đẩy công ty đến một "100x org", nơi lợi ích về năng suất được tối đa hóa. Phương pháp này không độc đáo đối với ClickUp, vì 80% công ty sử dụng công nghệ tự động đã cắt giảm việc làm, theo một cuộc khảo sát của Gartner. Trong khi một số nhà phê bình cho rằng các công ty đang sử dụng AI chưa được chứng minh như một lý do để giảm quy mô, ClickUp vẫn khẳng định họ đang thấy lợi ích về năng suất đáng kể từ các tác nhân AI. Công ty đang lên kế hoạch tích hợp những hiệu quả này vào một sản phẩm sắp tới cho khách hàng của mình, chuyển trọng tâm từ việc tiêu thụ token sang giá trị được tạo ra và thời gian được tiết kiệm.

8

Ferrari Luce điện đã đến cuối cùng

 [The Electric Ferrari Luce Is Finally Here](#)

 Wired [Đọc bài viết →](#)

Ferrari cuối cùng đã ra mắt chiếc xe điện đầu tiên của mình, Luce, sau nhiều năm chờ đợi. Chiếc xe này lần đầu tiên được gợi ý vào năm 2018 bởi cựu chủ tịch Ferrari Sergio Marchionne, người nhằm tạo ra một siêu xe điện để cạnh tranh với Tesla. Luce tự hào với các số liệu hiệu suất ấn tượng, bao gồm bốn động cơ sản xuất hơn 1.000 mã lực trong chế độ Boost, tăng tốc từ 0-62mph trong 2,5 giây và đạt tốc độ tối đa 192mph. Xe có pin 122 kWh, một trong những pin lớn nhất trong bất kỳ xe điện sản xuất nào, cung cấp phạm vi hơn 329 dặm cho mỗi lần sạc. Luce cũng có hệ dẫn động bốn bánh và hệ thống lái được lấy cảm hứng từ SUV Purosangue, với mỗi bánh xe có sức mạnh, phanh, giảm xóc và hệ thống lái riêng. Ferrari cũng đã phát triển một hệ thống âm thanh độc đáo, sử dụng gia tốc kế để thu và lọc rung động từ động cơ để tạo ra âm thanh động cơ tự nhiên. Thiết kế của Luce được tạo ra với sự hợp tác của LoveFrom, một cơ quan được thành lập bởi Jony Ive,

và có thiết kế bên ngoài khí động học và nội thất sang trọng với trọng tâm là vật liệu cao cấp và các yếu tố thiết kế sáng tạo.

⚡ TIPS & TRICKS CHO DEV

⚡ Tự động kiểm thử đơn vị

Vấn đề: Viết và bảo trì các trường hợp kiểm thử đơn vị tốn thời gian và công sức.

Cách làm: Sử dụng framework như Pytest để tự động hóa quá trình kiểm thử. Ví dụ, lệnh `pytest --cov` giúp tính toán mức độ bao phủ mã.

Đánh giá: Hiệu quả cao trong việc giảm thời gian kiểm thử, nên dùng cho dự án lớn.

⚡ Code Review Tự Động

Vấn đề: Kiểm tra mã nguồn thủ công tốn nhiều thời gian và dễ xảy ra lỗi.

Cách làm: Sử dụng công cụ như GitHub Code Review để tự động kiểm tra mã nguồn. Ví dụ, prompt "Check code style and syntax" giúp phát hiện lỗi cơ bản.

Đánh giá: Nhanh chóng và hiệu quả, nên dùng cho dự án có quy mô trung bình.

⚡ Tự động hóa kiểm thử giao diện

Vấn đề: Kiểm thử giao diện người dùng (UI) tốn nhiều thời gian và dễ xảy ra lỗi.

Cách làm: Sử dụng framework như Selenium để tự động hóa quá trình kiểm thử UI. Ví dụ, lệnh `selenium webdriver` giúp tự động tương tác với giao diện.

Đánh giá: Hiệu quả cao trong việc giảm thời gian kiểm thử UI, nên dùng cho dự án có giao diện phức tạp.

📖 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 tối ưu hóa chi phí và hiệu năng của mô hình ngôn ngữ lớn (LLM). Điều này giúp giảm thiểu chi phí vận hành và cải thiện trải nghiệm người dùng. Việc tối ưu hóa hiệu năng cũng giúp ứng dụng phản hồi nhanh chóng và chính xác.

3. Ví dụ, bằng cách sử dụng kỹ thuật fine-tuning và LoRA (Low-Rank Adaptation), chúng ta có thể giảm kích thước mô hình và cải thiện hiệu năng mà không mất đi độ chính xác. Điều này có thể được thực hiện thông qua việc sử dụng các thư viện như Hugging Face Transformers.

4. 💡 Tip hoặc bước tiếp theo: Hãy bắt đầu bằng việc đánh giá hiệu năng hiện tại 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 cải thiện hiệu năng và giảm chi phí.



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

Nguồn: Google News · Groq AI