BIS - MT · 11/07/2026

Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

> *"Nhà sạch thì mát, bát sạch ngon cơm."*
>
> — Tục ngữ Việt Nam

Trật tự và sự ngăn nắp tạo ra môi trường để làm việc hiệu quả và cuộc sống chất lượng hơn.

TIN TỨC NỔI BẬT

1 Các nhà nghiên cứu bảo mật cảnh báo về lỗ hổng trong code do AI tạo ra

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 lo ngại về các lỗ hổng trong code do AI tạo ra, loại code đang ngày càng được sử dụng nhiều trong phát triển phần mềm. Các nhà nghiên cứu cảnh báo rằng code do AI tạo ra có thể chứa các lỗi bảo mật, biến nó thành mối đe dọa tiềm tàng đối với các hệ thống và dữ liệu. Những lỗ hổng này có thể phát sinh từ các thuật toán AI được sử dụng để tạo code, vốn có thể không có khả năng xác định hoặc giảm thiểu các rủi ro bảo mật tiềm ẩn. Hơn nữa, việc phụ thuộc vào code do AI tạo ra có thể dẫn đến thiếu minh bạch và hiểu biết về code nền tảng, gây khó khăn trong việc xác định và khắc phục các vấn đề bảo mật. Các nhà nghiên cứu nhấn mạnh rằng code do AI tạo ra không nên được coi là sự thay thế cho code do con người viết, mà là một công cụ để bổ trợ và hỗ trợ các developer. Họ khuyến nghị các developer nên xem xét và kiểm tra kỹ lưỡng code do AI tạo ra trước khi triển khai nó trong môi trường production để đảm bảo mức độ bảo mật cao nhất.

2 AWS open-source MCP server cho Bedrock AgentCore nhằm tinh gọn 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 open-source một MCP (Multi-Cloud Platform) server cho Bedrock AgentCore, một framework để phát triển các AI agent. Động thái này nhằm mục đích tinh gọn quá trình phát triển AI agent bằng cách cung cấp một platform tiêu chuẩn hóa để xây dựng, kiểm thử và deploy các AI model trên nhiều môi trường cloud khác nhau. MCP server được thiết kế để hoạt động liền mạch với Bedrock AgentCore, cho phép các developer tạo và quản lý AI agent dễ dàng và linh hoạt hơn. Bằng cách open-source MCP server, AWS đang tạo điều kiện cho cộng đồng developer rộng lớn hơn đóng góp và hưởng lợi từ công nghệ này. Việc open-source MCP server được kỳ vọng sẽ thúc đẩy nhanh quá trình phát triển AI agent và khuyến khích sự hợp tác giữa các developer. Nó cũng nhấn mạnh cam kết của AWS trong việc thúc đẩy nghiên cứu và phát triển AI, đồng thời thúc đẩy khả năng tương tác (interoperability) và tính di động (portability) trên các cloud platform khác nhau. Động thái này được coi là một bước tiến quan trọng hướng tới việc làm cho quá trình phát triển AI trở nên dễ tiếp cận và hiệu quả hơn, và có khả năng tác động tích cực đến toàn bộ hệ sinh thái phát triển AI.

3

Các model open-weight tốt nhất của Trung Quốc — và những đối thủ mạnh nhất từ Mỹ

The best Chinese open-weight models — and the strongest US rivals

understandingai.org Đọc bài viết →

Bài viết thảo luận về các model open-weight hàng đầu của Trung Quốc và những đối thủ mạnh nhất từ Mỹ. Các model open-weight đề cập đến những phương tiện điện được thiết kế nhẹ, thường với cách tiếp cận tối giản về thiết kế nội thất và ngoại thất. Các nhà sản xuất Trung Quốc đã đi đầu trong xu hướng này, cung cấp các lựa chọn giá cả phải chăng và hiệu quả. Một số model open-weight hàng đầu của Trung Quốc bao gồm BYD Tang, Geely Geometry A và XPeng P5. Những phương tiện này ưu tiên hiệu suất, phạm vi hoạt động và khả năng chi trả, khiến chúng trở thành lựa chọn hấp dẫn cho người tiêu dùng. Ngược lại, các nhà sản xuất Mỹ đã chậm hơn trong việc áp dụng xu hướng model open-weight. Tuy nhiên, một số đối thủ đáng chú ý của Mỹ bao gồm Tesla Model 3, vốn là người tiên phong trên thị trường xe điện, và Rivian R1T, một chiếc xe bán tải mang đến sự kết hợp độc đáo giữa hiệu suất và tính năng. Bài viết nhấn mạnh sự cạnh tranh ngày càng tăng trên thị trường xe điện, với các nhà sản xuất Trung Quốc đang đẩy mạnh giới hạn về đổi mới và khả năng chi trả. Khi thị trường

tiếp tục phát triển, sẽ rất thú vị để xem các nhà sản xuất Mỹ sẽ phản ứng thế nào trước thách thức từ các đối thủ Trung Quốc của họ.

