BIS - MT · 07/06/2026

Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

“Nothing in life is to be feared, it is only to be understood.”

↪ Không có gì trong cuộc sống đáng sợ, chỉ cần được hiểu.

— Marie Curie

Hiểu biết là liều thuốc chữa nỗi sợ hãi — khi hiểu rõ bản chất của thử thách hay rủi ro, ta có thể đối mặt với chúng lý trí và hiệu quả hơn.

TIN TỨC NỔI BẬT

1

Các nhà nghiên cứu bảo mật phát tín hiệu cảnh báo về lỗ hổng trong mã được tạo bởi AI

Security Researchers Sound the Alarm on Vulnerabilities in AI-Generated Code

Infosecurity Magazine Đọc bài viết →

Các nhà nghiên cứu bảo mật đã bày tỏ mối quan ngại về các điểm yếu trong mã được tạo ra bởi AI, nhấn mạnh các rủi ro tiềm ẩn liên quan đến sự phụ thuộc ngày càng tăng vào trí tuệ nhân tạo trong phát triển phần mềm. Theo các nhà nghiên cứu, mã được tạo ra bởi AI có thể chứa các lỗi và điểm yếu mà các kẻ tấn công có thể khai thác, làm tổn hại đến bảo mật của hệ thống. Các nhà nghiên cứu chỉ ra rằng các model AI có thể tạo ra mã không chỉ kém hiệu quả mà còn dễ bị tấn công bởi các cuộc tấn công phổ biến như SQL injection và cross-site scripting. Hơn nữa, sự thiếu minh bạch và khả năng giải thích trong mã được tạo ra bởi AI làm cho việc xác định và sửa lỗi điểm yếu trở nên khó khăn. Các nhà nghiên cứu đang kêu gọi các developer thận trọng khi sử dụng mã được tạo ra bởi AI và phải thử nghiệm và xem xét kỹ lưỡng mã trước khi triển khai. Họ cũng nhấn mạnh sự cần thiết phải nghiên cứu thêm về các tác động bảo mật của mã được tạo ra bởi AI và phát triển các model AI mạnh mẽ và bảo mật hơn. Các phát hiện này nhấn mạnh tầm quan trọng của việc cân bằng giữa lợi ích của AI trong phát triển phần mềm với nhu cầu về các biện pháp bảo mật mạnh mẽ.

2 Lớp lệnh AI nhà kho đa agent cho phép đạt được sự xuất sắc trong hoạt động 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 phát triển một Lớp lệnh AI Nhà kho Đa tác nhân, một công nghệ tiên tiến được thiết kế để tối ưu hóa hoạt động nhà kho và nâng cao trí tuệ chuỗi cung ứng. Hệ thống đổi mới này tận dụng trí tuệ nhân tạo (AI) và học máy (ML) để tối ưu hóa quản lý nhà kho, cải thiện hiệu suất và năng suất. Lớp lệnh AI Nhà kho Đa tác nhân cho phép giám sát và kiểm soát thời gian thực các hoạt động nhà kho, cho phép đưa ra quyết định nhanh chóng và phản ứng với các điều kiện thay đổi. Bằng cách tích hợp AI và ML, hệ thống có thể phân tích lượng lớn dữ liệu từ các nguồn khác nhau, cung cấp thông tin chi tiết quý giá về hiệu suất chuỗi cung ứng. Công nghệ này có tiềm năng cách mạng hóa quản lý nhà kho, cho phép các doanh nghiệp đạt được sự xuất sắc về hoạt động và đưa ra quyết định dựa trên dữ liệu. Bằng cách tự động hóa các nhiệm vụ và tối ưu hóa quy trình làm việc, Lớp lệnh AI Nhà kho Đa tác nhân có thể giúp giảm chi phí, cải thiện sự hài lòng của khách hàng và tăng cường khả năng cạnh tranh trên thị trường. Khả năng phân tích nâng cao của hệ thống cũng cho phép các doanh nghiệp có được sự hiểu biết sâu sắc hơn về chuỗi cung ứng của họ, cho phép lập kế hoạch và thực hiện hiệu quả hơn.

