

Bản Tin AI Hằng Ngày

Cập nhật công nghệ AI mới nhất

“Ăn quả nhớ kẻ trồng cây.”

— Tục ngữ Việt Nam

Luôn ghi nhớ và biết ơn những người đã tạo ra thành quả mà ta đang được hưởng.

TIN TỨC NỔI BẬT

1

Khám phá AI mới nhất của Anthropic: Điều gì được hé lộ và điều gì còn bỏ ngỏ

What Anthropic's latest AI discovery does—and doesn't—show

MIT Tech Review

[Đọc bài viết →](#)

Anthropic, công ty AI giá trị nhất thế giới, đã có một khám phá mới về các large language model (LLM) của mình. Công ty tuyên bố đã tìm thấy một "cửa sổ nhìn vào suy nghĩ nội tại" của các model, cho phép họ hiểu cách các model đưa ra câu trả lời. Bước đột phá này là một phần trong trọng tâm của Anthropic về mechanistic interpretability, bao gồm việc phân tích các phép toán phức tạp đằng sau các AI model để tìm hiểu lý do chúng tạo ra các output nhất định. Nghiên cứu của Anthropic đã tiết lộ một "J-space" bên trong các LLM của họ, một không gian chứa đầy các từ không xuất hiện trong output nhưng lại ảnh hưởng đến quá trình ra quyết định của model. Những từ này có thể theo dõi tiến độ của model, đại diện cho sự nhận biết, hoặc đóng vai trò như bình luận nội bộ về các quyết định của model. Công ty cũng phát hiện ra rằng các LLM có thể mô tả và thao tác các từ trong không gian này, cho thấy chúng đang sử dụng nó. Khám phá này có ý nghĩa quan trọng vì nó cung cấp sự hiểu biết sâu sắc hơn về cách các LLM hoạt động, điều cần thiết để kiểm soát và cải thiện hiệu suất của chúng. CEO của Anthropic đã tuyên bố rằng việc tìm hiểu thêm về cách các LLM hoạt động là rất quan trọng để kiểm soát hoàn toàn chúng. Nghiên cứu của công ty đang tiếp diễn, và phát hiện mới này là một bước tiến tới việc đạt được mục tiêu đó.

2

Máy quay PV-460 của Panasonic: Giải pháp ổn định video rung lắc

Panasonic's PV-460 Camcorder Stabilized Shaky Videos

IEEE Spectrum

[Đọc bài viết →](#)

Máy quay PV-460 của Panasonic đã cách mạng hóa việc quay video tại nhà. Ra mắt vào năm 1988, PV-460 là chiếc máy quay video đầu tiên tích hợp bộ ổn định hình ảnh quang học (optical image stabilizer), giúp giảm đáng kể các cảnh quay rung lắc do chuyển động tự nhiên của tay và cơ thể. Đổi mới này đã thay đổi lĩnh vực quay phim cá nhân, giúp mọi người dễ dàng ghi lại các video chất lượng cao về những buổi họp mặt gia đình, kỳ nghỉ và sự kiện. Tính năng ổn định hình ảnh của PV-460 đã dân chủ hóa việc tạo video, cho phép những người bình thường thể hiện bản thân với sự tự do sáng tạo mới mẻ. Công nghệ này cũng có các ứng dụng quan trọng trong các lĩnh vực chuyên biệt, góp phần vào những tiến bộ trong truyền thông giáo dục và y tế từ xa (telemedicine). PV-460 đã được vinh danh là một IEEE Milestone vào ngày 9 tháng 7, công nhận tác động của nó đối với sự phát triển của công nghệ video. Việc ra mắt chiếc camcorder này dựa trên những đổi mới trước đó, bao gồm công nghệ phơi sáng tự động (automatic exposure) và lấy nét tự động (autofocus), được phát triển lần lượt vào những năm 1950 và 1970.

3

Chính phủ Mỹ cảnh báo: Hacker nhà nước Nga đang nhắm vào router của bạn

The US government warns that Russia state hackers are coming after your router

Ars Technica

[Đọc bài viết →](#)

Chính phủ Mỹ, thông qua Cơ quan An ninh Cơ sở hạ tầng và An ninh mạng (CISA), đã đưa ra cảnh báo cho người dùng router tại nhà và văn phòng nhỏ về việc bảo mật thiết bị của họ do các hacker nhà nước Nga đang liên tục cố gắng xâm nhập. Các hacker này, được theo dõi dưới nhiều tên gọi khác nhau, đã và đang khai thác các thiết bị networking được cấu hình kém và dễ bị tấn công trên toàn thế giới để giành quyền kiểm soát các mạng lưới thuộc lĩnh vực cơ sở hạ tầng quan trọng. Phương thức xâm nhập chính bao gồm việc hacker quét các dải IP có các agent của Simple Network Management Protocol (SNMP) đang hoạt động và chấp nhận thông tin xác thực (authentication credentials) phổ biến hoặc mặc định. Để tự bảo vệ, người dùng router được khuyến nghị tắt SNMP phiên bản 1 và 2, sử

dụng SNMP phiên bản 3, tắt Cisco Smart Install, sử dụng mật khẩu mạnh, cập nhật firmware thường xuyên và tránh sử dụng các networking protocol khác. Cảnh báo này được đồng phát hành bởi các chính phủ từ khắp nơi trên thế giới, bao gồm Úc, Đan Mạch, New Zealand và Vương quốc Anh.

