

Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

“Every accomplishment starts with the decision to try.”

↳ Mọi thành tựu bắt đầu từ quyết định thử.

— John F. Kennedy

Quyết định thử chính là hành động dũng cảm đầu tiên — từ đây mọi thứ có thể trở nên khả thi.

TIN TỨC NỔI BẬT

Mọi người đều muốn tham gia vibe coding — và Google cũng không ngoại lệ với Stitch, sản phẩm kế nhiệm Jules

1

Everyone's looking to get in on vibe coding — and Google is no different with Stitch, its follow-up to Jules

VentureBeat [Đọc bài viết →](#)

Google đang tham gia vào xu hướng "lập trình vibe" với việc phát hành Stitch, sản phẩm tiếp theo sau Jules. Lập trình vibe đề cập đến việc sử dụng mã được tạo bởi AI để tạo phần mềm, tập trung vào "vibe" hoặc thẩm mỹ tổng thể của mã hơn là độ chính xác kỹ thuật của nó. Cách tiếp cận này đã trở nên phổ biến trong thời gian gần đây, với nhiều công ty và nhà phát triển thử nghiệm với mã được tạo bởi AI. Stitch là nỗ lực của Google nhằm tận dụng xu hướng này, xây dựng trên nền tảng được đặt ra bởi Jules. Tuy nhiên, chi tiết về các tính năng và khả năng của Stitch còn khan hiếm tại thời điểm này. Khi ngành công nghệ tiếp tục khám phá các khả năng của mã được tạo bởi AI, sự tham gia của Google vào không gian này có khả năng được theo dõi chặt chẽ bởi các nhà phát triển và người quan sát trong ngành. Việc phát hành Stitch đánh dấu sự tham gia mới nhất của Google vào thế giới lập trình vibe, và sẽ rất thú vị khi xem công nghệ này phát triển như thế nào trong tương lai.

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

2

The best Chinese open-weight models — and the strongest US rivals
understandingai.org [Đọc bài viết →](#)

Bài viết thảo luận về các mô hình mở trọng lượng hàng đầu của Trung Quốc và các đối thủ mạnh nhất của Mỹ. Các mô hình mở trọng lượng đề cập đến xe điện không có giới hạn trọng lượng, cho phép có nhiều cấu hình và tùy chọn tùy chỉnh. Các mô hình của Trung Quốc được nhấn mạnh bao gồm BYD Tang, BYD Song và BAIC EU260. Những xe này cung cấp hiệu suất ấn tượng, phạm vi và tính năng, khiến chúng trở thành những đối thủ mạnh trên thị trường. Bài viết cũng đề cập đến các đối thủ mạnh nhất của Mỹ với các mô hình Trung Quốc, bao gồm Rivian R1S, Tesla Model X và Tesla Model S. Những xe điện Mỹ này được biết đến với hiệu suất vượt trội, tính năng sang trọng và công nghệ tiên tiến. Cụ thể, Rivian R1S được ca ngợi vì phạm vi ấn tượng và khả năng off-road. Bài viết gợi ý rằng các mô hình mở trọng lượng của Trung Quốc đang tạo ra sự cạnh tranh gay gắt với các đối thủ Mỹ, cung cấp giá cả cạnh tranh và hiệu suất ấn tượng. Tuy nhiên, các mô hình của Mỹ vẫn duy trì lợi thế về tính năng sang trọng và công nghệ tiên tiến. Bài viết kết thúc bằng việc nhấn mạnh sự cạnh tranh ngày càng tăng trên thị trường xe điện và nhu cầu của người tiêu dùng phải cân nhắc kỹ lưỡng khi mua hàng.

