

Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

“Không có gì là không thể.”

— Tục ngữ Việt Nam

Giới hạn lớn nhất của con người nằm ở chính tâm trí — khi tin vào bản thân, mọi thách thức đều có thể vượt qua.

TIN TỨC NỔI BẬT

1

11 lựa chọn thay thế GitHub Copilot mà các developer đã chuyển sang dùng trong năm 2026

11 GitHub Copilot Alternatives Developers Switched To in 2026

Security Boulevard [Đọc bài viết →](#)

Vào năm 2026, ngày càng nhiều developer đã tìm kiếm các giải pháp thay thế cho GitHub Copilot, một trợ lý mã hóa AI phổ biến. Sự thay đổi này được thúc đẩy bởi nhiều yếu tố, bao gồm lo ngại về bảo mật dữ liệu và mong muốn kiểm soát nhiều hơn đối với các gợi ý mã hóa. Kết quả là, một số giải pháp thay thế đã trở nên phổ biến trong cộng đồng developer. Những giải pháp thay thế này cung cấp các chức năng tương tự như GitHub Copilot, chẳng hạn như tính năng tự động hoàn thiện mã và gợi ý. Một số giải pháp thay thế đáng chú ý bao gồm Kite, TabNine và CodeWithMe. Các giải pháp thay thế khác như Code Scanner của GitHub, Codex của Microsoft và Llama của Meta cũng đã được các developer áp dụng. Những công cụ này tận dụng AI để cung cấp cho developer các gợi ý mã, cải thiện hiệu suất mã hóa và tăng cường hợp tác. Mặc dù GitHub Copilot vẫn là một công cụ được sử dụng rộng rãi, nhu cầu ngày càng tăng đối với các giải pháp thay thế này nhấn mạnh nhu cầu về sự linh hoạt và tùy chỉnh nhiều hơn trong các trợ lý mã hóa. Khi nhu cầu về các công cụ mã hóa AI tiếp tục tăng, các developer đang tìm kiếm các lựa chọn đáp ứng nhu cầu và sở thích cụ thể của họ.

2

Claude Code đấu với Cursor 2026: Đạt 80.8% SWE-bench, 1M Context [Đã thử nghiệm]

Trong một bài kiểm tra benchmarking gần đây, Claude Code và Cursor 2026 đã được so sánh trong một loạt các đánh giá. Kết quả cho thấy Claude Code đã đạt được điểm số ấn tượng 80,8% trên SWE-bench, một công cụ benchmark được sử dụng rộng rãi để đánh giá các model ngôn ngữ. Ngoài ra, Claude Code đã chứng minh khả năng của mình bằng cách xử lý một bài kiểm tra ngữ cảnh khổng lồ với 1 triệu token. SWE-bench là một công cụ đánh giá toàn diện đánh giá khả năng của một model ngôn ngữ trong việc hiểu và tạo ra văn bản giống con người. Điểm số 80,8% cho thấy Claude Code đã đạt được tiến bộ đáng kể trong lĩnh vực này, thể hiện tiềm năng của nó cho các ứng dụng thực tế. Bài kiểm tra ngữ cảnh 1 triệu token [\[1\]](#) nhấn mạnh khả năng của Claude Code trong việc xử lý lượng lớn thông tin và tạo ra các phản hồi mạch lạc. Thành tựu này là minh chứng cho khả năng mở rộng của model và khả năng xử lý dữ liệu phức tạp. Kết quả của bài kiểm tra benchmarking này chứng minh khả năng của Claude Code và tiềm năng của nó để trở thành một model ngôn ngữ hàng đầu trong tương lai.

Tình hình các Agent Framework: Chọn Runtime phù hợp cho việc triển khai AI trong doanh nghiệp

3

State of Agent Frameworks: Choosing the Right Runtime for Enterprise AI Execution

Medium [Đọc bài viết →](#)

Bài viết "Trạng thái của các Framework Trình điều khiển: Chọn Thời gian chạy Đúng cho Thực thi AI Doanh nghiệp" cung cấp cái nhìn tổng quan về trạng thái hiện tại của các framework trình điều khiển cho thực thi AI doanh nghiệp. Các framework trình điều khiển là các nền tảng phần mềm cho phép tích hợp trí tuệ nhân tạo (AI) và các model học máy (ML) vào các ứng dụng kinh doanh. Các framework này cung cấp một môi trường thời gian chạy cho các model AI để thực thi và tương tác với các hệ thống khác. Bài viết nhấn mạnh tầm quan trọng của việc chọn framework trình điều khiển đúng cho thực thi AI doanh nghiệp, xem xét các yếu tố như khả năng mở rộng, bảo mật và tích hợp với các hệ thống hiện có. Nó cũng thảo luận về các đặc điểm chính của các framework trình điều khiển phổ biến, bao gồm Rasa, Microsoft Bot Framework và Google Dialogflow. Bài viết nhằm giúp các nhà phát triển và tổ chức chọn framework trình điều khiển phù hợp nhất cho các dự án AI của họ, dựa trên nhu cầu và yêu cầu cụ thể của họ. Bằng cách hiểu rõ điểm mạnh và điểm yếu của từng

framework, các nhà phát triển có thể đưa ra quyết định thông minh và xây dựng các ứng dụng được hỗ trợ bởi AI hiệu quả hơn.

