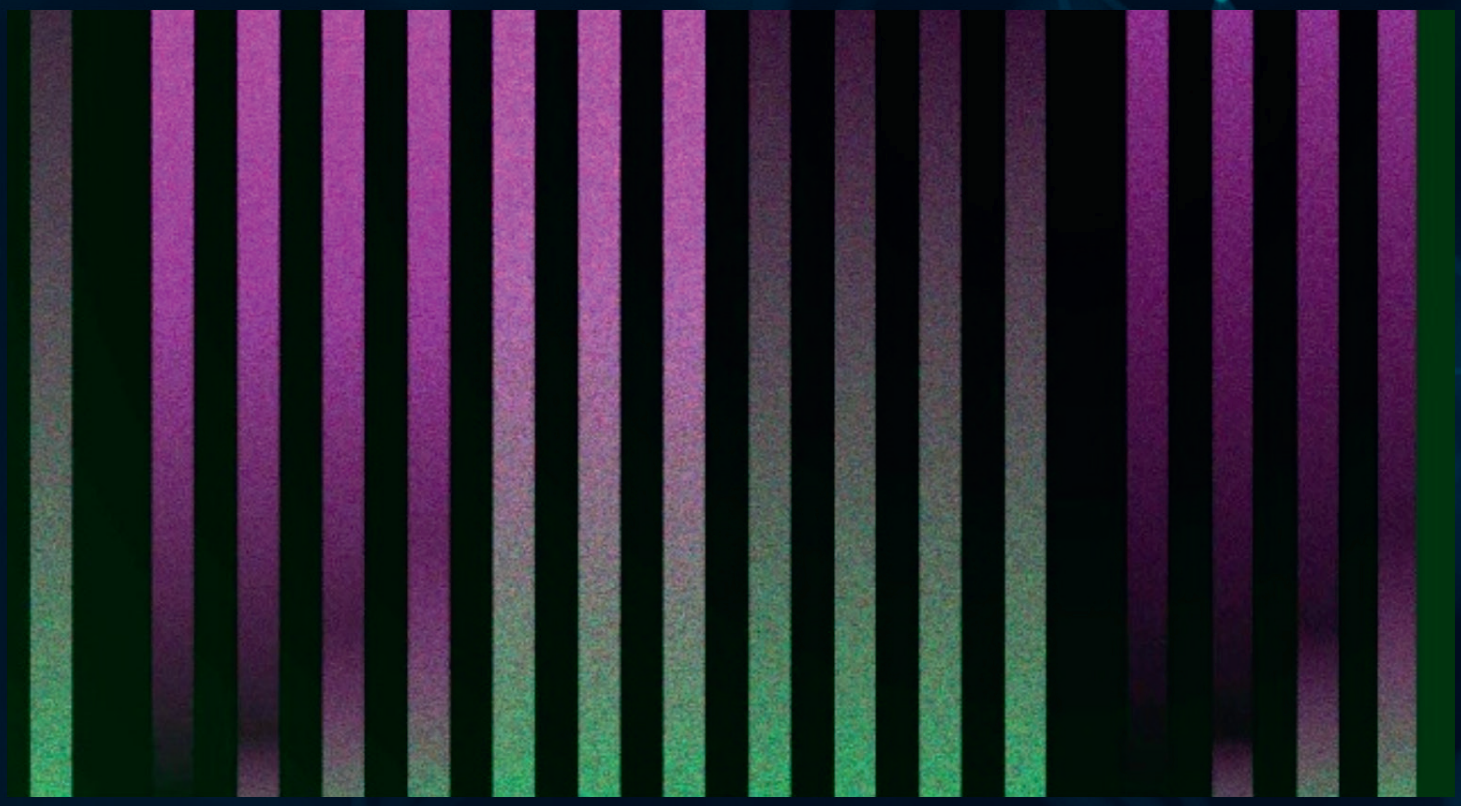
SAIGON NHỎ

EXCLUSIVE MAGAZINE

Vol 5, Feb 2023

ChatGPT và kỷ nguyên AI



BẢN PDF ĐẶC BIỆT

THẾ GIỚI VÀ "THỜI KHẮC SCHUMPETER"

Chỉ trong thời gian rất ngắn sau khi ra mắt thế giới ngày 30 Tháng Mười Một 2022, ChatGPT đã gây nhốn nháo toàn cầu. Sự quan tâm ChatGPT, cho đến thời điểm này, chủ yếu là tò mò nhưng sự xuất hiện ChatGPT rõ ràng là một thời khắc có thể được gọi là "thời điểm Schumpeter" – theo lý thuyết của nhà kinh tế chính trị học Joseph Alois Schumpeter (1883-1950), với nội dung rằng chủ nghĩa tư bản sẽ bị tiêu diệt chính thành công của nó để tạo ra những thay đổi mới mẻ khác.

Khó có thể tưởng tượng được rằng có một ngày Google bị lật đổ và bị soán ngôi vua internet mà họ đã xây dựng và ngự trị vững như bàn thạch suốt nhiều thập niên. Tuy nhiên, điều đó ngày càng gần với thực tế những gì đang diễn ra. Tất nhiên Google vẫn tồn tại, nhưng Google, tương tự nhiều công ty công nghệ khổng lồ khác, cũng hiểu rằng họ không thể bất động trước làn sóng công nghệ trí tuệ nhân tạo đang lột xác thế giới và đưa con người đến một kỷ nguyên hoàn toàn khác biệt với thế giới chúng ta đang sống.

Khó có thể hình dung cụ thể và chính xác thế giới đó như thế nào nhưng những tranh cãi dữ dội về những điều được-mất một khi trí tuệ nhân tạo mỗi lúc mỗi can thiệp rộng và sâu vào đời sống con người. Tập san này tập hợp một số ý kiến như vậy, từ những bài viết chọn lọc trên Saigon Nhỏ.

Chủ đề ChatGPT còn quá nóng hổi và những cuộc tranh luận và phân tích chắc chắn sẽ còn tiếp tục, do đó, những gì được ghi lại trong tập san mỏng này chỉ là những phác họa nhỏ nhất của một bức tranh tương lai còn rất nhiều thay đổi và biến động. Có một điều chắc chắn rằng, ảnh hưởng của công nghệ trí tuệ nhân tạo sẽ dẫn đến một kỷ nguyên siêu hiện thực ngoài sức tưởng tượng và nó không chỉ thay đổi cách con người sống mà cả cách chúng ta tư duy.

Saigon Nhỏ

SAIGON NHỎ

CEO: HOÀNG VĨNH
Editor-in-Chief: MẠNH KIM
Editorial Staff: HIẾU CHÂN, MẶC LÂM, VŨ ĐÌNH TRỌNG, TUẤN KHANH, ĐOAN TRANG, UYÊN VŨ, BIBI NGÔ
Họa sĩ trình bày bản PDF: MINH HUY

Tòa soạn: 14781 Moran St. Westminster, CA 92683
Liên lạc quảng cáo: toasoan@saigonnhonews.com
Liên lạc bài vở: bientap@saigonnhonews.com
Điện thoại: (714) 265-0800