3 Nhà máy agent: Kết nối các agent, 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 nhằm kết nối các tác nhân, ứng dụng và dữ liệu bằng cách sử dụng các tiêu chuẩn mở mới. Nền tảng này tận dụng Nền tảng Điện toán Đám mây (MCP) và Tiêu chuẩn Ứng dụng-sang-Ứng dụng (A2A) của Microsoft để tạo điều kiện cho sự tương tác liền mạch giữa các hệ thống khác nhau. Với Agent Factory, các nhà phát triển có thể tạo và triển khai các tác nhân thông minh có thể giao tiếp với các ứng dụng và nguồn dữ liệu khác nhau. Điều này cho phép tạo ra các giải pháp tinh vi và tích hợp hơn có thể tự động hóa các nhiệm vụ, cung cấp thông tin theo thời gian thực và nâng cao năng suất tổng thể. Các tiêu chuẩn mở của nền tảng

cho phép linh hoạt và khả năng tương tác cao hơn, giúp các nhà phát triển xây dựng và triển khai các giải pháp tùy chỉnh dễ dàng hơn. Bằng cách kết nối các tác nhân, ứng dụng và dữ liệu theo cách hiệu quả và tiêu chuẩn hóa hơn, Agent Factory có tiềm năng cách mạng hóa cách thức hoạt động và tương tác của các doanh nghiệp với khách hàng. Nền tảng Azure của Microsoft có khả năng đóng vai trò quan trọng trong việc hỗ trợ phát triển và triển khai các giải pháp Agent Factory.

4

Khi Claude thay đổi, mọi thứ thay đổi: Quản lý phạm vi ảnh hưởng của AI trong sản xuất

When Claude changed, everything changed: Managing AI blast radius in production

VentureBeat Đọc bài viết →

Một công ty đã xây dựng một hệ thống được hỗ trợ bởi AI, dịch các câu hỏi ngôn ngữ tự nhiên thành các cuộc gọi API, giúp người dùng dễ dàng truy cập dữ liệu từ nhiều nguồn khác nhau. Hệ thống này đã thành công, tạo ra hàng trăm báo cáo mỗi tháng, nhưng việc nâng cấp mô hình từ Claude Sonnet 4.0 lên 4.5 đã gây ra các vấn đề không lường trước. Mô hình mới bắt đầu gộp các tham số yêu cầu vào trường mô tả, dẫn đến các cuộc gọi API không chính xác và làm rõ các câu hỏi trong phản hồi. Sự thay đổi này đã có tác động đáng kể đến hệ thống, gây ra các lỗi ở hạ□ và yêu cầu quay lại phiên bản mô hình trước đó. Sự cố này làm nổi bật những thách thức khi làm việc với các Large Language Models (LLMs) trong môi trường sản xuất. Không giống như kỹ thuật phần mềm truyền thống, nơi các thay đổi có thể được dự đoán và giới hạn, LLMs có "vùng ảnh hưởng vô hạn" do không gian đầu vào không giới hạn và các chế độ thất bại tiềm năng. Công ty đã học được rằng lời nhắc của họ không được chỉ định đầy đủ và rằng các phiên bản mô hình trước đó đã suy luận các ràng buộc mà mô hình mới không tuân theo. Để giảm thiểu vấn đề này, các tác giả đề xuất xử lý bộ đánh giá như là thông số kỹ thuật chính thức của hệ thống, thay vì lời nhắc. Phương pháp này bao gồm việc viết các đánh giá để chỉ định hành vi dự kiến của hệ thống, bao gồm các thuộc tính và hàm tính điểm. Bằng cách xử lý các nâng cấp mô hình và thay đổi lời nhắc như các yêu cầu kéo mà phải vượt qua bộ đánh giá, các nhóm có thể giới hạn vùng ảnh hưởng của các thay đổi và đảm bảo rằng hệ thống hoạt động như mong đợi trong sản xuất.

