BIS - MT · 29/06/2026

# Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

*"Gần mực thì đen, gần đèn thì rạng."*

— Tục ngữ Việt Nam

*Môi trường và người xung quanh ảnh hưởng lớn đến nhân cách — hãy chọn bạn tốt, chọn môi trường tích cực.*

## TIN TỨC NỔI BẬT

1

### AWS Mở Mã Nguồn Server MCP cho Bedrock AgentCore để Tối Ưu 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 một máy chủ MCP (Nền tảng 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 các ứng dụng như robot, xe tự hành và nhiều hơn nữa. Máy chủ MCP được thiết kế để hoạt động mượt mà với Bedrock AgentCore, một framework để xây dựng các tác nhân AI. Bằng cách mở nguồn máy chủ MCP, AWS cung cấp cho các nhà phát triển một nền tảng tiêu chuẩn hóa để xây dựng, triển khai và quản lý các tác nhân AI trên nhiều môi trường đám mây. Điều này có thể giúp giảm thiểu sự phức tạp và chi phí liên quan đến việc phát triển và triển khai các tác nhân AI. Việc mở nguồn máy chủ MCP dự kiến sẽ đẩy nhanh sự phát triển và áp dụng các tác nhân AI trên nhiều ngành công nghiệp. Nó cũng thể hiện cam kết của AWS trong việc làm cho việc phát triển AI trở nên dễ tiếp cận và hiệu quả hơn cho các nhà phát triển.

2

### Tối Ưu Hóa Workflow 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 (Nền tảng Canvas Mô hình), cho phép các nhà phát triển tận dụng sức mạnh của AI để tự động hóa và tối ưu hóa các quy trình làm việc trên GitHub của họ. Với tích hợp này, các nhà phát triển có thể sử dụng Amazon Bedrock để tạo và quản lý các model, và sau đó triển khai chúng lên MCP để thực thi. Điều này cho phép sử dụng AI tạo sinh để tự động hóa các nhiệm vụ như xem xét mã, phát hiện lỗi và thậm chí là tạo mã. Tích hợp này nhằm mục đích đơn giản hóa quá trình tích hợp AI vào các quy trình làm việc trên GitHub, giúp các nhà phát triển dễ dàng áp dụng và tận dụng các công nghệ AI mới nhất. Bằng cách tự động hóa các nhiệm vụ lặp đi lặp lại và cải thiện chất lượng mã, các nhà phát triển có thể tập trung vào các nhiệm vụ cấp cao hơn và giao tiếp dự án một cách hiệu quả hơn. Tích hợp này dự kiến sẽ nâng cao trải nghiệm của nhà phát triển trên GitHub và đẩy nhanh việc áp dụng AI trong phát triển phần mềm.

3

## Agent Factory: Kết Nối Các Agent, App và Data với Các Open Standard 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 sáng kiến mới nhằm kết nối các tác nhân, ứng dụng và dữ liệu bằng cách sử dụng các tiêu chuẩn mở. Sáng kiến này tận dụng hai tiêu chuẩn mở mới: MCP (Microsoft Cloud Protocol) và A2A (Application to Application). Agent Factory cho phép giao tiếp và tích hợp liền mạch giữa các hệ thống, ứng dụng và dịch vụ khác nhau. Với Agent Factory, các nhà phát triển có thể xây dựng và triển khai các tác nhân có thể tương tác với các ứng dụng và nguồn dữ liệu khác nhau, bất kể công nghệ hoặc vị trí cơ bản của chúng. Điều này tạo điều kiện cho sự linh hoạt và khả năng tương tác cao hơn, cho phép các tổ chức tạo ra các quy trình làm việc hiệu quả và hiệu quả hơn. Việc giới thiệu các tiêu chuẩn MCP và A2A dự kiến sẽ đơn giản hóa quá trình tích hợp các hệ thống và ứng dụng khác nhau, giảm độ phức tạp và chi phí liên quan đến trao đổi và tích hợp dữ liệu. Bằng cách cung cấp một khuôn khổ chung cho giao tiếp và trao đổi dữ liệu, Agent Factory có tiềm năng mở ra những khả năng mới cho sự

đổi mới và hợp tác trên các ngành và lĩnh vực khác nhau, đặc biệt là trong lĩnh vực AI, API, LLM, model, token, và framework.