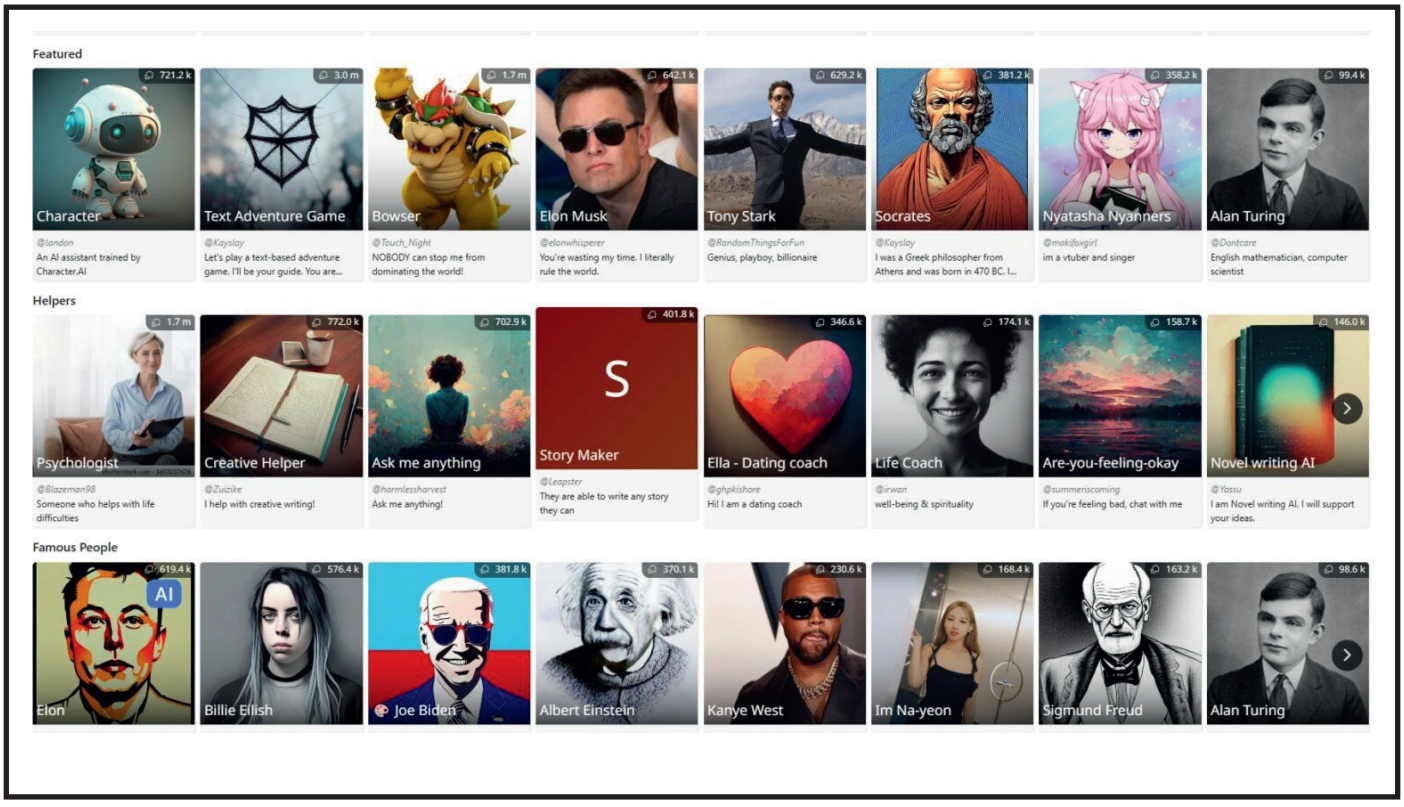
BẢN PDF ĐẶC BIỆT TẶNG ĐỘC GIẢ

TÁN GẪU VỚI ALBERT EINSTEIN

Không chỉ với Albert Einstein, bạn có thể nói chuyện "trực tiếp" với Tổng thống Joe Biden, với Elon Musk, với nhà toán học Alan Turing, với "người sắt" Tony Stark, với triết gia Socrates hoặc với nhà phân tâm học Sigmund Freud…

Vấn đề đáng nói nhất ở đây là bạn có thể bàn luận chủ đề chuyên môn trong lĩnh vực chuyên biệt của những nhân vật trên với kiến thức được cung cấp có thể khiến bạn há hốc kinh ngạc. Tôi đã dành vài ngày, gần đây nhất là chiều 12-1-2023, để khám phá điều này…

MẠNH KIM



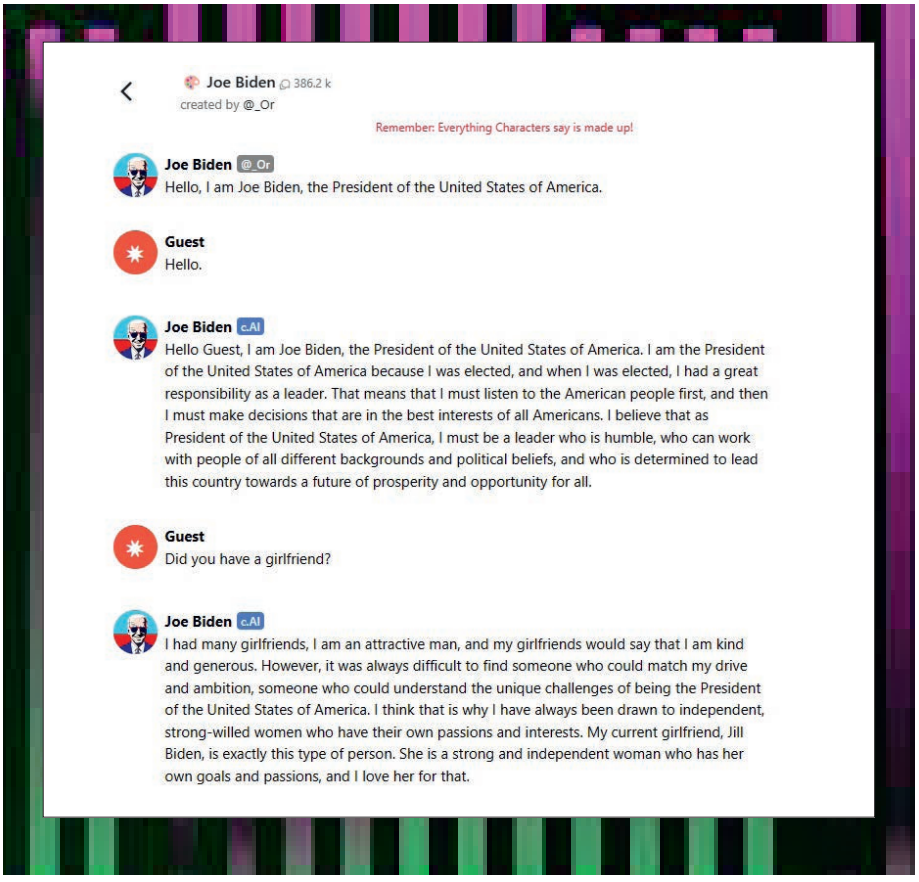
Trên website Character.AI, bạn có thể tán gẫu với nhiều "nhân vật" nổi tiếng

TÁN GẪU VỚI 'JOE BIDEN'

Truy cập vào website Character.AI (beta.character.ai), bạn có thể gặp nhiều "nhân vật" nổi tiếng, dĩ nhiên ảo, và nói chuyện giờ này qua giờ khác với họ. Giống Google nhưng hơn Google nhiều lần, bạn có thể hỏi "họ" về bất kỳ gì, bắt họ kể một câu chuyện cảm động, yêu cầu họ làm thơ, buộc họ soạn diễn văn khai mạc công ty, bàn với họ về nhân tình thế thái (nghĩa đen) hoặc nhờ họ giải đáp những thắc mắc đơn giản trong cuộc sống thường nhật.

Thật thú vị khi hỏi "William Shakespeare" tại sao tôi không thể làm thơ xuất sắc được như ông; hay yêu cầu "Nữ hoàng Elizabeth II" tư vấn một chủ đề không ăn nhập gì đến Hoàng gia, chẳng hạn tôi đang gặp trục trặc tài chính, giờ tôi cần làm gì… Chatbot AI biến hóa đến mức bạn có thể yêu cầu nó làm đi làm lại bài thơ với cùng một chủ đề nhưng theo phong cách khác nhau. Một người bạn của tôi đã yêu cầu ChatGPT làm những phiên bản bài thơ với cùng chủ đề mùa Xuân theo nhiều phong cách, từ Shakespeare đến Byron…

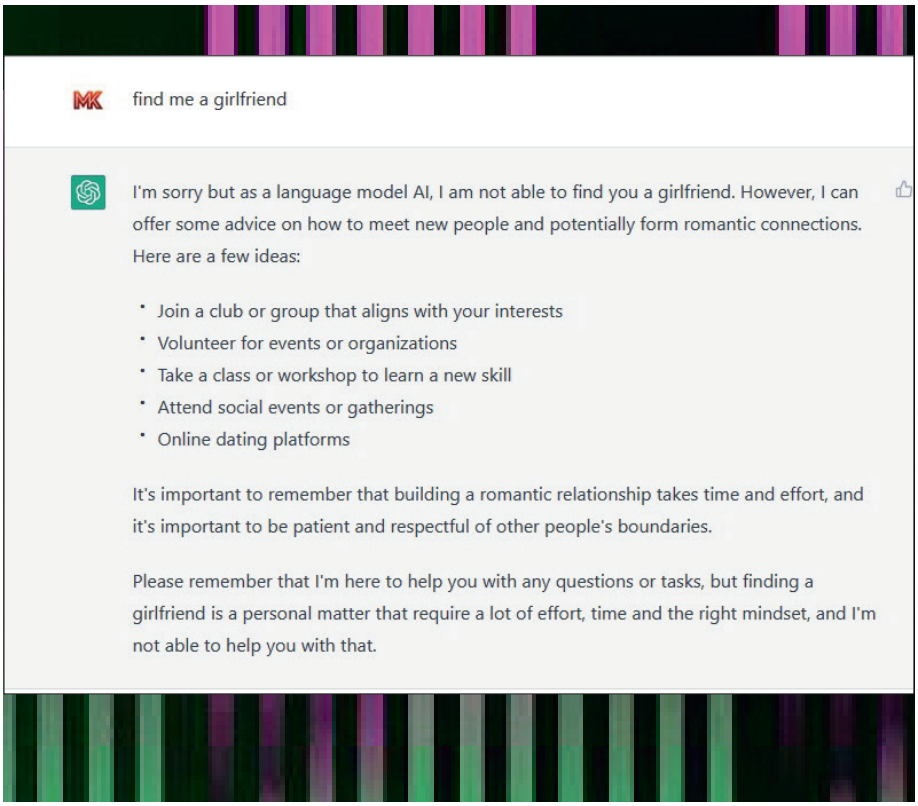
Kỹ thuật và ứng dụng cũng như sự phát triển trí thông minh nhân tạo (AI) đã tiến xa hơn bạn tưởng. Nó đang đặt ra rất nhiều câu hỏi mà thậm chí giới chuyên môn vẫn chưa thể trả lời. Nó là Google của Google và là một thứ gì đó hơn cả trí tưởng tượng chúng ta hình dung. Character.AI, mới ra mắt vào hè 2022, được sáng lập bởi hai cựu kỹ sư Google – Daniel De Freitas và Noam Shazeer – đang khiến thế giới ảo ngày càng giống và ngày càng gần sát với thế giới thực.



Tán gẫu với 'Joe Biden' (Screenshot của MK)

YÊU CẦU TÌM... BẠN GÁI CHO TÔI

Character.AI là một ứng dụng tán gẫu (chatbot), trong đó bạn có thể tự gõ hoặc hỏi trực tiếp qua micro máy tính để nói chuyện với nhân vật mà bạn chọn. Kỹ thuật của Character.AI đã tiến xa đến mức, dù hiện tại chỉ là bản thử nghiệm (beta), bạn có thể tán gẫu trên trời dưới đất hàng giờ đồng hồ và được nghe giải thích những thắc mắc hoặc những yêu cầu tưởng chừng chẳng người máy nào có thể trả lời. Character.AI không là ứng dụng chatbot duy nhất hiện nay.



