

Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

"Innovation distinguishes between a leader and a follower."

↳ Sáng tạo phân biệt người dẫn đầu với kẻ đi theo.

— Steve Jobs

Trong thế giới cạnh tranh, tư duy đổi mới và dám thách thức hiện trạng là điều tạo ra sự khác biệt thật sự.

TIN TỨC NỔI BẬT

1

9 Lựa chọn thay thế AI hỗ trợ lập trình trên Terminal: Hướng dẫn toàn diện 2026

9 AI Coding Alternatives for Terminal Development: Complete 2026 Guide

Pasquale Pillitteri

[Đọc bài viết →](#)

Bài viết "9 Giải pháp Lập trình AI cho Phát triển Terminal: Hướng dẫn Hoàn chỉnh 2026" cung cấp cái nhìn tổng quan về các công cụ lập trình được hỗ trợ bởi AI cho phát triển terminal. Những giải pháp thay thế này nhằm mục đích tối ưu hóa quá trình lập trình, tăng hiệu suất và năng suất. Hướng dẫn này bao gồm chín công cụ lập trình AI, mỗi công cụ có các tính năng và khả năng độc đáo. Những công cụ này có thể hỗ trợ các nhiệm vụ như hoàn thiện mã, gỡ lỗi và tối ưu hóa. Chúng thường sử dụng các thuật toán học máy để phân tích và hiểu mã, cho phép chúng cung cấp các đề xuất và khuyến nghị chính xác hơn. Bài viết không chỉ định tên của chín công cụ lập trình AI, nhưng nó gợi ý rằng chúng có thể được sử dụng cho các ngôn ngữ lập trình và nhiệm vụ phát triển khác nhau. Bằng cách tận dụng công nghệ AI, những công cụ này có thể giúp các nhà phát triển tiết kiệm thời gian và giảm thiểu lỗi, cuối cùng cải thiện trải nghiệm lập trình tổng thể. Hướng dẫn này được thiết kế để trở thành một tài nguyên toàn diện cho các nhà phát triển muốn khám phá các giải pháp lập trình được hỗ trợ bởi AI cho phát triển terminal vào năm 2026.

2

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

Trong một so sánh gần đây, Claude Code và GitHub Copilot đã được thử nghiệm trong một loạt các bài kiểm tra. Đánh giá tập trung vào tiêu chuẩn SWE-bench (Kỹ thuật Phần mềm), đánh giá hiệu suất của các công cụ mã hóa được hỗ trợ bởi AI. Kết quả cho thấy Claude Code vượt trội so với GitHub Copilot ở nhiều lĩnh vực, bao gồm chất lượng mã, hiệu quả và độ tin cậy. So sánh cũng xem xét các mô hình định giá của hai công cụ. Trong khi GitHub Copilot cung cấp một cấu trúc định giá phân cấp, với kế hoạch miễn phí có sẵn cho các dự án mã nguồn mở, Claude Code tính phí hàng tháng cố định. Chi tiết định giá của Claude Code không được chỉ định trong bài viết. Các bài kiểm tra được thực hiện để cung cấp cho các nhà phát triển một sự hiểu biết tốt hơn về điểm mạnh và điểm yếu của từng công cụ. Kết quả có thể hữu ích cho những người đang xem xét công cụ trợ lý mã hóa được hỗ trợ bởi AI nào để sử dụng cho các dự án của họ.

3

Xây dựng Workflow vận hành bởi agent sử dụng Cortex AI và Managed MCP Servers

Build Agent-Powered Workflows Using Cortex AI and Managed MCP Servers

Snowflake

[Đọc bài viết →](#)

Snowflake đã giới thiệu một tích hợp mới cho phép người dùng xây dựng các quy trình làm việc được hỗ trợ bởi các tác nhân sử dụng Cortex AI và máy chủ Managed MCP. Tích hợp này cho phép các nhà phát triển tự động hóa các nhiệm vụ và quy trình làm việc phức tạp bằng cách tận dụng khả năng của Cortex AI và khả năng mở rộng của máy chủ Managed MCP. Với tích hợp này, người dùng có thể tạo các quy trình làm việc tùy chỉnh sử dụng khả năng học máy và trí tuệ nhân tạo của Cortex AI để phân tích và xử lý dữ liệu. Các quy trình làm việc này có thể được kích hoạt bởi các sự kiện khác nhau và có thể tự động hóa các nhiệm vụ như xử lý dữ liệu, báo cáo và nhiều hơn nữa. Tích hợp với máy chủ Managed MCP cung cấp một môi trường có thể mở rộng và bảo mật để chạy các quy trình làm việc này. Máy chủ MCP có thể xử lý các lượng dữ liệu lớn và cung cấp mức hiệu suất cao, khiến nó trở nên lý tưởng cho các quy trình làm việc phức tạp. Tích hợp này được thiết kế để giúp các nhà phát triển tối ưu hóa quy trình làm việc của họ, giảm thiểu công việc thủ công và tăng năng suất. Bằng cách tận dụng khả năng của Cortex AI và máy chủ Managed MCP, người dùng có thể tạo ra các quy trình làm việc hiệu quả và tự động hơn có thể giúp họ đạt được mục tiêu kinh doanh của mình.