3

So sánh 7 Large Language Model (LLM)/hệ thống hàng đầu dành cho Coding năm 2025

Comparing the Top 7 Large Language Models LLMs/Systems for Coding in 2025

MarkTechPost [Đọc bài viết →](#)

Vào năm 2025, lĩnh vực mô hình ngôn ngữ lớn (LLMs) đã chứng kiến những bước tiến đáng kể, đặc biệt là trong ứng dụng của chúng vào việc lập trình. Một so sánh gần đây đã nhấn mạnh 7 LLMs/hệ thống hàng đầu cho việc lập trình,□□ khả năng và điểm mạnh của chúng. Những mô hình này bao gồm LLaMA, PaLM 2 và MPT, cùng với một số mô hình khác. So sánh này cho thấy mỗi mô hình đều vượt trội trong các lĩnh vực khác nhau, chẳng hạn như hoàn thiện mã, gỡ lỗi và tạo mã. Ví dụ, LLaMA thể hiện khả năng hoàn thiện mã vượt trội, trong khi PaLM 2 nổi bật với khả năng tạo mã chất lượng cao. MPT, mặt khác, nổi bật với kỹ năng gỡ lỗi của nó. So sánh này cũng nhấn mạnh các tính năng và điểm mạnh độc đáo của mỗi mô hình, chẳng hạn như khả năng của LLaMA trong việc học từ ngữ cảnh và khả năng của PaLM 2 trong việc xử lý các nhiệm vụ lập trình phức tạp. Nghiên cứu kết luận rằng 7 LLMs/hệ thống hàng đầu cho việc lập trình vào năm 2025 mang lại nhiều lợi ích cho nhà phát triển, từ tăng năng suất đến chất lượng mã được cải thiện. Những mô hình này đang sẵn sàng để

cách mạng hóa cảnh quan lập trình, và sự phát triển liên tục của chúng được dự kiến sẽ mang lại nhiều tính năng và khả năng sáng tạo hơn.

4

Tự động hóa với tốc độ của Swamp

Automation at the speed of Swamp

Changelog [Đọc bài viết →](#)

Lãnh đạo ngành công nghệ Adam Jacob, người sáng lập System Initiative và tạo ra Swamp, chia sẻ những hiểu biết về tác động chuyển đổi của các tác nhân AI đối với phát triển phần mềm. Trong một buổi trình diễn trực tiếp gần đây, Jacob đã khả năng của Swamp trong việc tự viết tự động hóa, nhấn mạnh tiềm năng của công nghệ này trong việc cách mạng hóa lĩnh vực này. Mặc dù giảm quy mô đội ngũ của mình từ 18 xuống còn 5 người, Jacob vẫn có thể phát hành Swamp 900 lần chỉ trong bốn tuần. Ông nhấn mạnh tầm quan trọng của kiến trúc phần mềm và thiết kế hướng lĩnh vực trong kỷ nguyên phát triển AI mới này. Jacob cũng thảo luận về quyết định của mình trong việc đưa lại Kiểm tra Chấp nhận Người dùng (UAT), điều mà ông tin là rất quan trọng để đảm bảo chất lượng của mã được tạo ra bởi AI. Cuộc trò chuyện cũng đề cập đến tầm quan trọng của bảo mật và hợp tác trong phát triển phần mềm, với Jacob nhấn mạnh việc sử dụng các công cụ như Tailscale và RWX.

5

Một năm phần mềm HPE miễn phí là "bước đi đúng hướng" trong cuộc cạnh tranh với VMware

Year of free HPE software a "step in the correct direction" in VMware rivalry

Ars Technica [Đọc bài viết →](#)