4

AGI không phải là Multimodal

AGI Is Not Multimodal

The Gradient [Đọc bài viết →](#)

Những tiến bộ gần đây trong các generative AI model đã khiến một số người tin rằng Artificial General Intelligence (AGI) sắp xuất hiện. Tuy nhiên, giả định này có thể sai lầm. Thành công của các model này phần lớn là do khả năng scale trên phần cứng hiện có, hơn là một cách tiếp cận thông minh có tính toán. Cách tiếp cận multimodal, kết hợp nhiều modality để tạo ra một AI tổng quát, khó có thể dẫn đến AGI ở cấp độ con người. Thay vào đó, các nhà nghiên cứu nên tập trung vào các

5

Dữ liệu dành cho Agent

Data for Agents

Hugging Face Blog [Đọc bài viết →](#)

Nemotron của NVIDIA đang giải quyết thách thức xây dựng các tác nhân trí tuệ nhân tạo (AI) bằng cách phát hành các sản phẩm dữ liệu mở. Không giống như trọng lượng mô hình, chỉ là một phần của câu chuyện, dữ liệu mở cho phép tái tạo, kiểm tra và giải thích hành vi của tác nhân. Điều này rất quan trọng vì các tác nhân AI tương tác với các hệ thống, công cụ và quy trình làm việc khác nhau, và các nhà phát triển cần hiểu dữ liệu đã định hình hành vi của chúng. Để mở rộng nghiên cứu AI, NVIDIA nhấn mạnh tầm quan trọng của dữ liệu tổng hợp, cho phép các đội bảo tồn các tín hiệu hữu ích mà không lộ nguồn gốc cơ bản. Cách tiếp cận này cho phép các công ty đóng góp vào một lớp dữ liệu chung phong phú hơn mà không ảnh hưởng đến lợi thế cạnh tranh của họ. NVIDIA đã phát hành hơn 10 nghìn tỷ token trước khi đào tạo và hàng triệu mẫu sau khi đào tạo trên các lĩnh vực và hình dạng dữ liệu khác nhau. Để giúp khám phá dữ liệu này, công ty đã phát triển Nemotron Post-Training v3 Prompt Atlas, một bản đồ trực quan tương tác cho phép người dùng tái tổ chức và kiểm tra các

ví dụ đại diện. Bài viết cũng nhấn mạnh tầm quan trọng của dữ liệu được nền tảng hóa cục bộ, lấy vào account các sắc thái của các ngôn ngữ và văn hóa khác nhau. Nemotron-Personas là một nỗ lực để giải quyết thách thức này bằng cách cung cấp dữ liệu tổng hợp được nền tảng hóa cục bộ. Bằng cách làm cho dữ liệu trở nên dễ tiếp cận và đa dạng hơn, NVIDIA nhằm mục đích nuôi dưỡng một hệ sinh thái AI tham gia, nơi các bên liên quan khác nhau có thể đóng góp và thúc đẩy nghiên cứu AI thông qua API, framework và LLM.

TIPS & TRICKS CHO DEV

Tối ưu hóa RAG

Vấn đề: Việc tạo nội dung mới tốn nhiều thời gian và công sức.

Cách làm: Sử dụng RAG với prompt như "Tạo một bài viết về chủ đề XYZ" và kết hợp với embeddings để tăng tốc độ tìm kiếm. Ví dụ: "Generate a blog post about AI applications in healthcare".

Đánh giá: Hiệu quả cao trong tạo nội dung mới, nhưng cần cân nhắc chất lượng và chính xác.

Sử dụng Embeddings

Vấn đề: Tìm kiếm thông tin không chính xác due to từ khóa không phù hợp.

Cách làm: Sử dụng embeddings như Sentence-BERT để mã hóa văn bản và tìm kiếm dựa trên ý nghĩa. Ví dụ: `sentence-transformers encode --input "Chủ đề XYZ"`.

Đánh giá: Hiệu quả cao trong tìm kiếm thông tin liên quan, nhưng cần có dữ liệu chất lượng.

Semantic Search

Vấn đề: Tìm kiếm thông tin không đầy đủ do thiếu thông tin ngữ nghĩa.

Cách làm: Sử dụng semantic search với prompt như "Tìm kiếm thông tin về chủ đề XYZ" và kết hợp với embeddings để tăng độ chính xác.

Đánh giá: Hiệu quả cao trong tìm kiếm thông tin chi tiết, nhưng cần có kiến thức về ngữ nghĩa và embeddings.

BÀI HỌC AI HÔM NAY CHO DEV

1. Tối ưu chi phí & hiệu năng LLM

2. Để phát triển ứng dụng hiệu quả, các nhà phát triển cần tối ưu hóa 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 phản hồi và giảm thiểu tài nguyên cần thiết.

3. Ví dụ, việc sử dụng kỹ thuật fine-tuning và LoRA (Low-Rank Adaptation of Weights) có thể giúp giảm kích thước mô hình và cải thiện hiệu năng. Code snippet minh họa: `model = LLM.from_pretrained('base_model');`
`model.fine_tune(dataset, epochs=5)`.

4. Tip hoặc bước tiếp theo: Hãy thử nghiệm với các kỹ thuật tối ưu hóa khác nhau, như quantization và pruning, để tìm ra phương pháp phù hợp nhất cho ứng dụng của bạn.

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

Nguồn: Google News · Groq AI