4

Tại sao agent harness quan trọng ngang model trong bảo mật AI

Why the agent harness matters as much as the model in AI security

BD Tech Talks

[Đọc bài viết →](#)

Một nghiên cứu gần đây của công ty an ninh mạng Lasso Security đã nhấn mạnh tầm quan trọng của bộ phận điều khiển (harness) trong an ninh AI, thách thức giả định phổ biến rằng nó là một lớp nền trung lập. Các nhà nghiên cứu đã xây dựng một tác nhân tấn công tự động và chỉ thay đổi khuôn khổ (framework) thời gian chạy bên dưới nó, cho thấy rằng bộ phận điều khiển ảnh hưởng đáng kể đến hiệu suất trong điều kiện đối thủ. Nghiên cứu cho thấy rằng việc thay thế bộ phận điều khiển có thể thay đổi chiến lược của một cuộc tấn công, xác định liệu các rào cản mạng có chặn hoặc cho phép các payload độc hại và thậm chí có thể chấm dứt các vòng thực hiện đa lượt. Các nhà nghiên cứu đã thử nghiệm năm mô hình tấn công trên hai bộ phận điều khiển, một khuôn khổ mã nguồn mở (deepagents của LangChain) và một công cụ của nhà cung cấp mã nguồn đóng (Claude Agent SDK của Anthropic), tạo ra mười tổ hợp mô hình-bộ phận điều khiển. Họ đã thiết kế 20 nhiệm vụ tấn công trên năm môi trường ứng dụng mô

phỏng, tạo ra một tập dữ liệu gồm 1.000 lần thực hiện tấn công rời rạc. Kết quả cho thấy hai bộ phận điều khiển này tạo ra các payload yêu cầu cơ bản khác nhau, với khuôn khổ mã nguồn mở đặt lời nhắc hệ thống của người dùng trước tiên và khuôn khổ mã nguồn đóng đặt dòng nhận dạng ở đầu. Kết quả của nghiên cứu này□ lộ một điểm mù an ninh quan trọng: an ninh không phải là thuộc tính của mô hình (model) alone, mà là của cặp mô hình-bộ phận điều khiển cụ thể. Điều này nhấn mạnh nhu cầu của các kỹ sư phải xem xét bộ phận điều khiển tác nhân như một thành phần quan trọng trong an ninh AI, chứ không phải là một lớp nền trung lập.

5

Ngĩ đến ACE? Chúng ta có thể làm được với ít token hơn

Thinking of ACE? We Can Do It with Fewer Tokens

Hugging Face Blog

[Đọc bài viết →](#)

Các nhà nghiên cứu đã phát triển hai hệ thống, ALTK-Evolve và ACE (Kỹ thuật Kỹ sư Context), để cải thiện hiệu suất của các tác nhân Large Language Model (LLM) bằng cách cho phép chúng học hỏi từ kinh nghiệm trong quá khứ. Cả hai hệ thống đều sử dụng bộ nhớ agentic để chuyển đổi các đường dẫn trong quá khứ thành các bài học có thể tái sử dụng và cung cấp chúng trở lại tại thời điểm suy luận, mà không cần nhắc người hoặc cập nhật trọng số. Sự khác biệt chính giữa hai hệ thống nằm ở cách chúng cung cấp các bài học này cho model. ACE tổ chức các bài học thành một cuốn sổ tay toàn diện, đang phát triển, trong khi ALTK-Evolve hợp nhất các bài học thành các hướng dẫn có thể thu hồi riêng lẻ. Hai hệ thống đều đồng ý rằng việc nén các bài học khó khăn của một tác nhân thành một bản tóm tắt gọn gàng không có lợi, và thay vào đó khuyến nghị đếm và giữ lại mỗi bài học. Cách tiếp cận này được phản ánh trong việc sử dụng các bộ đếm per-bullet của ACE và các bộ đếm hỗ trợ của ALTK-Evolve, theo dõi số lượng tập độc lập đã tạo ra mỗi bài học. Sự khác biệt chính giữa hai hệ thống nằm ở cách tiếp cận giao hàng. ACE thêm cuốn sổ tay toàn diện vào mỗi bước, trong khi ALTK-Evolve coi việc giao hàng như một nóm vắn, gửi một lõi cố định nhỏ của các hướng dẫn hỗ trợ cao và mở rộng nó cho mỗi nhiệm vụ với một số hướng dẫn được chọn. Cách tiếp cận này dẫn đến ALTK-Evolve hiệu quả hơn, với giảm 40% chi phí suy luận trên một model mạnh và tỷ lệ chính xác tương đương với ACE trên một model yếu.