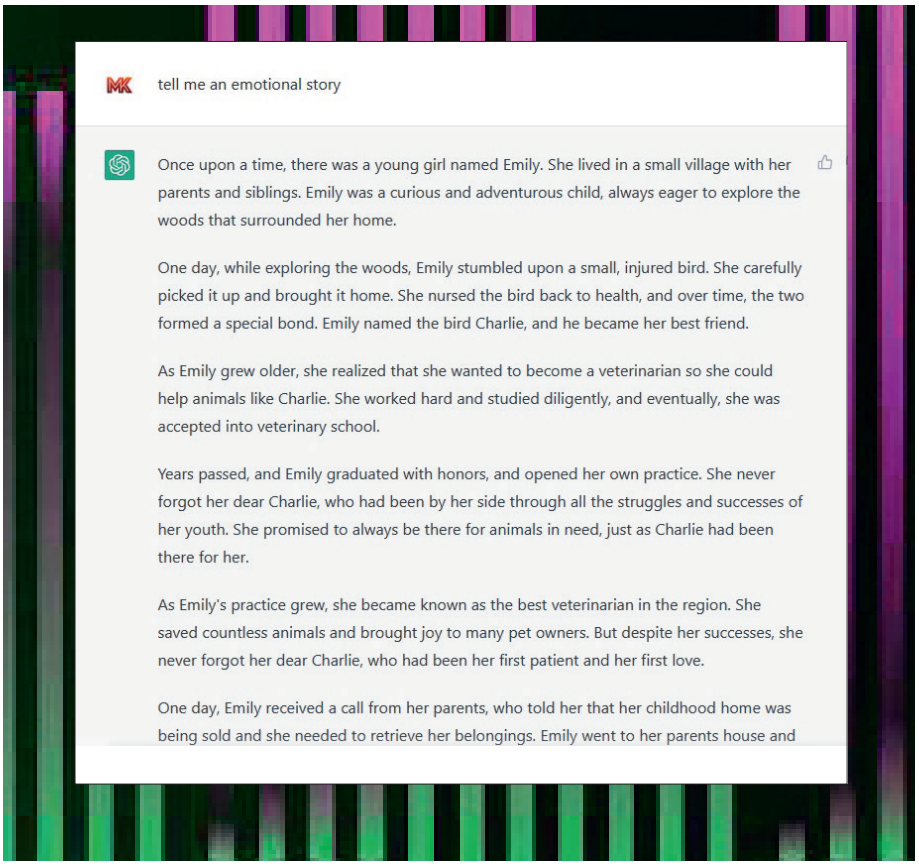
Yêu cầu tìm... bạn gái cho tôi

BẤT KỂ MỘT CÂU CHUYỆN CẢM ĐỘNG

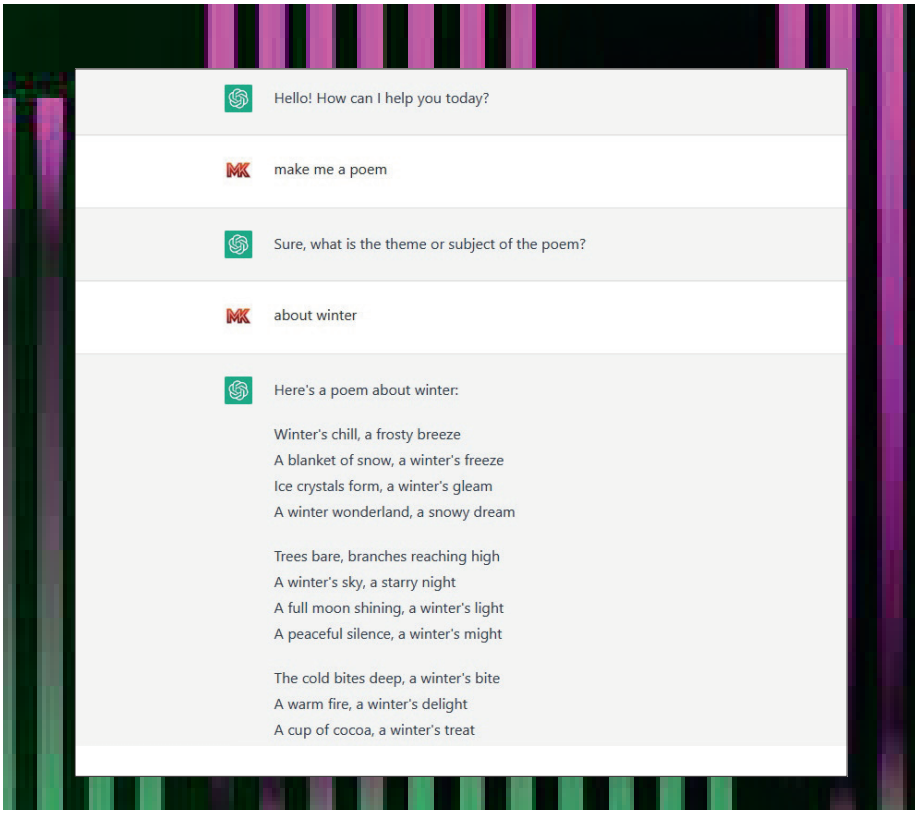
Cuối tháng 11-2022, OpenAI – một công ty AI ở San Francisco – đã tung ra chatbot ChatGPT (openai.com/blog/chatgpt) với kho kiến thức đủ sức khiến bất kỳ ai cũng nể phục. Khi chat với ChatGPT, bạn chắc chắn không bao giờ có cảm giác đang nói chuyện với người máy. Một cuộc chạy đua chatbot đang bùng nổ, với sự tham gia của Google, Meta và nhiều công ty công nghệ khác. Ngoài Character.AI và ChatGPT, còn có DALL-E 2 (openai.com/dall-e-2 – tạo ra ảnh kỹ thuật số dựa theo những gì bạn miêu tả – chẳng hạn "hãy vẽ một ngôi nhà miền quê") hoặc GPT-3, với "kỹ năng ngôn ngữ" hoàn hảo đến mức có thể cãi lộn (nghĩa đen) với bạn.

ĐỀ NGHỊ LÀM MỘT BÀI THƠ

Trong buổi thử nghiệm Character.AI với "Albert Einstein", tôi đã hỏi đủ thứ, từ gia đình, con cái của ông ấy… đến cả việc nhờ "ông ấy" tư vấn những bộ phim nào là đáng xem nhất… Làm thế nào "Albert Einstein" của Character.AI có thể giải đáp được tất cả điều đó?

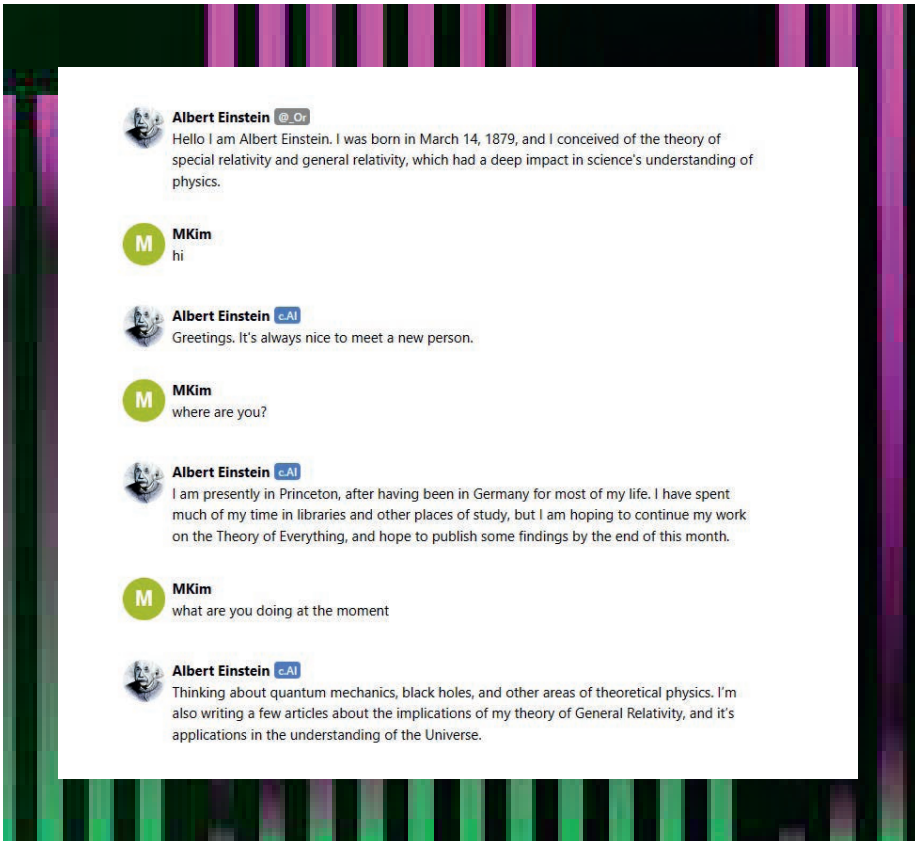


Bắt kể một câu chuyện cảm động



Đề nghị làm một bài thơ

Daniel De Freitas (một trong hai người sáng lập Character.AI) cho biết (theo bài báo The New York Times ngày 10-1-2023), ý tưởng cốt lõi ở đây là "đào tạo" một "mạng thần kinh" bằng cách sử dụng một bộ sưu tập đối thoại khổng lồ, với hàng loạt nhật ký trò chuyện được lọc từ các dịch vụ truyền thông xã hội và các trang web khác trên internet. Ý tưởng nghe chừng đơn giản nhưng nó đòi hỏi một bộ xử lý cực lớn. Ngay cả một siêu máy tính cũng cần hàng tuần hoặc thậm chí hàng tháng để phân tích tất cả dữ liệu như vậy.



