BIS - MT · 18/05/2026

🤖 Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

✨ *"Alone we can do so little; together we can do so much."*

↪ Một mình ta làm được rất ít; cùng nhau ta làm được rất nhiều.

— Helen Keller

💡 *Sức mạnh của sự hợp tác vượt xa tổng sức mạnh cá nhân — cộng tác chân thành là nền tảng của mọi thành tựu lớn.*

📰 TIN TỨC NỔI BẬT

1

Lớp điều khiển AI kho hàng đa agent: Tối ưu vận hành và thông minh chuỗi cung ứng

🇬🇧 *Multi-Agent Warehouse AI Command Layer Enables Operational Excellence and Supply Chain Intelligence*

📰 NVIDIA Developer 🔗 Đọc bài viết →

> NVIDIA đã giới thiệu Multi-Agent Warehouse AI Command Layer, được thiết kế để nâng cao hiệu quả vận hành và thông minh chuỗi cung ứng trong các kho hàng. Hệ thống được hỗ trợ bởi AI này cho phép giám sát và kiểm soát hoạt động kho hàng theo thời gian thực, giúp tối ưu hóa quản lý tồn kho, tinh gọn workflow và cải thiện năng suất. Multi-Agent Warehouse AI Command Layer sử dụng sự kết hợp giữa các thuật toán AI và machine learning để phân tích dữ liệu từ nhiều nguồn khác nhau, bao gồm sensor, camera và các thiết bị IoT khác. Dữ liệu này sau đó được dùng để đưa ra các quyết định sáng suốt, chẳng hạn như dự đoán nhu cầu, tối ưu hóa quy trình lưu trữ và truy xuất, cũng như xác định các nút thắt cổ chai tiềm ẩn. Bằng cách tận dụng sức mạnh của AI, hệ thống cũng có thể cung cấp những insight giá trị về hoạt động chuỗi cung ứng, giúp các doanh nghiệp đưa ra các quyết định dựa trên dữ liệu và cải thiện hiệu quả tổng thể. Multi-Agent Warehouse AI Command Layer của NVIDIA là một giải pháp tiên tiến dành cho các kho hàng muốn dẫn đầu đối thủ và đạt được operational excellence.

2

Agent Factory: Kết nối agent, app và data với các chuẩn mở mới (MCP, 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 sáng kiến mới nhằm kết nối các agent, app và data thông qua các open standard. Nền tảng này sử dụng các chuẩn Microsoft Cloud Program (MCP) và Application-to-Application (A2A) để tạo điều kiện tương tác liền mạch giữa các hệ thống khác nhau. Với Agent Factory, các developer có thể tạo và deploy các intelligent agent có khả năng giao tiếp với nhiều ứng dụng và nguồn data khác nhau, giúp các workflow hiệu quả và tự động hơn. Các open standard được Agent Factory sử dụng cho phép linh hoạt và interoperability cao hơn, giúp các hệ thống khác nhau dễ dàng làm việc cùng nhau. Việc giới thiệu Agent Factory là một phần trong nỗ lực rộng lớn hơn của Microsoft nhằm thúc đẩy các open standard và interoperability trong cloud. Bằng cách cung cấp một framework chung để kết nối các agent, app và data, Agent Factory có tiềm năng thúc đẩy innovation và cải thiện trải nghiệm người dùng tổng thể.

3

Giới thiệu Open Agent Specification (Agent Spec): Chuẩn 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 unified representation cho các AI agent. Specification này nhằm mục đích cung cấp một framework chung để các developer tạo, deploy và quản lý các AI agent trên nhiều nền tảng và ứng dụng khác nhau. Agent Spec được thiết kế để language-agnostic, cho phép các developer xây dựng AI agent bằng các programming language mà họ ưa thích. Specification này phác thảo một tập hợp các core component và interface cho các AI agent, bao gồm behavior, knowledge và các tương tác của chúng với environment. Cách tiếp cận tiêu chuẩn hóa này cho phép tích hợp và giao tiếp liền mạch giữa các AI agent, tạo điều kiện phát triển các hệ thống AI phức tạp và tinh vi hơn. Bằng cách áp dụng Agent Spec, các developer có thể tạo ra các AI agent portable, scalable và maintainable hơn. Specification cũng thúc đẩy interoperability và collaboration giữa các developer, researcher và tổ chức làm việc trong các dự án AI. Open Agent Specification là một

open-source initiative, cho phép cộng đồng đóng góp vào sự phát triển và tiến hóa của nó.

