BIS - MT · 28/06/2026

Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

> *"Cần cù bù thông minh."*
>
> — Tục ngữ Việt Nam
>
> *Sự chăm chỉ và kiên trì có thể bù đắp cho những hạn chế về năng khiếu bẩm sinh.*

TIN TỨC NỔI BẬT

1 Các nhà nghiên cứu bảo mật gióng lên hồi chuông cảnh báo về các 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 an ninh đã bày tỏ mối quan ngại về các điểm yếu trong mã được tạo bởi AI, nhấn mạnh các rủi ro tiềm ẩn liên quan đến công nghệ mới nổi này. Vấn đề này xuất phát từ thực tế rằng các công cụ tạo mã AI thường dựa vào các đoạn mã và thuật toán đã tồn tại, có thể chứa các điểm yếu đã biết. Các điểm yếu này sau đó có thể được giới thiệu vào mã được tạo bởi AI, có khả năng dẫn đến các vi phạm bảo mật. Các nhà nghiên cứu đã xác định một số vấn đề chính với mã được tạo bởi AI, bao gồm sự thiếu minh bạch và trách nhiệm giải trình trong quá trình phát triển. Điều này làm cho việc xác định nguồn gốc và chất lượng của mã trở nên khó khăn, tăng nguy cơ các điểm yếu được giới thiệu. Hơn nữa, việc sử dụng mã đã tồn tại cũng có thể dẫn đến sự lây lan của các điểm yếu đã biết, thay vì tạo ra mã mới và an toàn. Kết quả nghiên cứu của các nhà nghiên cứu đóng vai trò như một cảnh báo cho các tổ chức dựa vào mã được tạo bởi AI, nhấn mạnh nhu cầu đánh giá và kiểm tra cẩn thận công nghệ này để đảm bảo bảo mật và tính toàn vẹn của hệ thống của họ.

2 Claude Code so với GitHub Copilot 2026: SWE-bench, Giá cả [Đã thử nghiệm]

Claude Code vs GitHub Copilot 2026: SWE-bench, Pricing [Tested]

tech-insider.org Đọc bài viết →

Trong một so sánh gần đây, Claude Code và GitHub Copilot đã được kiểm tra trong một điểm chuẩn Kỹ thuật Phần mềm (SWE). Kết quả cho thấy Claude Code vượt trội so với GitHub Copilot ở nhiều lĩnh vực, bao gồm chất lượng mã, hiệu suất và khả năng thích nghi. Bài kiểm tra SWE-bench, đánh giá hiệu suất của các công cụ mã hóa được hỗ trợ bởi AI, đã tìm thấy rằng Claude Code có thể viết mã chính xác và hiệu quả hơn, với mức độ thích nghi cao hơn với các ngôn ngữ lập trình và kịch bản khác nhau. Khi nói đến giá cả, cả hai công cụ đều cung cấp các lựa chọn cạnh tranh. GitHub Copilot có sẵn dưới dạng dịch vụ dựa trên đăng ký, với thời gian thử nghiệm miễn phí và phí hàng tháng bắt đầu từ 19,95 đô la. Claude Code, mặt khác, cung cấp một mô hình giá nhiều cấp, với kế hoạch miễn phí có sẵn cho các nhà phát triển cá nhân, cũng như kế hoạch trả phí bắt đầu từ 49 đô la mỗi tháng cho các nhóm. Tổng thể, kết quả của bài kiểm tra SWE-bench cho thấy Claude Code có thể là một lựa chọn hiệu quả hơn cho các nhà phát triển đang tìm kiếm một công cụ mã hóa được hỗ trợ bởi AI đáng tin cậy và hiệu quả. Tuy nhiên, các lựa chọn giá cả của cả hai công cụ sẽ phụ thuộc vào nhu cầu và ngân sách cụ thể của nhà phát triển.