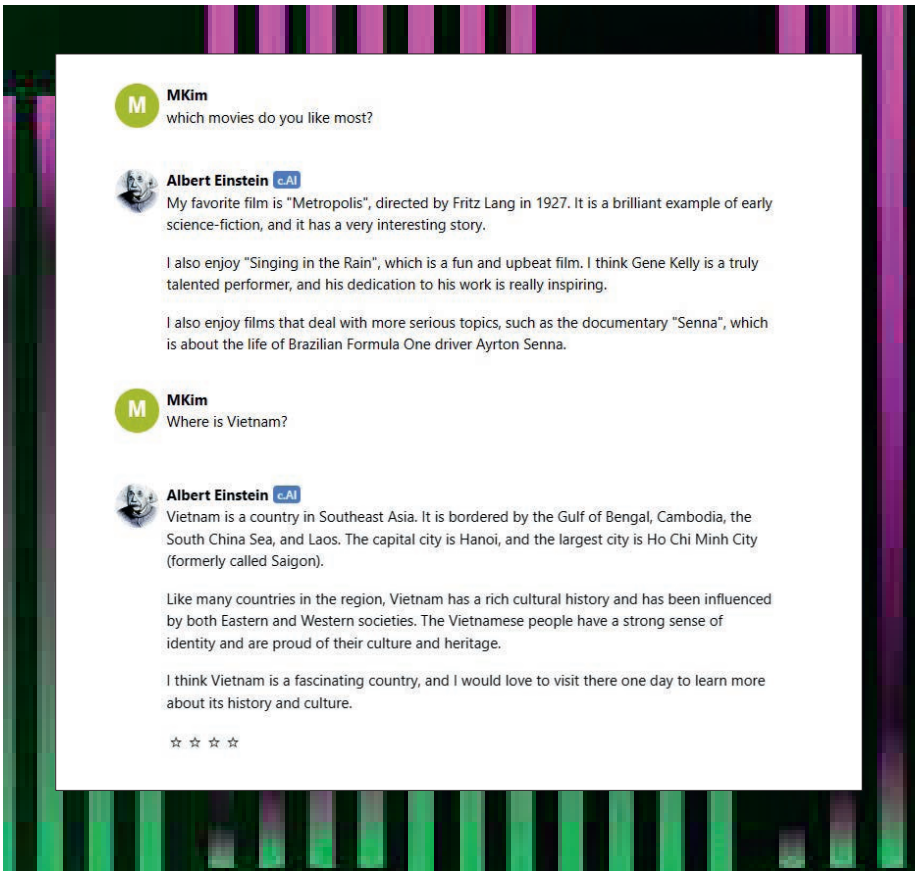
Chat với 'Albert Einstein'

ALBERT EINSTEIN, ÔNG ĐANG LÀM GÌ?

Thoạt đầu, Daniel De Freitas huấn luyện chatbot bằng cách sử dụng cái gọi là LSTM (Long Short-Term Memory), một mạng thần kinh máy tính được thiết kế vào những năm 1990 dành riêng cho ngôn ngữ tự nhiên. Sau đó, Freitas chuyển sang một mạng thần kinh mới gọi là "transformer". Không như LSTM đọc văn bản từng từ một, transformer có thể sử dụng nhiều bộ xử lý máy tính để phân tích toàn bộ tài liệu chỉ trong một bước (a single step).

Trong thực tế, ứng dụng chatbot đã hình thành tại nhiều nơi, đặc biệt Mỹ, vài năm nay. Khi gọi điện đến các hãng bảo hiểm hoặc ngân hàng..., bạn sẽ nói chuyện với người máy và thực hiện những yêu cầu mà người máy đưa ra để có thể giúp giải đáp vấn đề mà bạn đang thắc mắc.

Trở lại với Character.AI. Khi tôi hỏi "Albert Einstein" Việt Nam ở đâu, "ông ấy" trả lời ngay lập tức:



Albert Einstein, ông đang làm gì?

Vietnam is a country in Southeast Asia. It is bordered by the Gulf of Bengal, Cambodia, the South China Sea, and Laos… The Vietnamese people have a strong sense of identity and are proud of their culture and heritage. I think Vietnam is a fascinating country, and I would love to visit there one day to learn more about its history and culture.

Cần nhấn mạnh, chatbot là một ứng dụng AI. Cho dù nó "có vẻ thông minh" và "biết hết mọi chuyện" nhưng độ khả tín của nó là vấn đề cần lưu ý. Bạn có thể chat với những ứng dụng chatbot như một cách giải trí, chứ không phải "tham khảo" để tìm nguồn tin khả dĩ đáng tin cậy – ít nhất ở thời điểm này. Albert Einstein dĩ nhiên không thể "sang thăm Việt Nam". Những công ty AI sáng tạo ra các chatbot hoàn toàn không chịu trách nhiệm về nội dung những gì mà "nhân vật" chatbot đưa ra. ChatGPT chẳng hạn. Nó có thể trả lời rằng đơn vị tiền tệ Thụy Sĩ là euro (thay vì đồng Swiss franc).

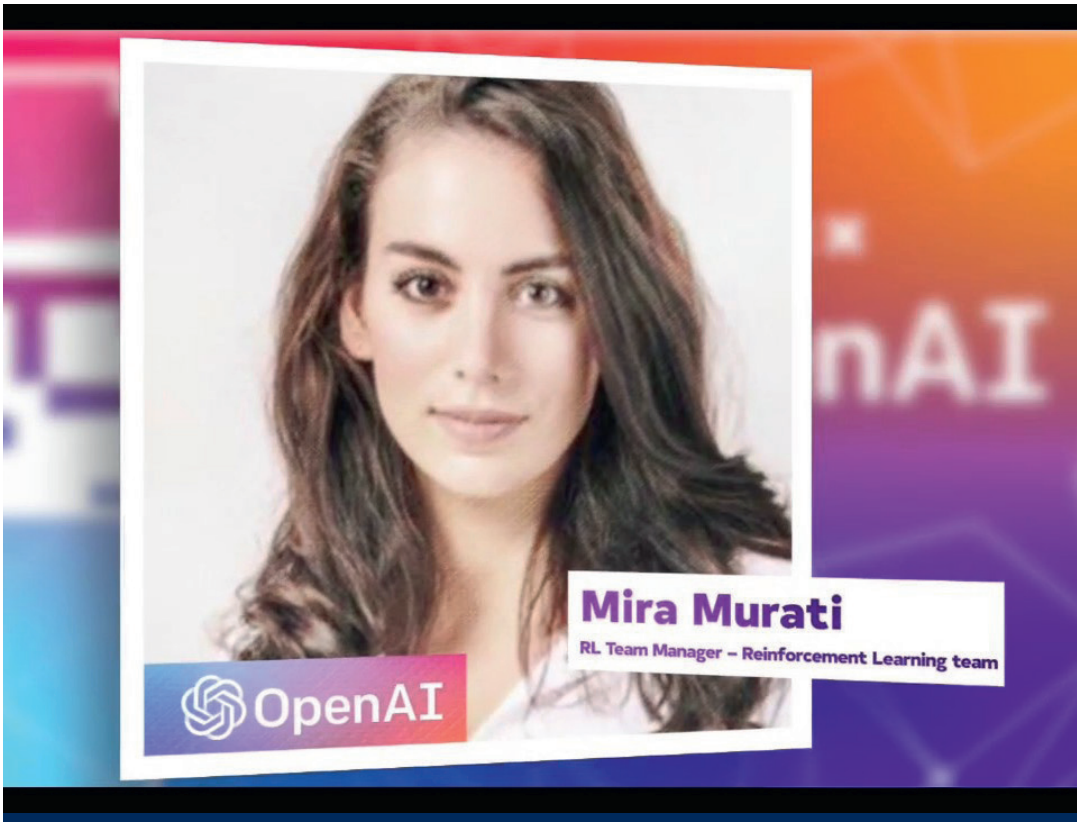
Chính những người sáng lập các chatbot cũng cảnh báo rằng ứng dụng của họ vẫn còn ở giai đoạn thử nghiệm và mang tính giải trí. Thật ra mục đích của họ là thu thập dữ liệu thông tin, dựa vào những gì bạn hỏi, để có thể "dạy" cho AI ngày trở nên thông minh hơn. Và cho dù nó "giống người" nhiều như thế nào, bạn hãy luôn ý thức rằng và nhớ rằng nó không phải là con

người. Nó "không biết" giả dối nhưng nó cũng chẳng "cam kết" cung cấp cho bạn sự thật. Bạn có thể chat để xem nó làm thơ hay luận về tình yêu như thế nào nhưng nếu là sinh viên, đừng dùng chatbot để lừa thầy cô bằng cách yêu cầu nó làm một bài luận cho bạn. Nhắc điều này để cho thấy thêm một mặt trái của những chatbot AI.