## 4 Prompt Injection Khai Thác Các Lỗ Hổng Thiết Kế Lớn Nhất Của Enterprise AI Bằng Cách Nhắm Mục Tiêu Vào Agent, RAG Pipeline và Model Router

*Prompt injection is exploiting enterprise AI's biggest design flaws by targeting agents, RAG pipelines and model routers*

VentureBeat Đọc bài viết →

Các doanh nghiệp đã ngày càng áp dụng các mô hình ngôn ngữ lớn (LLM) cho các nhiệm vụ khác nhau, nhưng điều này cũng đã dẫn đến một mối đe dọa ngày càng tăng của các cuộc tấn công tiêm lệnh (prompt injection attacks). Những cuộc tấn công này khai thác các điểm yếu thiết kế của LLM, vốn gặp khó khăn trong việc phân biệt giữa các lệnh và dữ liệu. Kết quả là, các kẻ tấn công có thể thao túng LLM bằng cách tạo ra các lệnh độc hại, dẫn đến hệ thống và dữ liệu bị xâm phạm. Trong năm 2025 và 2026, một số báo cáo đã nhấn mạnh mức độ nghiêm trọng của các cuộc tấn công tiêm lệnh, với danh sách OWASP LLM Top 10 xếp nó là điểm yếu nghiêm trọng nhất. Báo cáo Đe dọa Toàn cầu của CrowdStrike đã ghi nhận hơn 90 tổ chức bị ảnh hưởng bởi các cuộc tấn công tiêm lệnh, được sử dụng để đánh cắp thông tin đăng nhập và tiền điện tử. Các sự kiện thực tế, chẳng hạn như một điểm yếu trong Slack AI và một cuộc khai thác zero-click chống lại Microsoft 365 Copilot, chứng minh tác động hoạt động của các cuộc tấn công tiêm lệnh. Bề mặt tấn công đã mở rộng để bao gồm kiến trúc đa tác nhân, đường ống tạo ra tăng cường bằng cách thu thập, bộ định tuyến mô hình và khả năng bộ nhớ dài hạn. Để giải quyết mối đe dọa này, các doanh nghiệp phải thay đổi tư duy và coi LLM như các trình thông dịch không đáng tin cậy. Điều này bao gồm việc hạn chế quyền của mô hình, phân đoạn nội dung không đáng tin cậy, giám sát việc gọi công cụ, xác thực nguồn gốc nội dung, tăng cường bảo mật cho bộ định tuyến mô hình và coi LLM như các thành phần không đáng tin cậy.

## 5 HP Inc. Khởi Động Đối Tác Chiến Lược Frontier với OpenAI

*HP Inc. launches Frontier strategic partnership with OpenAI*

OpenAI Blog Đọc bài viết →

> HP Inc. đã công bố việc mở rộng quan hệ đối tác chiến lược với OpenAI, một công ty công nghệ trí tuệ nhân tạo (AI) hàng đầu. Đối tác này, được gọi là Frontier, nhằm tích hợp các khả năng AI trên nhiều khía cạnh của kinh doanh HP. Điều này bao gồm nâng cao trải nghiệm khách hàng, cải thiện phát triển phần mềm và tối ưu hóa hoạt động doanh nghiệp. Thông qua quan hệ đối tác này, HP Inc. tìm cách tận dụng chuyên môn của OpenAI trong lĩnh vực AI để thúc đẩy đổi mới và hiệu quả trên toàn bộ hoạt động. Việc triển khai các công nghệ AI dự kiến sẽ có tác động đáng kể đến nhiều lĩnh vực của công ty, cho phép nó cung cấp dịch vụ cá nhân hóa và phản hồi hơn cho khách hàng, tối ưu hóa quy trình phát triển phần mềm và cải thiện hiệu quả hoạt động tổng thể. Quan hệ đối tác mở rộng này là một bước phát triển quan trọng cho HP Inc., vì nó tìm cách tận dụng sức mạnh của AI để thúc đẩy tăng trưởng và cạnh tranh trên thị trường. Với các công nghệ AI tiên tiến của OpenAI, HP Inc. được đặt vào vị trí thuận lợi để đi đầu và mang lại giá trị vượt trội cho khách hàng và các bên liên quan của mình.