3 Lớp Command AI Kho hàng Đa Agent mang lại sự xuất sắc trong 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 Lớp lệnh AI Nhà kho Đa tác nhân, được thiết kế để nâng cao hiệu quả hoạt động và trí tuệ chuỗi cung ứng trong môi trường nhà kho. Giải pháp sáng tạo này tận dụng AI và học máy để tối ưu hóa hoạt động nhà kho, đơn giản hóa quy trình và cải thiện việc ra quyết định. 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 phản hồi nhanh chóng với các điều kiện thay đổi. Bằng cách tích hợp với các hệ thống và thiết bị khác nhau, nó có thể thu thập và phân tích lượng lớn dữ liệu, cung cấp thông tin chi tiết quý giá về hiệu suất chuỗi cung ứng. Các tính năng chính của Lớp lệnh AI Nhà kho Đa tác nhân bao gồm: - Giám sát và kiểm soát thời gian thực các hoạt động nhà kho - Tích hợp với các hệ thống và thiết bị khác nhau - Phân tích dữ liệu và trí tuệ chuỗi cung ứng - Cải thiện khả năng ra quyết định Công nghệ này có tiềm năng cách mạng hóa hoạt động nhà kho, cho phép các doanh nghiệp phản hồi nhanh chóng với các điều kiện thị trường thay đổi và cải thiện hiệu quả chuỗi cung ứng tổng thể. Bằng cách tận dụng

sức mạnh của AI và học máy, Lớp lệnh AI Nhà kho Đa tác nhân có thể giúp các tổ chức đạt được sự xuất sắc hoạt động và dẫn đầu trong một thị trường cạnh tranh.

4

Tháng 6 năm 2026: Bản tin LangChain — Fleet On-Call Copilot, Deep Agents Rubrics, và nhiều hơn nữa

June 2026: LangChain Newsletter — Fleet On-Call Copilot, Deep Agents Rubrics, and More

LangChain Blog Đọc bài viết →

LangChain đã phát hành một số bản cập nhật cho nền tảng LangSmith của mình, nâng cao khả năng xây dựng và triển khai các tác nhân AI. Một người đồng hành trực tuyến mới cho việc phân loại cho phép các tác nhân soạn thảo các bản cập nhật để xem xét, trong khi các tác nhân hiện có thể sử dụng máy tính ảo cách ly để truy cập mã và tệp. Nền tảng này cũng có tính năng phát âm thanh trực tuyến để gỡ lỗi nhanh hơn, theo dõi tiến trình trực tiếp cho các thí nghiệm và công cụ RubricMiddleware để các Deep Agents đánh giá công việc của chính mình. Ngoài ra, LangChain đã giới thiệu một mẫu tác nhân sẵn có và một khóa học về Triển khai LangSmith, bao gồm việc triển khai, quản lý và kiểm soát các tác nhân trên quy mô lớn. Nền tảng này cũng đã được các công ty như Box và Harmonic sử dụng để tăng tốc độ cải tiến và lặp lại của tác nhân, cũng như cải thiện khả năng giữ chân khách hàng. LangChain sẽ tổ chức một số sự kiện sắp tới, bao gồm các buổi hội thảo và gặp mặt tại Chicago, Berlin, Washington DC và Las Vegas, nơi những người tham dự có thể tìm hiểu về việc xây dựng và triển khai các tác nhân AI. Những sự kiện này sẽ có các buổi nói chuyện từ các chuyên gia trong ngành và các phiên thực hành với nền tảng của LangChain.

5

CEO của SoftBank không phải là người duy nhất đặt câu hỏi về sự cường điệu trung tâm dữ liệu quỹ đạo của Elon Musk

SoftBank's CEO isn't the only one with questions about Elon Musk's orbital data center hype

TechCrunch AI Đọc bài viết →

CEO Masayoshi Son của SoftBank đã thể hiện sự hoài nghi về tầm nhìn của Elon Musk về các trung tâm dữ liệu quỹ đạo. Tại một cuộc họp cổ đông gần đây, Son đã đặt câu hỏi về hiệu quả chi phí và thời gian xây dựng các trung tâm dữ liệu trong không gian, cho rằng chúng