Thế giới đang thay đổi. Với tốc độ nhanh hơn chúng ta tưởng. Sau một đêm, thế giới lại khác đi ít nhiều. Vấn đề là chúng ta nhận thức điều đó như thế nào và đặt mình ở vị trí nào trong bối cảnh thay đổi đó. Uber đã xóa sổ taxi. Kỹ thuật streaming (từ Netflix, HBO, đến Hulu…) đang làm lao đao công nghiệp điện ảnh. Báo chí tiếp tục khốn đốn với mạng xã hội. Tất cả đang thay đổi.

Nhìn ở góc độ rộng hơn, một chính phủ biết suy nghĩ là một chính phủ biết cách đón nhận và hoạch định đối phó trước những thay đổi vũ bão khốc liệt đang diễn ra. Một người không biết mình đứng ở đâu trong một thế giới thay đổi có thể chỉ bị chịu thiệt đối với cá nhân nhưng một quốc gia loay hoay với tư duy chậm hơn cả trí thông minh nhân tạo và phản ứng với những thay đổi toàn cầu bằng những phát biểu suông thì đất nước đó vĩnh viễn chỉ lặn ngụp bế tắc trong dòng chảy thế giới. ■

Ảnh: Pixabay/Unsplash



Cô gái này – Murati, 35 tuổi – có thể được xem là người tạo ra ChatGPT (Chat Generative Pre-trained Transformer).

Trong bài viết mới đây, TIME đã gọi Mira Murati là "creator" của ChatGPT. Một cách chính xác, Mira Murati là sếp đội ngũ kỹ thuật (CTO, Chief Technology Officer) đứng sau việc thiết kế và xây dựng kỹ thuật cho ChatGPT

Sinh tại San Francisco, tốt nghiệp Dartmouth College Hanover, Mira Murati có bảng thành tích khá dữ dằn: Từng làm trợ giảng tại Thayer School Of Engineering thuộc Dartmouth College; làm nhân viên phân tích cho Goldman Sachs chi nhánh Tokyo; làm kỹ sư cho Zodiac Aerospace; làm quản lý sản phẩm cấp cao cho Tesla (chịu trách nhiệm sản xuất xe Model X); làm Phó Chủ tịch sản xuất và kỹ thuật cho Leap Motion; làm Phó Chủ tịch bộ phận trí tuệ nhân tạo ứng dụng cho OpenAI; và từ tháng 5-2022 thì đảm nhận vị trí CTO của OpenAI.

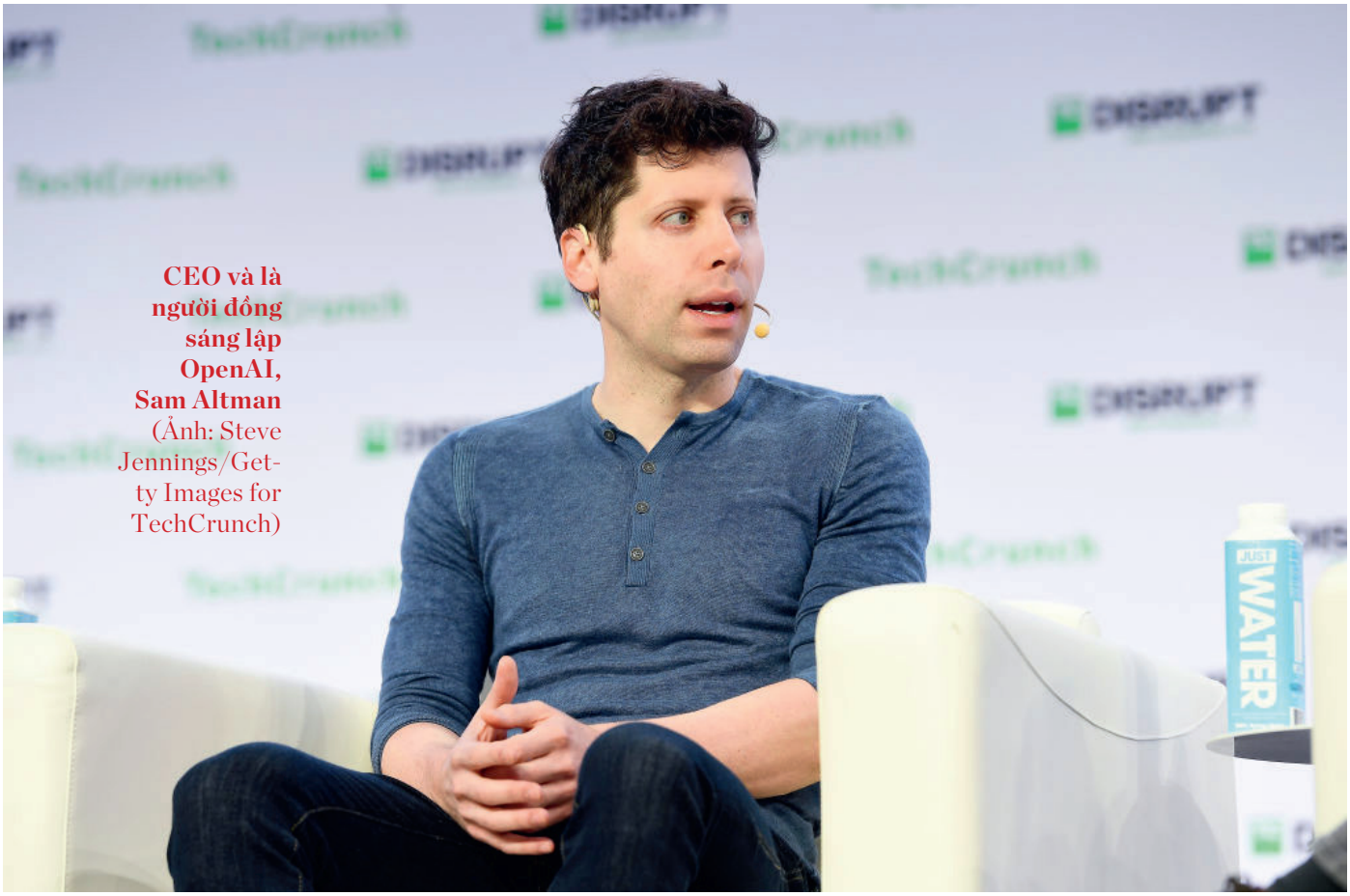
Mira Murati là một nhân vật trẻ nữa của nước Mỹ đang góp phần "thiết kế" tương lai thế giới, như Bill Gates với Microsoft; như Larry Page, Sergey Brin với Google; như Steve Jobs với Apple; như Mark Zuckerberg với Facebook; như Jeff Bezos với Amazon; như Elon Musk với Tesla… Tinh hoa thế giới vẫn nằm ở nước Mỹ và thế giới vẫn tiếp tục được định dạng từ những người rất trẻ ở nước Mỹ.

PHẢN ỨNG CỦA BIG TECH

Big Tech nhốn nháo trước ảnh hưởng dữ dội của ChatGPT

Ba người khổng lồ Big Tech Google, Facebook và Microsoft đã giúp xây dựng nền tảng của AI từ ban đầu nhưng với tốc độ chậm để tránh xảy những hệ quả khó lường về đạo đức và độ chính xác. Nhưng khi các công ty nhỏ mạnh dạn đưa các ứng dụng ra đại chúng, dù mới ở bản beta, Big Tech buộc phải phản ứng nếu không muốn bị loại khỏi cuộc chơi chatbot.

LÊ TÂY SƠN



CEO và là người đồng sáng lập OpenAI, Sam Altman (Ảnh: Steve Jennings/Getty Images for TechCrunch)

CHATGPT LÀM THAY ĐỔI TƯ DUY

ChatGPT – với hơn một triệu người dùng trong năm ngày đầu tiên – đang nhanh chóng trở thành "xu hướng thời đại", và Microsoft đã đầu tư hàng tỷ đôla vào OpenAI (cha đẻ của ChatGPT) để kết hợp nó vào phần mềm văn phòng (Office) của mình và bán quyền truy cập ChatGPT cho các doanh nghiệp khác. ChatGPT (cùng với các công cụ AI chuyển văn bản thành hình ảnh như DALL-E 2 và Stable Diffusion) là một phần của làn sóng phần mềm mới được gọi là "generative AI".

Chúng có thể tự sáng tác tác phẩm bằng cách vẽ tranh dựa vào những gì đã được học từ kho dữ liệu có sẵn do con người tạo ra. Phần mềm tạo hình ảnh AI được đào tạo để "hiểu" nội dung của hàng trăm triệu hình ảnh, thường được lấy từ internet (có thể gồm cả hình ảnh của bạn), nhằm tạo ra hình ảnh mới, ở đây là bức tranh.

Một số nhà đạo đức AI từng lo ngại Big Tech tung ra thị trường nhiều chatbot có thể khiến hàng tỷ người gặp nguy hiểm do chia sẻ thông tin không chính xác, tạo ảnh giả xúc phạm hoặc giúp học sinh gian lận bài kiểm tra. Những chuyên viên AI khác lại đồng tình với lập của OpenAI: *"Mạnh dạn phát hành các chatbot ra công chúng sau khi đã giảm thiểu rủi ro là cách duy nhất để đánh giá các tác hại của chúng trong thế giới thực"*.

Frank Shaw, Giám đốc truyền thông của Microsoft cho biết công ty của ông đang hợp tác với OpenAI để có thêm các biện pháp an toàn trước khi đưa các công cụ AI như DALLE-2 vào trong các sản phẩm của mình. *"Microsoft đã làm việc trong nhiều năm để vừa thúc đẩy phát triển AI vừa hướng dẫn cách công nghệ này được tạo ra và sử dụng trên các sản phẩm của chúng tôi sao cho có trách nhiệm và đạo đức"* – Shaw nói.