Hội nghị thượng đỉnh Kỹ thuật IEEE hỗ trợ Chuyển đổi số của Bhutan

IEEE Engineering Summit Supports Bhutan's Digital Transformation

IEEE Spectrum [Đọc bài viết →](#)

Hội nghị thượng đỉnh kỹ thuật IEEE đã được tổ chức tại Paro, Bhutan, vào ngày 9-10 tháng 6, với mục tiêu hỗ trợ quá trình chuyển đổi số của đất nước. Sự kiện này, được tổ chức hợp tác với chính phủ Bhutan, đã tập hợp các nhà lãnh đạo học thuật toàn cầu, chuyên gia công nghệ và quan chức chính phủ để thảo luận về sự giao thoa giữa chuyển đổi số, giáo dục kỹ thuật và phát triển bền vững. Hội nghị thượng đỉnh tập trung vào việc phát triển chương trình đào tạo kỹ thuật từ lý thuyết sang ứng dụng, xây dựng năng lực nghiên cứu có giá trị cao và tích hợp trí tuệ nhân tạo (AI) vào giáo dục đại học. Công chúa Chimi Yangzom Wangchuck của Bhutan đã nhấn mạnh tầm quan trọng của việc kết hợp đổi mới công nghệ với triết lý Hạnh phúc Quốc gia Tổng thể (GNH) của quốc gia, ưu tiên sự thịnh vượng, bền vững và đạo đức. Sự kiện này đã thiết lập một kế hoạch hợp tác cho một nền kinh tế tri thức có giá trị cao tại Bhutan thông qua một số lĩnh vực trọng tâm. Trong khuôn khổ sáng kiến này, IEEE đã cung cấp cho các trường đại học và cơ sở của Bhutan một thử nghiệm miễn phí sáu tháng của hai tài nguyên kỹ thuật. Chủ tịch và Giám đốc điều hành của IEEE, Mary Ellen Randall, cho biết tổ chức này nhằm hỗ trợ các trường đại học và sinh viên địa phương phát triển công nghệ đáng tin cậy, sẵn sàng cho tương lai, tôn vinh cam kết của quốc gia đối với sự phát triển bền vững và hạnh phúc của con người.

monday.com tái kiến trúc Sidekick để hỗ trợ AI chủ động với các subagent chuyên biệt

LangChain Blog [Đọc bài viết →](#)

Vào thứ hai, trợ lý AI của monday.com, Sidekick, ban đầu có kiến trúc đại lý chung cho phép nó thực hiện nhiều nhiệm vụ khác nhau. Tuy nhiên, khi thêm nhiều công cụ, hệ thống trở nên phức tạp, mơ hồ và tốn kém để vận hành. Đội ngũ nhận ra rằng mục tiêu tạo ra một trợ lý AI chủ động có thể hỗ trợ người dùng trong công việc của họ bị cản trở bởi kiến trúc phẳng. Để giải quyết vấn đề này, đội ngũ đã tái kiến trúc Sidekick bằng cách phân chia trách nhiệm trên toàn hệ thống, giới thiệu các công cụ bị giới hạn, hộp cát và các đại lý chuyên dụng.

Cách tiếp cận này cho phép họ vẽ ra ranh giới rõ ràng hơn giữa lập kế hoạch, lý luận chuyên ngành, sử dụng công cụ và thực thi. Đội ngũ sử dụng LangChain Deep Agents để phân chia các vấn đề phức tạp thành các nhiệm vụ có thể quản lý, cho phép đại lý chính hiểu mục tiêu của người dùng và ủy quyền nhiệm vụ cho các đại lý chuyên dụng hoặc môi trường thực thi cách ly. Kiến trúc mới đã cải thiện hiệu suất và khả năng của Sidekick, cho phép nó xử lý nhiều loại nhiệm vụ, từ tìm kiếm đơn giản đến các quy trình làm việc phức tạp. Đội ngũ đã học được rằng hộp cát là một nguyên thủy quan trọng cho đại lý, cho phép nó thực hiện công việc phức tạp và lặp lại. Bằng cách tái kiến trúc Sidekick, đội ngũ đã thực hiện một bước tiến đáng kể towards tạo ra một trợ lý AI chủ động có thể hỗ trợ người dùng trong công việc của họ.