4

Tự động hóa với tốc độ của Swamp

Automation at the speed of Swamp

Changelog Đọc bài viết →

Trong một cuộc phỏng vấn gần đây, Adam Jacob, founder của System Initiative và là người tạo ra Swamp, đã thảo luận về tác động của các AI agent đối với phát triển phần mềm. Với Swamp, một platform tự động hóa phát triển phần mềm, đội ngũ 18 người của Adam đã giảm xuống còn 5 người, nhưng họ vẫn có thể ship platform 900 lần chỉ trong bốn tuần. Ông cho rằng thành công này là nhờ sự trở lại của User Acceptance Testing (UAT) và tầm quan trọng ngày càng tăng của software architecture và domain-driven design. Một buổi demo trực tiếp đã trình diễn khả năng của Swamp trong việc tự viết automation, để lại ấn tượng sâu sắc. Adam cũng nhấn mạnh tầm quan trọng của security, dẫn chứng việc ông sử dụng Tailscale để truy cập tài nguyên dễ dàng và an toàn. Ngoài ra, ông còn nêu bật sự cần thiết của một CI/CD platform như RWX, cho phép các agent viết code trong vài phút và validate ngay lập tức.

5

Công cụ tốt hơn lại khiến code review của Copilot tệ hơn. Đây là cách chúng tôi thực sự cải thiện nó.

Better tools made Copilot code review worse. Here's how we actually improved it.

GitHub Blog Đọc bài viết →

GitHub gần đây đã cải thiện quy trình code review của Copilot. Ban đầu, team kỳ vọng một bản nâng cấp "sạch" khi họ thay thế các công cụ chia sẻ, được bảo trì tốt hơn để khám phá code. Tuy nhiên, các benchmark đã tiết

6

The Download: Cơ chế hoạt động bên trong của Claude và "super app" của OpenAI

The Download: Claude's inner workings and OpenAI's "super app"

MIT Tech Review Đọc bài viết →

Trong phiên bản ngày hôm nay của The Download, các nhà nghiên cứu tại Anthropic đã có một khám phá đột phá về hoạt động bên trong của các mô hình ngôn ngữ lớn (LLM). Họ đã phát triển một công cụ gọi là Jacobian lens, công cụ này đã tiết lộ một khu vực ẩn trong mô hình chủ chốt của họ, Claude, được gọi là J-space. J-space chứa các từ liên quan đến phản hồi mà mô hình đang làm việc nhưng có thể không sản xuất cuối cùng, cung cấp một cái nhìn về quá trình suy nghĩ của mô hình. Trong khi đó, OpenAI đã ra mắt "siêu ứng dụng" của mình, ChatGPT Work, kết hợp trình chatbot, công cụ mã hóa và các mô hình mới để hỗ trợ các nhiệm vụ làm việc. Ứng dụng này được thiết kế để làm việc cùng người dùng, giúp các nhiệm vụ trở nên hiệu quả hơn. Ngoài ra, OpenAI đã phát hành các mô hình GPT 5.6 và đang phát triển một nhà nghiên cứu tự động hoàn toàn. Các câu chuyện đáng chú ý khác bao gồm humanoid thực hiện phẫu thuật từ xa trên động vật sống, danh sách phá kỷ lục của SK Hynix tại Mỹ và thỏa thuận của Tencent để hủy bỏ việc Meta mua lại Manus. Meta cũng đã bắt đầu thu phí truy cập AI và dự định sản xuất một chip AI.

7 Bản patch cho lỗ hổng 0-day của Windows Defender có thể cho phép kẻ tấn công làm đầy ổ cứng

Patch for Windows Defender 0-day could allow attackers to fill hard disk

Ars Technica Đọc bài viết →

Một bản vá gần đây được Microsoft phát hành để sửa lỗi zero-day trong bộ máy bảo mật Windows Defender có thể có những hậu quả không mong muốn. Bản vá, được thiết kế để giải quyết lỗ hổng RoguePlanet (CVE-2026-50656), có thể cho phép kẻ tấn công làm đầy đĩa cứng bằng cách ghi một lượng lớn dữ liệu vào đó. Điều này là do một hành vi được giới thiệu bởi bản vá "cập nhật bảo vệ theo chiều sâu" gây ra bộ máy bảo vệ malware của Microsoft rò rỉ dữ liệu khi cố gắng mở một tệp. Vấn đề này liên quan đến cách bộ máy xử lý tệp siêu dữ liệu Zone.Identifier, được sử dụng để đánh dấu nguồn gốc và khu vực bảo mật của tệp được tải xuống từ internet. Một tác nhân độc hại có thể khai thác hành vi này bằng cách sử dụng Server Message Block (SMB) để kích hoạt bộ máy ghi tệp lớn vào đĩa, có khả năng gây ra hệ thống hết dung lượng. Microsoft chưa xác nhận sự tồn tại của hành vi này, nhưng nhà nghiên cứu đã phát hiện ra lỗ hổng, NightmareEclipse, đã cung cấp thông tin chi tiết về cách khai thác nó.