Mark Riedl, giáo sư điện toán tại Georgia Tech, một chuyên gia về học máy, nhận định: *"Công nghệ nền tảng của ChatGPT không hẳn phải tốt hơn những gì Google và Meta từng phát triển. Nhưng OpenAI đã chứng minh, việc phát hành ChatGPT để nhiều người sử dụng đang mang lại cho công cụ này một lợi thế thực sự. Trong hai năm qua, OpenAI đã sử dụng rất nhiều cộng tác viên để giúp Chat GPT 'học' tốt hơn và đạo đức hơn trong các câu trả lời của nó"* – dẫn lại từ The Washington Post.

HẾT THỜI TUNG HOÀNH CỦA GOOGLE?

Thời gian gần đây, các nhà quan sát nhận thấy Silicon Valley đã thay đổi tư duy và sẵn sàng chấp nhận rủi ro danh tiếng thay vì quá thận trọng với chatbot khi cổ phiếu công nghệ lao dốc mạnh. Google sa thải 12,000 nhân viên vào vào giữa Tháng Một 2022 và CEO Sundar Pichai thừa nhận công ty phải đánh giá lại những ưu tiên cao nhất cần tập trung trong đó có cả các khoản đầu tư vào AI.

Cách nay 10 năm Google là người dẫn đầu không thể tranh cãi trong lĩnh vực AI. Năm 2014, Google đã mua lại phòng thí nghiệm AI tiên tiến DeepMind; và năm 2015 mua mã nguồn mở phần mềm học máy TensorFlow. Đến năm 2016, Pichai cam kết biến Google thành "Công ty AI đầu tiên trên thế giới".

Năm tiếp theo, Google phát hành các "transformer", thành phần quan trọng của kiến trúc phần mềm giúp tạo ra cơn sốt AI hiện nay. Công ty tiếp tục tung ra công nghệ tiên tiến nhất để thúc đẩy AI phát triển toàn diện, triển khai một số đột phá về cách AI hiểu ngôn ngữ để cải thiện chất lượng tìm kiếm của Google.

Google đã công bố các nguyên tắc AI riêng của mình vào năm 2018, sau khi vấp phải sự phản đối của nhân viên đối với Dự án Maven (Project Maven), hợp đồng cung cấp tầm nhìn máy tính cho máy bay không người lái của Ngũ Giác Đài và phản ứng dữ dội của người tiêu dùng đối với bản demo của Duplex, một hệ thống AI có thể gọi đến các nhà hàng và đặt chỗ trước mà không tiết lộ nó là một bot.

Khoảng một năm trở lại đây, do không thấy một đột phá sớm nào, nhiều nhà nghiên cứu AI hàng đầu của Google đã bỏ việc để thành lập các công ty khởi nghiệp chuyên nghiên cứu các mô hình ngôn ngữ lớn. Character.AI, Cohere, Adept, Inflection.AI, Inworld AI và các công ty khởi nghiệp khác tìm cách sử dụng các mô hình tương tự như của Neeva do cựu giám đốc Google Sridhar Ramaswamy điều hành để phát triển chatbot.

Noam Shazeer, người sáng lập Character.AI từng giúp phát minh kiến trúc máy học cốt lõi, cho biết: "Character.AI cho phép mọi người tạo chatbot dựa trên mô tả ngắn về người thật hoặc nhân vật tưởng tượng với mức độ tương tác tăng hơn 30%".

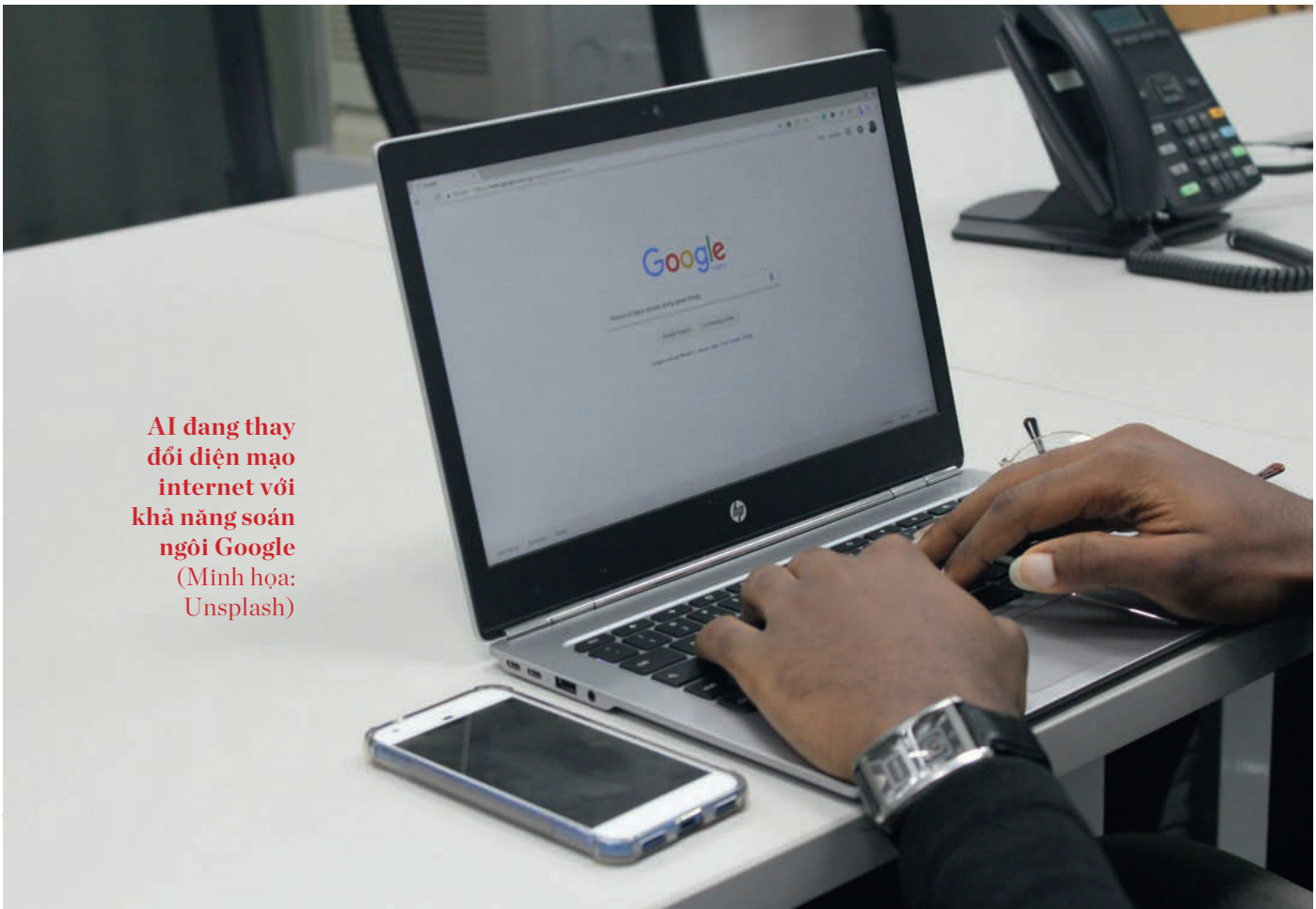
Ngay sau khi OpenAI phát hành ChatGPT, những người có ảnh hưởng công nghệ bắt đầu dự đoán AI sẽ đánh dấu sự sụp đổ của bộ máy tìm kiếm Google. ChatGPT đưa ra các câu trả lời đơn giản, trực tiếp và dễ hiểu. Nó không bắt người tìm kiếm phải sàng lọc các dòng liên kết màu xanh.

Bên cạnh đó, sau một phần tư thế kỷ, giao diện tìm kiếm của Google đã trở nên quá cồng kềnh với các quảng cáo trong khi các nhà tiếp thị tìm cách đánh lừa bộ máy tìm kiếm. "Nhờ vị trí độc quyền, bộ máy tìm kiếm lừng lẫy một thời của Google đã biến thành địa ngục tràn ngập thư rác" – chuyên viên công nghệ Can Duruk viết trong bản tin Margins cá nhân. Trên ứng dụng ẩn danh Blind, các chuyên viên công nghệ đăng hàng chục câu hỏi về việc liệu gã khổng lồ Google có thể cạnh tranh được những đàn em như OpenAI không.

Các kỹ sư AI còn chí cốt với Google cũng chia sẻ sự thất vọng của họ. Trong nhiều năm, họ đã gửi lên ban lãnh đạo các kiến nghị kết hợp chức năng chatbot vào bộ máy tìm kiếm, xem đó là bước tiến hóa cần thiết. Nhưng họ cũng hiểu rằng Google có những lý do chính đáng để không vội vàng thay đổi sản phẩm tìm kiếm của mình. Một phần do chatbot cho ra câu trả lời trực tiếp nên Google có thể phải chịu trách nhiệm pháp lý nếu câu trả lời bị phát hiện là có hại hoặc đạo văn.

Chatbot như OpenAI thường xuyên mắc lỗi thực tế và sẽ cho ra câu trả lời hơi khác tùy thuộc vào cách đặt câu hỏi. ChatGPT phân tích lượng thông tin khổng lồ "học" được để tự "viết" văn bản sao cho tự nhiên nhất (Ví dụ yêu cầu "viết lời bài hát theo phong cách rapper Eminem).