4

Nâng tầm: Chất lượng, trách nhiệm chung và tương lai chương trình bug bounty của GitHub

🇬🇧 *Raising the bar: Quality, shared responsibility, and the future of GitHub's bug bounty program*

📰 GitHub Blog 🔗 Đọc bài viết →

GitHub đang cập nhật chương trình bug bounty của mình để ưu tiên các submission chất lượng và làm rõ ranh giới shared responsibility. Công ty đặt mục tiêu phát triển cách thức thưởng cho các phát hiện rủi ro thấp, công nhận số lượng ngày càng tăng của các security researcher đóng góp vào sự an toàn của nền tảng. Với hơn 180 triệu developer tin cậy GitHub, bảo mật của nền tảng là ưu tiên hàng đầu. Chương trình bug bounty đã chứng kiến sự gia tăng đáng kể về số lượng submission, được thúc đẩy bởi sự ra đời của các tool mới, bao gồm artificial intelligence, đã giúp security research dễ tiếp cận hơn. GitHub tin rằng collaboration với các researcher bên ngoài là rất quan trọng để cải thiện bảo mật và vẫn cam kết với chương trình bug bounty của mình. Công ty đang thích nghi với bối cảnh thay đổi, tập trung vào các submission chất lượng và shared responsibility để đảm bảo nền tảng vẫn an toàn cho các developer.

5

Siri cải tiến của Apple có thể có tính năng tự động xóa chat

🇬🇧 *Apple's Siri revamp could include auto-deleting chats*

📰 TechCrunch AI 🔗 Đọc bài viết →

Siri được cải tiến của Apple có thể bao gồm tính năng tự động xóa chat. Theo Mark Gurman của Bloomberg, Siri mới sẽ tập trung vào privacy, một lĩnh vực then chốt mà Apple muốn tạo sự khác biệt so với các công ty artificial intelligence khác. Siri được cập nhật sẽ là một standalone app, được cung cấp sức mạnh bởi Google Gemini, mang đến trải nghiệm chatbot tương tự ChatGPT. Tuy nhiên, không giống như các chatbot khác, Siri sẽ có nhiều hạn chế hơn về việc lưu trữ và sử dụng user data. Cụ thể, người dùng sẽ có thể tự động xóa các cuộc trò chuyện sau 30 ngày hoặc một năm, hoặc giữ chúng vô thời hạn. Việc nhấn mạnh vào privacy này có thể là một động thái chiến lược của Apple để làm nổi bật cam kết bảo vệ user data, mặc dù Google sẽ xử lý một số khía cạnh bảo mật của Siri mới.

6

Giới thiệu LangChain Labs

🇬🇧 *Introducing LangChain Labs*

📰 LangChain Blog 🔗 Đọc bài viết →

LangChain Labs, một nỗ lực applied research mới, đã được ra mắt để thúc đẩy open và applied research cho agent-building. Mục tiêu là làm cho công nghệ continual learning trở nên hữu ích cho cộng đồng agent-building rộng lớn hơn. LangChain Labs đặt mục tiêu thu thập và chuyển đổi các signal hữu ích từ agent data, có thể được áp dụng ở các layer khác nhau của agent stack. Nhóm đang làm việc với các đối tác trong nhiều ngành khác nhau, bao gồm Harvey, Nvidia, Prime Intellect, Fireworks và Baseten, để đạt được mục tiêu này. Nghiên cứu tập trung vào một số lĩnh vực, bao gồm cải thiện agent bằng cách khai thác thông tin từ large-scale agent data, các agent hiệu quả tại Pareto frontier, xây dựng có hệ thống các evaluation và simulation environment, và prompt optimization. LangChain Labs cũng nhằm mục đích giúp việc tạo và chạy các environment để evaluation, simulation và reinforcement learning dễ dàng hơn. Hệ sinh thái open-source của nhóm sẽ tiếp tục là một phần cốt lõi trong cách các builder học hỏi lẫn nhau, với mục tiêu thúc đẩy nhiều open research hơn để cung cấp sức mạnh cho thế hệ self-improving agent tiếp theo. LangSmith, nền tảng agent engineering, sẽ giúp các developer debug và deploy agent hiệu quả hơn.