> sẽ không giải quyết nhu cầu hiện tại về các trung tâm dữ liệu và sẽ mất quá nhiều thời gian để phát triển. Sự hoài nghi này là iron khi xem xét lịch sử của SoftBank trong việc đặt "cược điên rồ". Trên podcast TechCrunch Equity, các host đã thảo luận về nhận xét của Son, lưu ý rằng kế hoạch của Musk để tạo ra một chùm sao vệ tinh để tạo thành một "trung tâm dữ liệu quỹ đạo" thực sự sẽ tạo ra nhiều doanh nghiệp hơn cho SpaceX. Cuộc trò chuyện cũng đề cập đến các chủ đề khác, bao gồm chip tùy chỉnh của OpenAI, việc chipmaker Groq nhận được 650 triệu đô la tài trợ mới, và xu hướng ngày càng tăng của các công ty chuyển sang "neo-clouds" để đáp ứng nhu cầu ngày càng tăng về sức mạnh tính toán.

6

Mở khóa cơ hội kiếm tiền Halal cho Freelancer công nghệ Pakistan thông qua AI và Automation

Unlocking Halal Earning Opportunities for Pakistani Tech Freelancers through AI and Automation

Dev.to AI Đọc bài viết →

Các nhà tự do làm việc trong lĩnh vực công nghệ của Pakistan đã nổi lên như một lực lượng kiên cường trong nền kinh tế toàn cầu, tận dụng cơ hội làm việc từ xa để kiếm ngoại tệ. Tuy nhiên, việc điều hướng thế giới phức tạp của các công cụ được hỗ trợ bởi AI và tự động hóa có thể là một thách thức, đặc biệt là khi đảm bảo thu nhập halal. Một cuộc khảo sát gần đây của Upwork cho thấy sự tăng trưởng đáng kể trong nhu cầu về các dịch vụ được thúc đẩy bởi AI như chú thích dữ liệu, kiểm duyệt nội dung và phát triển chatbot. Các nhà tự do làm việc trong lĩnh vực công nghệ của Pakistan phải đối mặt với một thách thức quan trọng trong việc phát triển kỹ năng số liên quan đến khách hàng quốc tế, cạnh tranh với một nhóm tài năng toàn cầu khổng lồ. Để vượt qua điều này, các công cụ được hỗ trợ bởi AI có thể được tận dụng để mở ra cơ hội mới cho thu nhập halal. Ví dụ, các lời nhắc AI có thể giúp các nhà tự do làm việc trong lĩnh vực giáo dục tạo ra nội dung giáo dục chất lượng cao, trong khi các dịch vụ tiếp thị số được thúc đẩy bởi AI có thể giúp các doanh nghiệp tối ưu hóa sự hiện diện trực tuyến của họ. Các nhà tự do làm việc trong lĩnh vực công nghệ của Pakistan có thể bắt đầu bằng cách xác định các lĩnh vực mà tự động hóa có thể thêm giá trị vào công việc của họ, chẳng hạn như sử dụng các thuật toán học máy để phân tích dữ liệu khách hàng hoặc tạo ra các chatbot tự động. Bằng cách tận dụng tự động hóa và các thuật toán học máy một cách có trách nhiệm và minh bạch, các nhà tự do làm việc trong lĩnh vực công nghệ của Pakistan có thể mở ra cơ hội mới cho sự tăng trưởng, đổi mới và thịnh vượng kinh tế.

7

Agent đang thay đổi công việc như thế nào

How agents are transforming work

OpenAI Blog Đọc bài viết →

Một bài nghiên cứu gần đây của OpenAI đã nhấn mạnh tác động đáng kể của các tác nhân AI đối với lực lượng lao động hiện đại. Theo nghiên cứu, những tác nhân này đang cách mạng hóa cách thức công việc được thực hiện bằng cách cho phép nhân viên giải quyết các nhiệm vụ dài hơn và phức tạp hơn với sự dễ dàng lớn hơn. Kết quả là năng suất lao động đang được mở rộng trên nhiều vai trò khác nhau, dẫn đến hiệu quả và sản lượng tăng cao. Các phát hiện cho thấy các tác nhân AI đang đóng vai trò quan trọng trong việc chuyển đổi cách thức công việc được thực hiện, cho phép cá nhân tập trung vào các nhiệm vụ và trách nhiệm có giá trị cao hơn. Mặc dù các chi tiết cụ thể về phương pháp và kết quả của bài nghiên cứu không được cung cấp,