5

OpenAI ra mắt Chế độ khóa để bảo vệ dữ liệu nhạy cảm khỏi các cuộc tấn công tiêm lệnh

OpenAI unveils Lockdown Mode to protect sensitive data from prompt injection attacks

TechCrunch AI Đọc bài viết →

> OpenAI đã giới thiệu Chế độ khóa (Lockdown Mode), một tính năng mới được thiết kế để bảo vệ dữ liệu nhạy cảm khỏi các cuộc tấn công tiêm lệnh (prompt injection attacks). Những cuộc tấn công này liên quan đến các lệnh trò chuyện độc hại ẩn trong nội dung web có thể làm suy yếu bảo mật của các trò chuyện bot như ChatGPT. Chế độ khóa vô hiệu hóa việc duyệt web trực tiếp, lấy hình ảnh từ web, nghiên cứu sâu và chế độ đại lý, giảm thiểu rủi ro chia sẻ dữ liệu nhạy cảm. Tuy nhiên, OpenAI lưu ý rằng ngay cả khi Chế độ khóa được bật, ChatGPT vẫn có thể dễ bị tấn công bởi các lệnh tiêm. Tính năng này được dành cho các cá nhân và tổ chức xử lý dữ liệu nhạy cảm yêu cầu bảo vệ nghiêm ngặt hơn khỏi rủi ro mất dữ liệu. Chế độ khóa đang được triển khai cho các tài khoản ChatGPT Business tự phục vụ và các tài khoản cá nhân đủ điều kiện.

6

Endava đang thiết kế lại việc giao hàng phần mềm xung quanh các agent AI

How Endava is redesigning software delivery around AI agents

OpenAI Blog Đọc bài viết →

> Endava, một công ty công nghệ, đang trải qua một sự chuyển đổi đáng kể trong quy trình giao hàng phần mềm của mình bằng cách tích hợp các tác nhân AI. Công ty đang tận dụng khả năng của ChatGPT Enterprise và Codex để thúc đẩy hiệu quả và đổi mới. Bằng cách sử dụng những công nghệ này, Endava nhằm mục đích tăng tốc giao hàng phần mềm và tự động hóa các quy trình làm việc khác nhau. Sự chuyển đổi này hướng tới các quy trình được thúc đẩy bởi AI cũng dự kiến sẽ thúc đẩy một nền văn hóa AI bản địa trong toàn tổ chức.

7

Tại sao tương lai của AI agent lại xoay quanh việc kiểm soát

Why the future of agentic AI is all about the harness

BD Tech Talks Đọc bài viết →

Tương lai của AI có khả năng hành động, cho phép máy móc thực hiện hành động và đưa ra quyết định, đang chuyển dịch khỏi việc xây dựng các model lớn hơn và cung cấp cho chúng nhiều dữ liệu hơn. Theo một bài báo mới từ UC Berkeley, điểm nghẽn chính tiếp theo trong AI có khả năng hành động là "system scaling", hay nói cách khác là mở rộng "harness" giúp chuyển đổi câu trả lời của model thành hành vi trong thế giới thực. Harness này là một thành phần quan trọng của hệ thống AI, bao gồm sáu thành phần tương tác: bộ nhớ, constructor ngữ cảnh, lớp định tuyến kỹ năng, vòng lặp điều phối, lớp xác minh và quản trị, và hơn thế nữa. Các hệ thống AI khác nhau, chẳng hạn như Claude Code và OpenClaw, có các thành phần tương tự nhưng ưu tiên các khía cạnh khác nhau, chẳng hạn như độ tin cậy hoặc khả năng tái tạo. Các hệ thống AI có khả năng hành động hoạt động trên nhiều thang thời gian khác nhau, với các lệnh xác định mục tiêu ngay lập tức, kỹ năng thực hiện các nhiệm vụ cụ thể, và bộ nhớ lưu giữ các sự kiện theo thời gian. Để tối ưu hóa hiệu suất, điều quan trọng là phải đánh giá toàn bộ hệ thống, bao gồm cả khung sườn của nó, chứ không chỉ là model riêng lẻ.