## 6 Chiến Dịch Toàn Cầu Giáng Đòn Kép Phá Vỡ "Dây Chuyền Lắp Ráp" Tội Phạm Mạng

*One-two punch delivered in global operation disrupts cybercrime "assembly line"*

Ars Technica Đọc bài viết →

> Trong một hoạt động lớn, các cơ quan chức năng quốc tế và các công ty công nghệ tư nhân đã phá vỡ một "dây chuyền lắp ráp" tội phạm mạng cho phép các hacker thu thập hàng triệu thông tin đăng nhập và đánh cắp hơn 47 triệu đô la. Được đặt tên là "Operation Endgame", việc nhắm mục tiêu đồng thời vào hai công cụ tội phạm được sử dụng rộng rãi, Amadey và StealC, đã làm tê liệt nghiêm trọng mạng lưới phân phối malware. Amadey là một nền tảng malware-as-a-service được sử dụng để xâm phạm thiết bị và phân phối payload độc hại, trong khi StealC là một nền tảng infostealer-as-a-service thu thập thông tin nhạy cảm. Các công cụ, đã được sử dụng từ ít nhất năm 2018, phụ thuộc vào cùng một cơ sở hạ tầng cơ bản, cho phép Microsoft tìm kiếm một lệnh gián đoạn chung. Hoạt động này đã dẫn đến việc gián đoạn hơn 200 máy chủ command-and-control, cắt đứt quyền kiểm soát hơn 18.000 máy tính bị nhiễm và thu hồi 27 triệu thông tin đăng nhập bị đánh cắp. Các công ty khác, bao gồm ESET, Proofpoint và IBM X-Force, đã hỗ trợ trong hoạt động này, bao gồm các quốc gia như Canada, Đan Mạch, Đức, Hà Lan, Anh và Mỹ.

## 7 Cách Mạng Năng Lượng Tái Tạo: Common Energy Đang Định Hình Lại Bức Tranh Năng Lượng Toàn Cầu Như Thế Nào

*Renewable Energy Revolution: How Common Energy is Reshaping the Global Energy Landscape*

Dev.to AI Đọc bài viết →

> Thế giới đang ở một ngã rẽ quan trọng, với nhu cầu về các giải pháp năng lượng bền vững ngày càng trở nên cấp thiết do biến đổi khí hậu. Hệ thống năng lượng hiện tại, dựa trên nhiên liệu hóa thạch, đang gây hại cho môi trường và□□ sự bất bình đẳng xã hội và kinh tế. Để giải quyết vấn đề này, Common Energy đang dẫn đầu trong cuộc cách mạng năng lượng tái tạo. Tổ chức này có sứ mệnh thúc đẩy các dự án năng lượng mặt trời cộng đồng, thay thế nhiên liệu hóa thạch bằng năng lượng sạch địa phương, dẫn đến giảm lượng khí thải carbon và giảm chi phí năng lượng cho khách hàng. Tầm nhìn của Common Energy là tạo ra một thế giới nơi năng lượng là một lực lượng tốt, chứ không phải là nguồn ô nhiễm và phá hủy. Bằng cách trao quyền cho các cộng đồng kiểm soát nhu cầu năng lượng của mình, tổ chức này

đang giảm sự phụ thuộc vào nhiên liệu hóa thạch và tạo ra một tương lai sạch hơn, bền vững hơn.