nhưng ý nghĩa tổng thể là rõ ràng: các tác nhân AI đang sẵn sàng có tác động sâu sắc đến tương lai của công việc, cho phép nhân viên trở nên năng suất và hiệu quả hơn trong vai trò của họ. Kết luận của nghiên cứu nhấn mạnh tiềm năng của AI trong việc thúc đẩy những thay đổi đáng kể trong cách thức công việc được thực hiện, và ý nghĩa của nó đối với lực lượng lao động hiện đại có khả năng sẽ rất rộng lớn.

8

MCP với Code Mode

MCP on Code Mode

Changelog Đọc bài viết →

Tuần này, Matt Carey từ Cloudflare thảo luận về Code Mode và mối quan hệ của nó với MCP (Model-Code-Program). Carey tiết lộ rằng hầu hết mọi người đã hiểu lầm MCP, và cách Code Mode phía máy chủ cho phép một máy chủ MCP lộ hơn 2.500 điểm cuối API của Cloudflare bằng cách sử dụng khoảng 1.000 token ngữ cảnh. Cuộc trò chuyện cũng đề cập đến trình tải Worker động mà chạy mã được viết bởi model một cách an toàn trong môi trường V8 isolate, quy trình làm việc của Carey với Claude, và vai trò của bộ nhớ trong tương lai của các agent. Ngoài ra, tập này cũng nhắc đến các nhà tài trợ khác nhau, bao gồm Coder.com, Tailscale, RWX và Fly.io, cung cấp môi trường bảo mật, truy cập dựa trên danh tính, nền tảng CI/CD và dịch vụ triển khai, tương ứng.

TIPS & TRICKS CHO DEV

Quản lý context window

Vấn đề: Giới hạn kích thước context window làm giảm hiệu suất mô hình.

Cách làm: Sử dụng kỹ thuật chunking, chia dữ liệu thành phần nhỏ hơn. Ví dụ, Claude có thể xử lý 131k tokens.

Đánh giá: Hiệu quả khi xử lý dữ liệu lớn, nhưng cần tối ưu hóa để tránh mất thông tin.

Tối ưu hóa long-context

Vấn đề: Mô hình không thể xử lý long-context dẫn đến sai sót.

Cách làm: Sử dụng mô hình như Longformer, hỗ trợ xử lý long-context lên đến 4096 tokens.

Đánh giá: Hiệu quả khi cần xử lý văn bản dài, nhưng yêu cầu tài nguyên lớn.

Cải thiện memory

Vấn đề: Giới hạn bộ nhớ làm giảm hiệu suất mô hình.

Cách làm: Sử dụng kỹ thuật model pruning, giảm kích thước mô hình mà không ảnh hưởng đến hiệu suất.

Đánh giá: Hiệu quả khi cần giảm kích thước mô hình, nhưng cần thử nghiệm để đảm bảo hiệu suất.

BÀI HỌC AI HÔM NAY CHO DEV

1. **Tối ưu chi phí & hiệu năng LLM**
2. Lập trình viên cần biết cách tối ưu chi phí và hiệu năng của mô hình ngôn ngữ lớn (LLM) để giảm thiểu chi phí và tăng tốc độ xử lý. Điều này đặc biệt quan trọng khi tích hợp AI vào ứng dụng. Việc tối ưu hóa giúp giảm tải hệ thống và tăng trải nghiệm người dùng.
3. Ví dụ, sử dụng kỹ thuật fine-tuning và LoRA (Low-Rank Adaptation) có thể giúp giảm kích thước mô hình và tăng tốc độ xử lý. Ví dụ code: `model = LLM.from_pretrained('base_model'); model.fine_tune(dataset, epochs=5)`.
4. Tip: Sử dụng các thư viện như Hugging Face Transformers để tối ưu hóa LLM và thử nghiệm với các kỹ thuật khác nhau như quantization và pruning để đạt được hiệu suất tốt nhất.