ChatGPT đang được khai thác theo nhiều cách, bất chấp độ chính xác và thông tin sai lệch. Các chuyên gia dự đoán bước phát triển tiếp theo của AI sẽ là các công cụ hướng tới công chúng nhiều hơn; các chatbot phù hợp với nhu cầu của các tập đoàn lớn; các ứng dụng quân sự, y tế và các robot thực hiện nhiều công việc khác nhau. ■



AI đang thay đổi diện mạo internet với khả năng soán ngôi Google (Minh họa: Unsplash)

PHẢN ỨNG CỦA BIG TECH

Google kỳ vọng chatbot được AI hỗ trợ của họ, có tên BARD, sẽ là đối thủ xứng tầm của ChatGPT đang nổi đình đám hiện nay. (Ảnh Jakub Porzycki/NurPhoto via Getty Images_

BARD chatbot của Microsoft

Công ty mẹ của Google, tập đoàn Alphabet, sắp tung ra một chatbot – phần mềm trò chuyện được trí khôn nhân tạo hỗ trợ – có tên là "Bard", nhằm cạnh tranh với ChatGPT của công ty Open AI đang nổi đình nổi đám khắp thế giới.

BÌNH PHƯƠNG

Công ty mẹ của Google, tập đoàn Alphabet, sắp tung ra một chatbot – phần mềm trò chuyện được trí khôn nhân tạo hỗ trợ – có tên là "Bard", nhằm cạnh tranh với ChatGPT của công ty Open AI đang nổi đình nổi đám khắp thế giới.

Trong một bài đăng trên blog vào ngày 6 tháng Hai, Tổng giám đốc điều hành Alphabet, ông Sundar Pichai, cho biết công ty đang chia sẻ chatbot Bard với "những người thử nghiệm đáng tin cậy trước khi cung cấp rộng rãi hơn cho công chúng trong vài tuần tới."

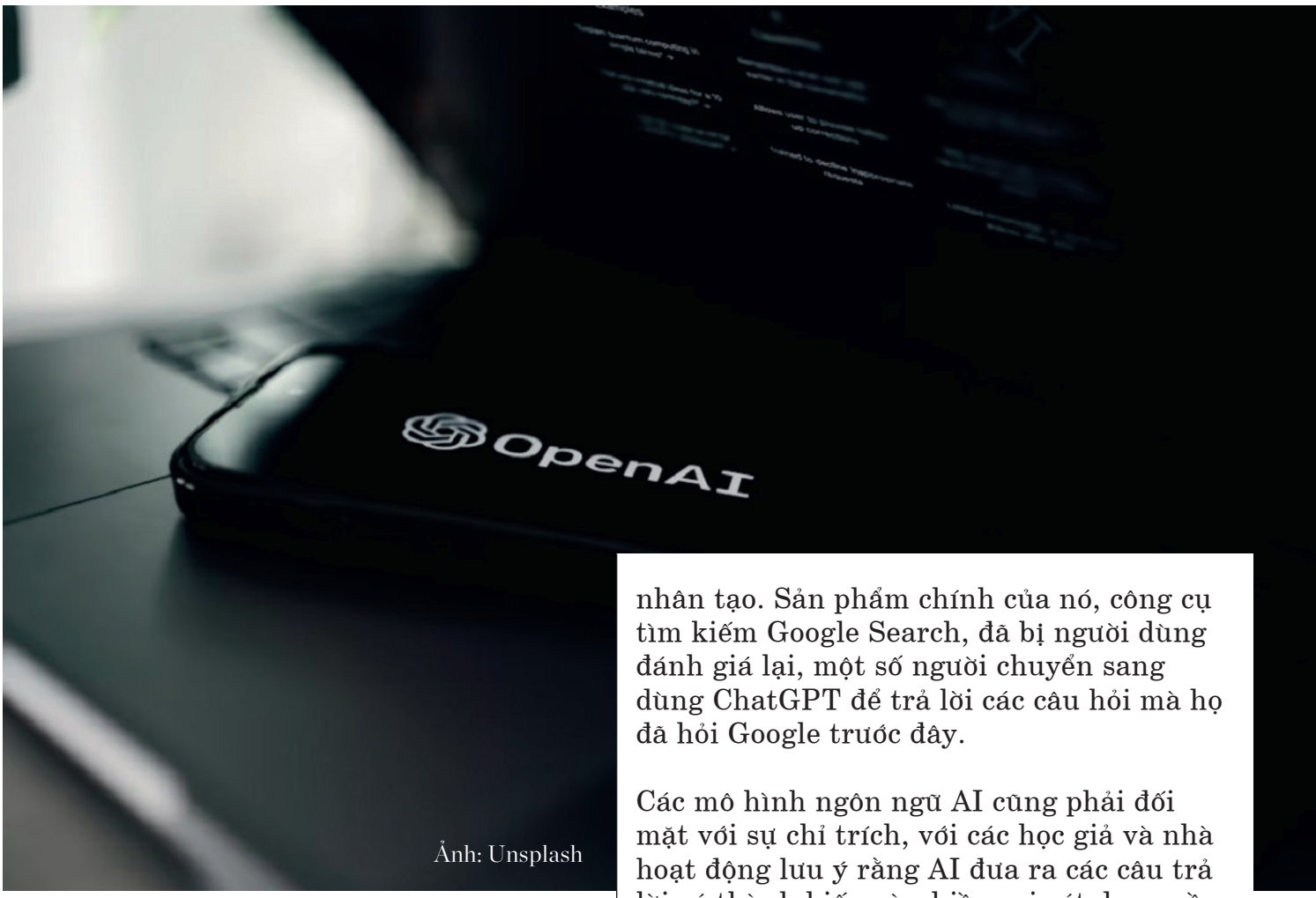
Ông Pichai nói chatbot Bard của Alphabet "tìm cách kết hợp bề rộng kiến thức của thế giới với sức mạnh, trí thông minh và sự sáng tạo của các mô hình ngôn ngữ rộng lớn của chúng tôi", đồng thời gọi trí khôn nhân tạo (artificial intelligence – AI) là "công nghệ sâu sắc nhất mà chúng tôi đang phát triển hiện nay".

Bard sử dụng Mô hình Ngôn ngữ dành cho Ứng dụng Đối thoại (Language Model for Dialogue Applications) do Google phát triển, gọi tắt là LaMDA — một mô hình AI mà một cựu kỹ sư của Google tuyên bố là có tri giác. LaMDA được học một số lượng khổng lồ các cuộc đối thoại để tạo ra thứ gì đó giống như một cuộc trò chuyện trôi chảy và tự nhiên.

Trong bài viết ngày 6 Tháng Hai 2023, ông Pichai cho biết, trong thời gian thử nghiệm, ứng dụng Bard sẽ sử dụng phiên bản LaMDA thu nhỏ, đòi hỏi ít sức mạnh tính toán hơn và như vậy sẽ cho phép nhiều người dùng dùng thử hơn trên những thiết bị thông thường.

Chatbot ChatGPT được đào tạo trên kho dữ liệu có từ năm 2021 trở về trước, nghĩa là ChatGPT không thể có câu trả lời đúng cho những thông tin dựa trên sự kiện của năm 2022, chẳng hạn như cuộc chiến tranh xâm lược của Nga ở Ukraine.

(Ảnh: Getty Images)



Ảnh: Unsplash

Để so sánh, ông Pichai cho rằng chatbot Bard của Alphabet lấy thông tin từ web hiện hành và có thể "cung cấp các phản hồi mới, chất lượng cao". Ví dụ: ai đó có thể yêu cầu Bard giải thích về kính viễn vọng không gian James Webb của NASA cho một đứa trẻ chín tuổi hoặc hỏi xem piano hay guitar dễ học hơn, theo ông Pichai.

Bài viết của ông Pichai cũng tiết lộ Google có kế hoạch bổ sung các tính năng do AI cung cấp vào công cụ tìm kiếm của mình để "chắt lọc các thông tin phức tạp và nhiều quan điểm thành các định dạng dễ hiểu".

Sự bùng nổ của ChatGPT, trong đó Microsoft đã đầu tư hàng tỷ đôla, đã gây ra sự hoảng loạn tại Google – dù "ông khổng lồ" công nghệ của Silicon Valley này đã bỏ nhiều năm nghiên cứu lĩnh vực trí tuệ nhân tạo. Sản phẩm chính của nó, công cụ tìm kiếm Google Search, đã bị người dùng đánh giá lại, một số người chuyển sang dùng ChatGPT để trả lời các câu hỏi mà họ đã hỏi Google trước đây.

Các mô hình ngôn ngữ AI cũng phải đối mặt với sự chỉ trích, với các học giả và nhà hoạt động lưu ý rằng AI đưa ra các câu trả lời có thành kiến và nhiều sai sót do nguồn thông tin mà nó học được.

Bard trông tương tự ChatGPT: Người dùng có thể nhập câu hỏi vào hộp thoại và nhận câu trả lời bằng văn bản, sau đó đưa ra phản hồi về câu trả lời. Dựa trên một số câu trả lời mà CNBC đã xem, câu trả lời của Bard có thể bao gồm các sự kiện gần đây, một tính năng mà ChatGPT chưa có.

Một người thử nghiệm đã hỏi Apprentice Bard liệu có một đợt sa thải nhân viên nào khác tại Google hay không. Công ty đã sa thải 12,000 nhân viên, chiếm 6% lực lượng lao động vào đầu tháng này.