## 8 Sóng Nhiệt Gây Rối Loạn Não Bộ. Các Nhà Khoa Học Đang Tìm Hiểu Lý Do.

*Heat waves mess with your brain. Scientists are trying to figure out why.*

MIT Tech Review  Đọc bài viết →

Một đợt sóng nhiệt khắc nghiệt đã ập đến Tây Âu, với Vương quốc Anh ghi nhận nhiệt độ tháng 6 cao nhất từ trước đến nay là 36,1 °C (97 °F). Nhiệt độ cực đoan không chỉ gây hại về thể chất mà còn ảnh hưởng đến sức khỏe tâm lý và chức năng não. Nghiên cứu cho thấy trẻ em và cá nhân mắc rối loạn sức khỏe tâm lý đặc biệt dễ bị tổn thương bởi tác động của sóng nhiệt. Các nghiên cứu đã chỉ ra rằng khi nhiệt độ tăng, mọi người trở nên dễ cáu gắt và bạo lực hơn, và những người mắc chứng rối loạn sức khỏe tâm lý dễ bị tổn thương bởi tác động vật lý của nhiệt. Một nghiên cứu gần đây cho thấy có sự tăng 9,7% trong số lượng nhập viện của những người mắc chứng rối loạn sức khỏe tâm lý trong thời kỳ sóng nhiệt. Các nhà nghiên cứu vẫn đang cố gắng hiểu các cơ chế đằng sau những tác động này, nhưng các nghiên cứu ban đầu cho thấy rằng việc tiếp xúc với nhiệt có thể làm suy giảm kỹ năng nhận thức, chẳng hạn như tập trung và chú ý, và những tác động này có thể kéo dài trong một thời gian dài. Cần có thêm nghiên cứu để xác định tác động lâu dài của sóng nhiệt đối với sức khỏe tâm lý và để phát triển các chiến lược bảo vệ dân số dễ bị tổn thương.

TIPS & TRICKS CHO DEV

### Tối ưu hóa mã với GitHub Copilot

**Vấn đề:** Mã nguồn chưa được tối ưu hóa khiến hiệu suất thấp.

**Cách làm:** Sử dụng GitHub Copilot để đề xuất cải tiến, nhập `/ TODO: optimize /` và Copilot sẽ gợi ý cải thiện mã.

**Đánh giá:** Hiệu quả cao trong việc tối ưu hóa mã, nên dùng khi muốn cải thiện hiệu suất.

### Tự động hoàn thành mã với Aider

**Vấn đề:** Nhập mã nguồn tốn nhiều thời gian.

**Cách làm:** Sử dụng Aider để tự động hoàn thành mã, nhập `print(` và Aider sẽ đề xuất hoàn thành câu lệnh.

**Đánh giá:** Tiết kiệm thời gian, nên dùng khi cần nhập mã nhanh chóng.

### Debug mã với Continue.dev

**Vấn đề:** Debug mã nguồn tốn nhiều công sức.

**Cách làm:** Sử dụng Continue.dev để phân tích và debug mã, nhập `console.log(` và Continue.dev sẽ giúp tìm ra vấn đề.

**Đánh giá:** Hiệu quả trong việc debug, nên dùng khi cần tìm ra nguyên nhân lỗi.

---

BÀI HỌC AI HÔM NAY CHO DEV

1. **Tối ưu chi phí & hiệu năng LLM**
2. Để tối ưu hóa chi phí và hiệu năng của mô hình ngôn ngữ lớn (LLM), các nhà phát triển cần biết cách tinh chỉnh và tối ưu hóa mô hình cho từng trường hợp sử dụng cụ thể. Điều này giúp giảm thiểu chi phí tính toán và tăng tốc độ xử lý. Việc tối ưu hóa hiệu năng LLM cũng giúp cải thiện hiệu suất của ứng dụng.
3. Ví dụ, chúng ta có thể sử dụng kỹ thuật fine-tuning và LoRA (Low-Rank Adaptation) để tinh chỉnh mô hình LLM cho từng nhiệm vụ cụ thể, 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 bắt đầu bằng cách đánh giá hiệu suất của mô hình LLM hiện tại và xác định các điểm cần cải thiện, sau đó áp dụng các kỹ thuật tinh chỉnh và tối ưu hóa để đạt được hiệu suất tốt nhất.