4

Xem trước GPT-5.6 Sol: một model thế hệ tiếp theo

Previewing GPT-5.6 Sol: a next-generation model

OpenAI Blog [Đọc bài viết →](#)

OpenAI đã ra mắt bản xem trước của mô hình mới nhất của họ, GPT-5.6 Sol, một hệ thống AI thế hệ tiếp theo được thiết kế để vượt trội trong nhiều lĩnh vực. Mô hình tiên tiến này có khả năng nâng cao trong lĩnh vực lập trình, khoa học và an ninh mạng, khiến nó trở thành một công cụ mạnh mẽ cho các chuyên gia và tổ chức trong các lĩnh vực này. Ngoài hiệu suất được cải thiện, GPT-5.6 Sol cũng được trang bị bộ tính năng an toàn tiên tiến nhất của OpenAI. Tính năng an toàn này có khả năng được thiết kế để giảm thiểu các rủi ro tiềm ẩn liên quan đến hệ thống AI, đảm bảo một trải nghiệm an toàn và đáng tin cậy hơn. Việc ra mắt GPT-5.6 Sol cho thấy OpenAI đang đẩy ranh giới của công nghệ AI, nhằm tạo ra các mô hình tinh vi và hữu ích hơn có thể mang lại lợi ích cho nhiều ngành công nghiệp và ứng dụng khác nhau. Mặc dù các chi tiết cụ thể về khả năng và tính năng của mô hình chưa được công bố, bản xem trước của GPT-5.6 Sol cho thấy một bước tiến đáng kể trong sự phát triển của AI.

5

Điểm tin: Sóng nhiệt kỷ lục và những hạn chế chưa từng có từ OpenAI

The Download: brain-melting heatwaves and unprecedented OpenAI restrictions

MIT Tech Review [Đọc bài viết →](#)

Một đợt sóng nhiệt khắc nghiệt đang ảnh hưởng 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 này đang gây ra gián đoạn cho nông nghiệp, cơ sở hạ tầng và hệ thống chăm sóc sức khỏe. Nghiên cứu cho thấy rằng nhiệt độ tăng cao có thể có tác động tiêu cực đến não bộ, dẫn đến tăng tính dễ cáu gắt và bạo lực, cũng như khó khăn trong việc tập trung cho lính cứu hỏa. Tác động của nhiệt đối với não bộ vẫn chưa được hiểu đầy đủ, nhưng các nghiên cứu trên động vật thí nghiệm đã chỉ ra rằng nhiệt độ quá cao có thể thay đổi chức năng của tín hiệu hóa học trong não. Trong tin tức khác, chính quyền Trump đã yêu cầu OpenAI hạn chế việc phát hành mô hình tiếp theo của họ, GPT

5.6, và kiểm tra người dùng đầu tiên trước khi ra mắt rộng rãi. Đây là lần đầu tiên một công ty Mỹ bị yêu cầu hạn chế một mô hình AI trước khi phát hành. Ngoài ra, Apple và Xbox đã tăng giá cho một số sản phẩm của họ do chi phí chip tăng cao được động bởi AI, điều này đã dẫn đến tình trạng khan hiếm và tăng giá cho bộ nhớ và lưu trữ.

6

Google Interactions API: Công nghệ AI thu hẹp khoảng cách phối hợp

Google Interactions API: AI Technology That Closes the Coordination Gap

Dev.to AI [Đọc bài viết →](#)

Google đã bổ sung khả dụng chung của Interactions API, một điểm cuối thống nhất để tương tác với các mô hình Gemini và tác nhân. API này hiện là giao diện chính cho cả mô hình Gemini và tác nhân, thay thế nhu cầu của các nhóm phải xây dựng các lớp điều phối của riêng họ bằng cách sử dụng các công cụ như LangChain và hàng đợi tùy chỉnh. Interactions API cung cấp một loạt các khả năng chính, bao gồm Managed Agents, trạng thái phía máy chủ, thực thi nền, kết hợp công cụ và tạo đa phương thức. Nó cho phép thực hiện một cuộc gọi API duy nhất để cung cấp một hộp cát Linux từ xa có thể duyệt, mã hóa và quản lý tệp. Sự thay đổi này trong cơ sở hạ tầng nhằm giải quyết "khoảng trống phối hợp AI", độ tin cậy và chi phí kỹ thuật phát sinh từ việc quản lý trạng thái, định tuyến công cụ và chuyển giao tác nhân trong các hệ thống AI. Bằng cách di chuyển lớp phối hợp này sang cơ sở hạ tầng của Google, các nhóm có thể đơn giản hóa các quy trình làm việc của họ và giảm độ phức tạp khi tích hợp các mô hình AI vào các ứng dụng của họ.