8 IEEE tưởng nhớ nhà khoa học máy tính tiên phong Peter G. Neumann

IEEE Remembers Pioneering Computer Scientist Peter G. Neumann

IEEE Spectrum Đọc bài viết →

> Peter G. Neumann, một nhà khoa học máy tính nổi tiếng và nhà phân tích rủi ro được kính trọng, đã qua đời vào ngày 17 tháng 5 tại tuổi 93. Neumann là một người tiên phong trong lĩnh vực máy tính, dành gần 70 năm để làm việc về rủi ro, độ tin cậy của hệ thống, bảo mật và khả năng chịu lỗi. Ông đã dành năm thập kỷ làm nhà khoa học chính tại SRI International ở California, nơi ông làm việc cho đến khi qua đời. Neumann được biết đến với khả năng xác định các điểm yếu hệ thống trước khi chúng trở nên được công nhận rộng rãi, và ông nhấn mạnh tầm quan trọng của trách nhiệm, độ tin cậy và nhận thức rủi ro trong đổi mới. Ông đã có những đóng góp đáng kể cho sự phát triển của các kiến trúc máy tính bảo mật, bao gồm cả thiết kế của hệ điều hành Multics. Neumann cũng là người sáng lập và điều hành Diễn đàn Risks của ACM, một cộng đồng trực tuyến được kính trọng để thảo luận về các sự cố máy tính, điểm yếu và các mối đe dọa công nghệ mới nổi. Những nhận xét của ông về tầm quan trọng của sự đơn giản trong thiết kế và nhu cầu bảo mật phải được tích hợp vào kiến trúc hệ thống từ đầu vẫn còn phù hợp trong cảnh quan máy tính ngày nay, đặc biệt là trong việc phát triển các mô hình AI, API, LLM và framework mới.

TIPS & TRICKS CHO DEV

Tích hợp MCP với Git

Vấn đề: Cần kết nối AI với Git để đẩy mã nguồn lên repository.

Cách làm: Sử dụng MCP với Claude Code, nhập lệnh `git push` sau khi mã hóa bằng Claude. Ví dụ: "Push mã nguồn lên GitHub bằng Claude Code".

Đánh giá: Hiệu quả khi cần phiên bản mã nguồn, nên dùng khi làm việc nhóm.

MCP với Database

Vấn đề: Cần truy xuất dữ liệu từ database để huấn luyện mô hình AI.

Cách làm: Sử dụng MCP với Claude Desktop, nhập lệnh `SELECT * FROM table` để truy xuất dữ liệu. Ví dụ: "Truy xuất dữ liệu từ MySQL bằng Claude Desktop".

Đánh giá: Hiệu quả khi cần dữ liệu lớn, nên dùng khi huấn luyện mô hình.

MCP với APIs

Vấn đề: Cần kết nối AI với APIs để thu thập dữ liệu từ dịch vụ ngoài.

Cách làm: Sử dụng MCP với Cursor, nhập lệnh `curl https://api.example.com/`

`data` để thu thập dữ liệu. Ví dụ: "Thu thập dữ liệu từ API OpenWeatherMap bằng Cursor".
Đánh giá: Hiệu quả khi cần dữ liệu thời gian thực, nên dùng khi cần tích hợp dịch vụ ngoài.

BÀI HỌC AI HÔM NAY CHO DEV

1. **Fine-tuning & LoRA cho use case cụ thể**
Fine-tuning và LoRA (Low-Rank Adaptation) là các kỹ thuật quan trọng giúp cải thiện hiệu suất của mô hình ngôn ngữ lớn (LLM) trong các ứng dụng cụ thể. Dev cần biết về những kỹ thuật này để tối ưu hóa hiệu suất của LLM trong dự án của mình.

Ví dụ, khi sử dụng LLM cho việc phân tích cảm xúc văn bản, fine-tuning có thể giúp mô hình học hỏi các mẫu ngôn ngữ và từ vựng cụ thể trong lĩnh vực đó.

3. Ví dụ thực tế: `transformers.Trainer(model, ...)` có thể được sử dụng để fine-tune mô hình.

4. Tip hoặc bước tiếp theo: Hãy thử nghiệm với các kỹ thuật fine-tuning và LoRA khác nhau để tìm ra phương pháp phù hợp nhất cho dự án của bạn.