7

Google ra mắt LLM Gemma 4 với khả năng dự đoán multi-token

🇬🇧 *Google brings multi-token prediction Gemma 4 LLMs*

📰 BD Tech Talks 🔗 Đọc bài viết →

> Google đã nâng cấp đáng kể dòng Large Language Model (LLM) open-weight Gemma 4 của mình, tăng cường tốc độ inference của chúng. Cải tiến này đến từ việc tích hợp multi-token prediction (MTP) vào kiến trúc, phá vỡ cách tiếp cận truyền thống một token tại một thời điểm. Điều này cho phép model dự đoán nhiều token

8

GDS lên tiếng về việc NHS rút lui khỏi Open Source

🇬🇧 *GDS weighs in on the NHS's decision to retreat from Open Source*

📰 Simon Willison 🔗 Đọc bài viết →

> Dịch vụ Kỹ thuật Số của Chính phủ Anh (GDS) đã tham gia vào quyết định của Dịch vụ Y tế Quốc gia (NHS) về việc hạn chế truy cập vào các kho mã nguồn mở của họ. Trước đó, NHS đã đóng truy cập do các lỗ hổng bảo mật được báo cáo. GDS đã xuất bản một báo cáo, "AI, mã nguồn mở và rủi ro lỗ hổng trong lĩnh vực công", trong đó khuyến nghị rằng tính mở nên vẫn là tư thế mặc định. Điều này có nghĩa là thông tin nên được công khai trừ khi có lý do cụ thể để hạn chế nó. Báo cáo lập luận rằng việc làm mọi thứ riêng tư có thể tăng chi phí và giảm tái sử dụng và kiểm tra thông tin. Việc GDS thể hiện quan điểm này được coi là một sự leo thang đáng kể trong cuộc tranh luận, với một số người giải thích nó như một dấu hiệu của căng thẳng trong Dịch vụ Dân sự của Anh.

⚡ TIPS & TRICKS CHO DEV

⚡ Theo dõi hiệu suất

Vấn đề: Không thể theo dõi hiệu suất của mô hình AI trong thời gian thực.

Cách làm: Sử dụng LangSmith để theo dõi hiệu suất, ví dụ: `langsmith track --model my_model`.

Đánh giá: Hiệu quả trong việc theo dõi hiệu suất, nên dùng khi cần tối ưu hóa mô hình.

⚡ Kiểm soát chi phí

Vấn đề: Chi phí sử dụng mô hình AI vượt quá dự kiến.

Cách làm: Sử dụng Langfuse để kiểm soát chi phí, ví dụ: `langfuse cost --model`

`my_model --budget 100`.

Đánh giá: Giúp kiểm soát chi phí hiệu quả, nên dùng khi cần quản lý ngân sách.

⚡ Tối ưu hóa mô hình

Vấn đề: Mô hình AI không đạt hiệu suất như mong đợi.

Cách làm: Sử dụng Arize Phoenix để tối ưu hóa mô hình, ví dụ: `arize optimize --model my_model --metric accuracy`.

Đánh giá: Tối ưu hóa mô hình hiệu quả, nên dùng khi cần cải thiện hiệu suất.

📚 BÀI HỌC AI HÔM NAY CHO DEV

1. **Tích hợp AI API vào ứng dụng**

Dev cần biết cách tích hợp AI API vào ứng dụng để tận dụng sức mạnh của trí tuệ nhân tạo trong việc tự động hóa và cải thiện trải nghiệm người dùng. Việc tích hợp AI API có thể giúp dev tạo ra các ứng dụng thông minh hơn và phản hồi nhanh hơn.

3. Ví dụ thực tế: tích hợp API của Google Cloud Vision vào ứng dụng di động để nhận dạng hình ảnh và tự động gắn thẻ.

4. 💡 Tip hoặc bước tiếp theo: sử dụng thư viện như TensorFlow hoặc PyTorch để tích hợp AI API vào ứng dụng và đảm bảo tính bảo mật và hiệu suất cao.