7

OpenAI hạn chế phát hành GPT-5.6 theo yêu cầu của chính phủ, khẳng định các hạn chế không nên là thông lệ

OpenAI limits GPT-5.6 rollout after government request, says restrictions shouldn't be the norm

TechCrunch AI [Đọc bài viết →](#)

OpenAI đã tạm thời hạn chế việc phát hành các mô hình AI mới nhất của mình, GPT-5.6, cho một nhóm nhỏ các đối tác đáng tin cậy do yêu cầu từ chính phủ Mỹ. Các mô hình này, bao gồm Sol, Terra và Luna, đã được lên kế hoạch phát hành cho công chúng nhưng giờ chỉ có sẵn cho các đối tác đã được chính phủ kiểm duyệt. Quyết định này được đưa ra khi chính phủ Mỹ tăng áp lực lên các công ty AI để hạn chế truy cập

vào các hệ thống tiên tiến nhất của họ. OpenAI đã thể hiện sự không hài lòng với thỏa thuận này, tuyên bố rằng họ không tin rằng các quy trình truy cập của chính phủ nên trở thành mặc định lâu dài. Công ty cho rằng hạn chế này sẽ giữ các công cụ tốt nhất khỏi người dùng và các nhà phát triển cần chúng. GPT-5.6 Sol là mô hình mạnh nhất của OpenAI cho đến nay, với khả năng cải tiến trong lĩnh vực mã hóa, sinh học và an ninh mạng. Công ty đang làm việc với chính quyền để phát triển một khuôn khổ mới cho việc phát hành mô hình trong tương lai và đã nhấn mạnh các tính năng bảo mật mạnh mẽ của mô hình mới, bao gồm khả năng ưu tiên công việc an ninh mạng phòng thủ hơn các lỗ hổng tấn công.

8 **GitHub và UNDP hợp tác thúc đẩy các ưu tiên phát triển tại Ghana bằng open source**

GitHub and UNDP team up to advance development priorities in Ghana with open source

GitHub Blog [Đọc bài viết →](#)

GitHub, một nền tảng hàng đầu cho các nhà phát triển, đã hợp tác với Chương trình Phát triển Liên Hợp Quốc (UNDP) để thúc đẩy các ưu tiên phát triển tại Ghana thông qua việc sử dụng phần mềm mã nguồn mở. Đối tác này nhằm khám phá cách quản trị mã nguồn mở có thể hỗ trợ nỗ lực cải cách số đầy tham vọng của Ghana. Bộ Truyền thông, Công nghệ Kỹ thuật số và Đổi mới (MoCDTI) của Ghana đang thực hiện một cuộc tái thiết số để tạo việc làm, phát triển doanh nghiệp và trao quyền cho giới trẻ. Nỗ lực này bao gồm các cải cách lập pháp trong các lĩnh vực như an ninh mạng, bảo vệ dữ liệu và công nghệ mới nổi. Đối tác giữa GitHub và UNDP sẽ giúp Ghana tận dụng các phương pháp mã nguồn mở để thúc đẩy tác động phát triển và cải thiện cuộc sống của công dân. Bằng cách áp dụng phần mềm mã nguồn mở, Ghana có thể hưởng lợi từ một phương pháp tiếp cận hợp tác và thích nghi với phát triển phần mềm, điều này có thể dẫn đến các giải pháp hiệu quả và hiệu quả hơn. Đối tác này là một bước quan trọng hướng tới thúc đẩy chuyển đổi số và tăng trưởng kinh tế tại Ghana.

TIPS & TRICKS CHO DEV

Sử dụng LangSmith

Vấn đề: Khó theo dõi hiệu suất của mô hình AI.

Cách làm: Sử dụng LangSmith để theo dõi và phân tích hiệu suất của mô hình. Ví

dụ, tạo dashboard với `langsmith create-dashboard`.

Đánh giá: Hiệu quả trong việc theo dõi mô hình, nên dùng khi cần tối ưu hóa.

Tích hợp Langfuse

Vấn đề: Chi phí sử dụng mô hình AI quá cao.

Cách làm: Tích hợp Langfuse để tối ưu hóa chi phí. Ví dụ, sử dụng `langfuse optimize-model` để giảm chi phí.

Đánh giá: Giúp giảm chi phí, nên dùng khi cần tiết kiệm.

Arize Phoenix

Vấn đề: Khó debug và theo dõi mô hình AI.

Cách làm: Sử dụng Arize Phoenix để theo dõi và debug mô hình. Ví dụ, sử dụng `arize phoenix log` để xem log.

Đánh giá: Hiệu quả trong việc debug, nên dùng khi cần tìm 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 tối ưu hóa kiến trúc và đào tạo mô hình. Điều này giúp giảm thiểu chi phí tính toán và tăng tốc độ xử lý.

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 đào tạo.

4. Tip hoặc bước tiếp theo: Sử dụng thư viện như Hugging Face Transformers để thực hiện fine-tuning và LoRA cho mô hình LLM của bạn, giúp giảm chi phí và tăng hiệu năng.

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

Nguồn: Google News · Groq AI