8

Chrome áp dụng biện pháp bảo vệ có thể là tốt nhất chống lại việc chiếm đoạt tài khoản

Chrome adopts what may be the best protection yet against account takeovers

Ars Technica

[Đọc bài viết →](#)

Trình duyệt Chrome của Google đã giới thiệu một tính năng mới gọi là thông tin đăng nhập phiên bị ràng buộc thiết bị (DBSCs) để bảo vệ chống lại việc chiếm đoạt tài khoản. Tính năng này lưu trữ một khóa mã hóa duy nhất trong một vị trí an toàn trên thiết bị, được gọi là Trusted Platform Module (TPM) trên Windows hoặc một khu vực an toàn trên macOS và iOS. DBSCs ngăn chặn việc đánh cắp cookie phiên, được sử dụng để tăng tốc duyệt trang trên các trang web nhạy cảm yêu cầu xác thực người dùng. Khi xác thực hai yếu tố và passkey trở nên phổ biến, các kẻ tấn công đã chuyển sang đánh cắp cookie phiên để truy cập vào các trang web bị hạn chế. DBSCs hoạt động bằng cách yêu cầu trình duyệt ký cookie phiên với khóa riêng được lưu trữ trong vị trí an toàn, khiến cho các kẻ tấn công không thể sử dụng cookie bị đánh cắp để truy cập. Tính năng mới này hiện có sẵn trong phiên bản Chrome 147 cho Windows và 150 cho macOS, nhưng chỉ được bật cho một tập hợp người dùng hạn chế. Có khả năng các trình duyệt dựa trên Chromium khác sẽ triển khai DBSCs trong tương lai.

TIPS & TRICKS CHO DEV

Chain-of-Thought Prompting

Vấn đề: Khó khăn khi tạo prompt phức tạp cho mô hình AI.

Cách làm: Sử dụng kỹ thuật chain-of-thought prompting, ví dụ "Tôi muốn tính thuế thu nhập, trước tiên tôi cần tính thu nhập chịu thuế, sau đó áp dụng tỷ lệ thuế".

Đánh giá: Hiệu quả khi tạo prompt phức tạp, giúp mô hình AI hiểu rõ vấn đề.

Few-Shot Learning

Vấn đề: Cần nhiều dữ liệu để huấn luyện mô hình AI.

Cách làm: Sử dụng kỹ thuật few-shot learning, ví dụ "Tạo mô hình phân loại văn bản với chỉ 5 ví dụ".

Đánh giá: Hiệu quả khi dữ liệu hạn chế, giúp mô hình AI học hỏi nhanh chóng.

System Prompt Design

Vấn đề: Khó khăn khi thiết kế prompt cho mô hình AI.

Cách làm: Sử dụng kỹ thuật system prompt design, ví dụ "Tạo prompt với định dạng đầu vào và đầu ra cụ thể".

Đánh giá: Hiệu quả khi cần tạo prompt chính xác, giúp mô hình AI hiểu rõ yêu cầu.

BÀI HỌC AI HÔM NAY CHO DEV

1. Tối ưu chi phí & hiệu năng LLM

2. Để phát triển ứng dụng AI hiệu quả, các dev cần biết cách tối ưu chi phí và hiệu năng của mô hình ngôn ngữ lớn (LLM). Điều này giúp giảm thiểu chi phí vận hành và cải thiện trải nghiệm người dùng. Việc tối ưu hóa LLM cũng giúp tăng tốc độ xử lý và giảm tải trên hệ thống.

3. Ví dụ, có thể sử dụng kỹ thuật fine-tuning để điều chỉnh mô hình LLM cho phù hợp với use case cụ thể, giảm thiểu số lượng tham số và tăng tốc độ huấn luyện.

4. Tip: Sử dụng các công cụ như LoRA (Low-Rank Adaptation) để fine-tuning mô hình LLM và giảm thiểu chi phí vận hành, đồng thời cải thiện hiệu năng của ứng dụng AI.

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

Nguồn: Google News · Groq AI