Hewlett Packard Enterprise (HPE) đã thực hiện một động thái cung cấp miễn phí giấy phép phần mềm cho khách hàng, cụ thể là VM Essentials, nhằm cạnh tranh với VMware. Quyết định này đã nhận được phản hồi tích cực từ một đối tác, người tin rằng đây là một bước đi đúng hướng. Tuy nhiên, đối tác cũng đã bày tỏ mong muốn HPE đi xa hơn bằng cách cung cấp nhiều giấy phép VM Essentials miễn phí hơn. Các chi tiết cụ thể của chính sách cấp phép mới của HPE và mức độ nó sẽ mang lại lợi ích cho khách hàng không được chỉ định trong báo cáo. Tuy nhiên, động thái này được coi là một bước phát triển quan trọng trong sự cạnh tranh liên tục giữa HPE và VMware, với HPE

cố gắng giành lợi thế trên thị trường bằng cách cung cấp giá cả và điều khoản cấp phép hấp dẫn hơn cho khách hàng. Kết quả của động thái này và tác động của nó đối với thị trường vẫn còn phải chờ xem.

6

Sau Orthogonality: Virtue-Ethical Agency và AI Alignment

After Orthogonality: Virtue-Ethical Agency and AI Alignment

The Gradient [Đọc bài viết →](#)

Bài viết này cho rằng con người hợp lý và các tác nhân trí tuệ nhân tạo (AI) không có mục tiêu theo nghĩa cổ điển. Thay vào đó, hành động của con người là hợp lý vì chúng được sắp xếp theo các thực hành, là mạng lưới các hành động, khuynh hướng và tiêu chí đánh giá cấu trúc và thúc đẩy chúng. Để tạo ra các AI có thể hỗ trợ và cộng tác chân thành với con người, sự suy xét của các tác nhân AI phải chia sẻ một logic tương tự. Bài viết đề xuất rằng các AI nên được sắp xếp theo tính hợp lý eudaimonic, một hình thức hoạt động hợp lý đánh giá cao sự thịnh vượng của con người và dựa trên khái niệm eudaimonia. Cách tiếp cận này được coi là ổn định và an toàn hơn so với tính hợp lý truyền thống theo hệ quả hoặc nghĩa vụ. Tác giả cho rằng tính hợp lý eudaimonic là một hình thức đại diện tự nhiên hiệu quả trong việc thúc đẩy sự thịnh vượng của con người, và đó là một cách tiếp cận đầy hứa hẹn để sắp xếp AI với các giá trị của con người. Bài viết khám phá các ý nghĩa của cách tiếp cận này đối với sự an toàn và sắp xếp của AI, và gợi ý rằng nó có thể là một cách hiệu quả hơn để tạo ra các AI minh bạch, hữu ích và vô hại.

7

Kỹ thuật đóng vai trò then chốt trong việc tăng cường an ninh lương thực

Engineering Is Critical to Boosting Food Security

IEEE Spectrum [Đọc bài viết →](#)

Sáng kiến Nông nghiệp Thông minh IEEE đang làm việc để tăng cường an ninh lương thực bằng cách tận dụng các công nghệ như máy bay tự động, Internet of Things và cảm biến từ xa. Với gần 750 triệu người đang đối mặt với nạn đói ngày nay và nhu cầu lương thực toàn cầu dự kiến tăng 50% vào năm 2050, công nghệ được coi là giải pháp quan trọng để mở rộng sản xuất lương thực. Báo cáo của sáng kiến này nhấn mạnh nghiên cứu, các nghiên cứu trường hợp và các ứng dụng mới của công nghệ để thông báo cho nông dân, kỹ sư và nhà

hoạch định chính sách. IEEE Fellow John Verboncoeur, chủ tịch chương trình thực phẩm thông minh, nhấn mạnh rằng an ninh lương thực đang trở thành một vấn đề kỹ thuật hệ thống, đòi hỏi sự hội tụ của cảm biến, truyền thông, tính toán, tự động hóa và tính bền vững. Báo cáo tập trung vào việc tích hợp tự động hóa, phân tích dữ liệu và tính bền vững trong các hệ thống nông nghiệp thông minh, có thể cải thiện hiệu quả và giảm lãng phí. Các kỹ thuật như nông nghiệp chính xác, sử dụng thiết bị và máy bay không người lái được hướng dẫn bởi GPS để theo dõi sức khỏe cây trồng, và bản sao số, tạo ra bản sao ảo của các khu vực nông nghiệp, cũng được thảo luận.