8

Năm phòng thí nghiệm, năm tâm trí: Xây dựng một vở kịch tài chính đa model trên các model nhỏ

Five labs, five minds: building a multi-model finance drama on small models

Hugging Face Blog Đọc bài viết →

Một bản cập nhật gần đây cho trò chơi "Thousand Token Wood" đã biến nó từ một trải nghiệm quan sát thụ động thành một trò chơi tương tác nơi người chơi đóng vai trò của một nhà tài chính, được gọi là Patron của Rừng. Trò chơi có năm sinh vật rừng, mỗi sinh vật được trang bị một mô hình nhỏ khác nhau từ các phòng thí nghiệm khác nhau, bao gồm OpenAI, NVIDIA và mô hình tinh chỉnh của chính tác giả. Những mô hình này được đào tạo trên các dữ liệu khác nhau và có các quá trình hậu đào tạo khác nhau, dẫn đến các hành vi và quá trình ra quyết định thực sự khác nhau. Cơ chế cốt lõi của trò chơi là mẹo nội bộ, nơi người chơi có thể cung cấp cho sinh vật thông tin đúng hoặc sai, ảnh hưởng đến kết quả của trò chơi. Để duy trì bảo mật của trò chơi và ngăn chặn sinh vật truy cập thông tin nhạy cảm, một lớp phân tích và sửa chữa JSON dung thứ đã được thực hiện để xử lý các khác biệt về tokenizer và thói quen định dạng. Trò chơi cũng có các mối quan hệ liên tục giữa các sinh vật, bị ảnh hưởng bởi các sự kiện và hành động của người chơi. Tuy nhiên, các nhà phát triển trò chơi đã

> lưu ý đến tầm quan trọng của lạm phát prompt, nơi lịch sử thô tăng không giới hạn, và đã thực hiện một giải pháp để giới hạn lịch sử trong prompt, cho phép trò chơi chạy mượt mà với các mô hình nhỏ.

TIPS & TRICKS CHO DEV

Sử dụng LangSmith

Vấn đề: Không thể theo dõi hiệu suất của mô hình AI.
Cách làm: Sử dụng LangSmith để theo dõi và phân tích dữ liệu. Ví dụ, sử dụng lệnh `langsmith track` để theo dõi hiệu suất mô hình.
Đánh giá: Hiệu quả khi cần theo dõi hiệu suất mô hình AI, nên dùng khi triển khai mô hình trong sản xuất.

Tích hợp Langfuse

Vấn đề: Không thể tích hợp mô hình AI với các công cụ khác.
Cách làm: Sử dụng Langfuse để tích hợp mô hình AI với các công cụ khác. Ví dụ, sử dụng lệnh `langfuse integrate` để tích hợp mô hình với các dịch vụ khác.
Đánh giá: Hiệu quả khi cần tích hợp mô hình AI, nên dùng khi cần kết nối với các công cụ khác.

Triển khai Arize Phoenix

Vấn đề: Không thể kiểm soát chi phí của mô hình AI.
Cách làm: Sử dụng Arize Phoenix để theo dõi và kiểm soát chi phí. Ví dụ, sử dụng lệnh `arize phoenix optimize` để tối ưu hóa chi phí mô hình.
Đánh giá: Hiệu quả khi cần kiểm soát chi phí, nên dùng khi cần tối ưu hóa chi phí vận hành mô hình AI.

BÀI HỌC AI HÔM NAY CHO DEV

1. **Tích hợp AI API vào ứng dụng**
2. Tích hợp AI API vào ứng dụng là một chủ đề quan trọng giúp các nhà phát triển tận dụng khả năng của trí tuệ nhân tạo trong các dự án của mình. Điều này cho phép họ tạo ra các ứng dụng thông minh hơn, tự động hóa các tác vụ và cải thiện trải nghiệm người dùng. Việc tích hợp AI API cũng giúp giảm thiểu thời gian và công sức phát triển.
3. Ví dụ, một ứng dụng có thể sử dụng API của Google Cloud Vision để phân tích hình ảnh và nhận diện đối tượng, từ đó cung cấp thông tin chi tiết cho người dùng.
4. Tip hoặc bước tiếp theo: Để bắt đầu tích hợp AI API vào ứng dụng, hãy xác định rõ yêu cầu của dự án và lựa chọn API phù hợp, sau đó tham khảo tài liệu và các ví dụ code để thực hiện tích hợp một cách hiệu quả.

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

Nguồn: Google News · Groq AI