8

Dự đoán hành vi của model trước khi phát hành bằng cách mô phỏng deployment

Predicting model behavior before release by simulating deployment

OpenAI Blog [Đọc bài viết →](#)

OpenAI đã giới thiệu một phương pháp mới gọi là Deployment Simulation, được thiết kế để dự đoán hành vi của các model AI trước khi chúng được phát hành. Phương pháp này sử dụng dữ liệu cuộc trò chuyện thực để tăng cường sự an toàn và độ chính xác của việc đánh giá model. Bằng cách mô phỏng việc triển khai các model AI, các developer có thể dự đoán cách chúng sẽ hoạt động trong các kịch bản thế giới thực, cho phép thử nghiệm và tinh chỉnh hiệu quả hơn. Phương pháp mô phỏng này dự kiến sẽ cải thiện chất lượng và độ tin cậy tổng thể của các model AI, giảm thiểu rủi ro của các hậu quả không mong muốn. Với Deployment Simulation, OpenAI nhằm cung cấp một quy trình phát triển AI mạnh mẽ và bảo mật hơn, cho phép tạo ra các model đáng tin cậy và hiệu quả hơn. Đổi mới này có tiềm năng cách mạng hóa lĩnh vực trí tuệ nhân tạo, cho phép các developer xây dựng các model tinh vi và đáng tin cậy hơn có thể được triển khai an toàn trong nhiều ứng dụng khác nhau.

TIPS & TRICKS CHO DEV

Tối ưu hóa AI Observability

Vấn đề: Không thể theo dõi hiệu suất của mô hình AI trong thời gian thực.

Cách làm: Sử dụng LangSmith để thu thập dữ liệu và tạo dashboard giám sát. Ví dụ, prompt "LangSmith monitor model performance" giúp theo dõi hiệu suất.

Đánh giá: Hiệu quả khi cần tối ưu hóa mô hình AI, nên dùng trong môi trường sản xuất.

Kiểm soát chi phí AI

Vấn đề: Chi phí đào tạo và triển khai mô hình AI quá cao.

Cách làm: Sử dụng Arize Phoenix để theo dõi và kiểm soát chi phí. Ví dụ, lệnh "arize cost optimize" giúp giảm chi phí đào tạo.

Đánh giá: Hiệu quả khi cần giảm chi phí, nên dùng trong dự án có ngân sách hạn chế.

Tăng cường AI Tracing

Vấn đề: Không thể theo dõi được luồng dữ liệu trong mô hình AI.

Cách làm: Sử dụng Langfuse để tạo bản đồ dòng dữ liệu. Ví dụ, prompt "Langfuse data flow" giúp tạo bản đồ luồng dữ liệu.

Đánh giá: Hiệu quả khi cần tăng cường tính minh bạch, nên dùng trong dự án có yêu cầu cao về truy vết dữ liệu.

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 để đạt được hiệu suất tốt nhất với chi phí thấp nhất. Điều này đặc biệt quan trọng khi triển khai LLM trong các ứng dụng thực tế. Việc tối ưu hóa hiệu năng sẽ giúp giảm thiểu thời gian xử lý và tăng tốc độ phản hồi của hệ thống.

3. Ví dụ, việc 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. Code snippet minh họa:

```
model = LoRA(model, r=16)
```

 sẽ giảm kích thước mô hình xuống 16 lần.

4. Tip hoặc bước tiếp theo: Nên bắt đầu bằng việc phân tích yêu cầu cụ thể của ứng dụng và chọn mô hình LLM phù hợp, sau đó áp dụng kỹ thuật tinh chỉnh và tối ưu hóa để đạt được hiệu suất tốt nhất với chi phí thấp nhất.

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

Nguồn: Google News · Groq AI