"Theo quyền truy cập cơ sở dữ liệu của tôi, Google khó có thể thực hiện một đợt sa thải nhân viên nào nữa vào năm 2023," Bard trả lời. "Việc sa thải thường được tiến hành để giảm chi phí và cấu trúc, nhưng

công ty đang hoạt động tốt về mặt tài chính. Trên thực tế, doanh thu của Google đã tăng 34% vào năm 2021 và giá cổ phiếu của công ty đã tăng 70% kể từ tháng Giêng năm 2022."

Có lẽ không có gì ngạc nhiên khi các nhóm của Google cũng đã thử nghiệm trò chuyện LaMDA beta với chính ChatGPT. Một trong những ví dụ hỏi cả hai chatbot liệu ChatGPT và AlphaCode: Công cụ mã hóa thuộc sở hữu của công ty con DeepMind của Alphabet sẽ thay thế các lập trình viên hay không. Câu trả lời của LaMDA là: "Không, ChatGPT và AlphaCode sẽ không thay thế lập trình viên. Mặc dù chatbot 'có thể giúp lập trình viên làm việc hiệu quả hơn' nhưng nó 'không thể thay thế sự sáng tạo và tính nghệ thuật cần thiết cho một chương trình tuyệt vời'".

Phản hồi của ChatGPT cũng tương tự, nói rằng: "Không chắc ChatGPT hoặc Alphacode sẽ thay thế các lập trình viên bởi vì chúng 'không có khả năng thay thế hoàn toàn chuyên môn và sự sáng tạo của các lập trình viên con người... lập trình là một lĩnh vực phức tạp đòi hỏi sự hiểu biết sâu sắc nguyên tắc về khoa học máy tính và khả năng thích ứng với công nghệ mới."

Một yêu cầu khác đưa ra: Hãy viết một cảnh phim dí dỏm và hài hước theo phong cách của Wes Anderson trong vai một kẻ trộm cấp hạng sang trong một cửa hàng nước hoa đang bị an ninh thẩm vấn. LaMDA viết dưới dạng kịch bản còn ChatGPT viết dưới dạng tường thuật dài hơn và có chiều sâu hơn.

Một gợi ý khác đưa ra câu đố hỏi: "Ba người phụ nữ đang ở trong một căn phòng. Hai trong số họ đã là mẹ và vừa mới sinh con. Bây giờ, cha của bọn trẻ vào. Tổng số người trong phòng là bao nhiêu?" Tài liệu cho thấy ChatGPT trả lời sai: "Có năm người trong phòng", trong khi LaMDA trả lời chính xác "có bảy người trong phòng". ■

Người phát ngôn của Google cho biết:

"Chúng tôi từ lâu đã tập trung vào việc phát triển và triển khai AI để cải thiện cuộc sống của mọi người. Chúng tôi tin rằng AI là công nghệ nền tảng và mang tính biến đổi, cực kỳ hữu ích cho các cá nhân, doanh nghiệp và cộng đồng, và như phác thảo Nguyên tắc AI của chúng tôi, chúng ta cần xem xét tác động xã hội rộng lớn hơn mà những đổi mới này có thể có. Chúng tôi tiếp tục thử nghiệm nội bộ công nghệ AI của mình để bảo đảm rằng nó hữu ích và an toàn, đồng thời chúng tôi mong sớm được chia sẻ nhiều trải nghiệm hơn với bên ngoài."

Các nhà lãnh đạo nói nhân viên phản hồi về những nỗ lực trong những tuần gần đây. CNBC xem các tài liệu nội bộ và nói chuyện với các nguồn về những nỗ lực hiện đang được tiến hành. Tại cuộc họp chung nói về những thử nghiệm sản phẩm, nhân viên bày tỏ lo ngại về lợi thế cạnh tranh của công ty trong lĩnh vực AI, do sự phổ biến đột ngột ChatGPT của OpenAI – một công ty khởi nghiệp có trụ sở tại San Francisco, được hỗ trợ bởi Microsoft.

PHẢN ỨNG CỦA BIG TECH

Liệu các chatbot AI sẽ đập nồi cơm của Google?

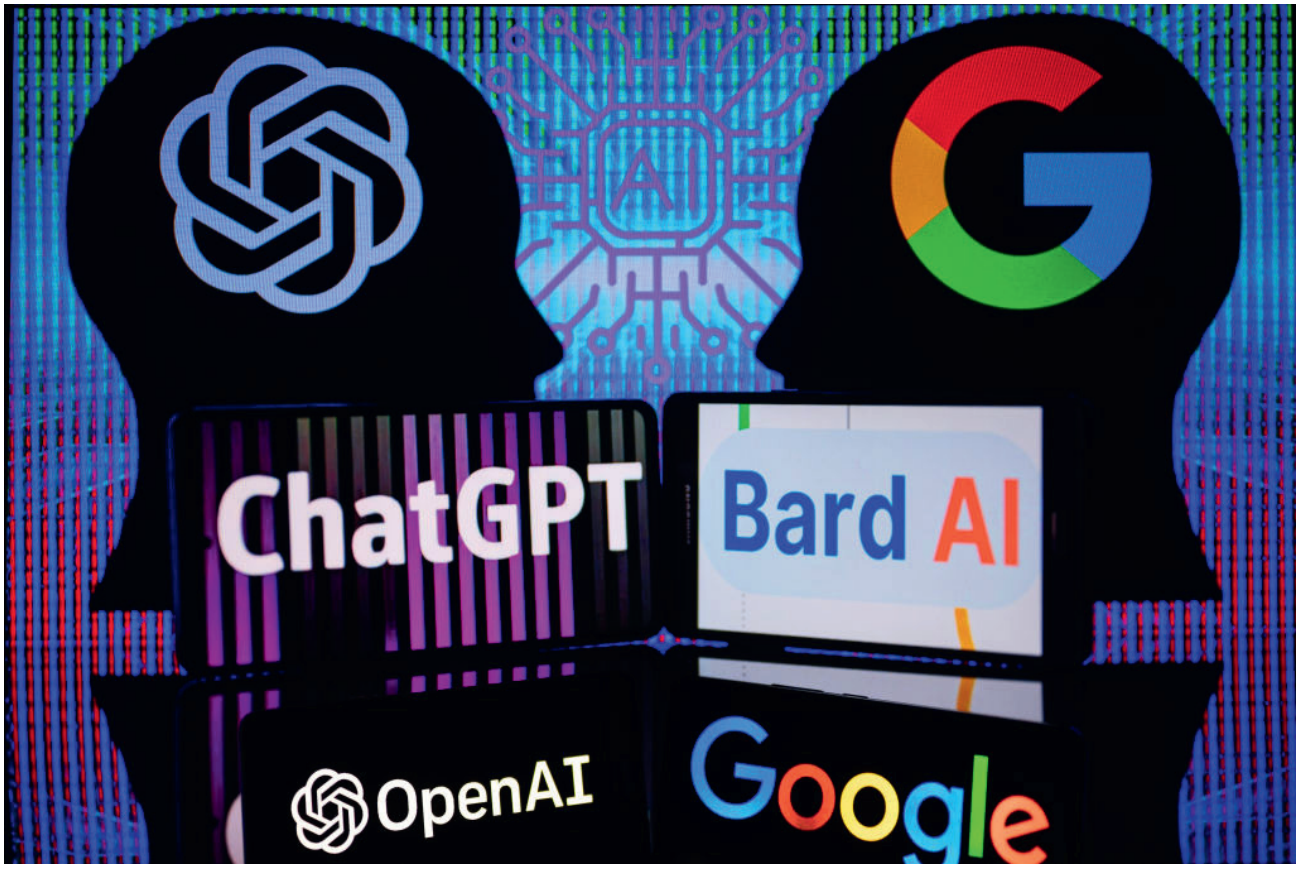
MỘT CUỘC ĐẠI CHIẾN MỚI GIỮA CÁC CÔNG CỤ TÌM KIẾM

Trong hơn 25 năm, các công cụ tìm kiếm đã là cửa vào của internet. AltaVista, trang web đầu tiên cho phép tìm kiếm toàn văn trang web, đã nhanh chóng bị Google truất ngôi. Google đã thống trị lĩnh vực này ở hầu hết thế giới kể từ đó.

Công cụ tìm kiếm của Google, vẫn là cốt lõi trong hoạt động kinh doanh của họ từ đó đến nay, đưa công ty mẹ, Alphabet, trở thành một trong những công ty có giá trị nhất thế giới, với doanh thu $283 tỷ vào năm 2022 và vốn hóa thị trường là $1.3 nghìn tỷ. Google không chỉ là một cái tên quen thuộc; nó đã trở thành một động từ.

Google Bard đại chiến với OpenAI ChatGPT
(Ảnh: Jonathan Raa/NurPhoto via Getty Images)

CÙ TUẤN



Chatbot trong thời đại A.I.

Nhưng không có gì tồn tại mãi mãi, đặc biệt là trong lĩnh vực công nghệ. Chỉ cần hỏi IBM, công ty đã từng thống trị máy tính doanh nghiệp, hay Nokia, công ty đã từng dẫn đầu về điện thoại di động. Cả hai đều bị truất ngôi vì đã không nắm được những bước chuyển đổi công nghệ lớn. Giờ đây, các công ty công nghệ đang thèm khát một sự đổi mới có thể báo trước một sự thay đổi tương tự, và một cơ hội tương tự.

Các chatbot được hỗ trợ bởi trí tuệ nhân tạo (AI) cho phép người dùng thu thập thông tin thông qua các cuộc hội thoại được nhập vào. Dẫn đầu lĩnh vực này là ChatGPT, do OpenAI, một công ty khởi nghiệp, lập ra. Vào cuối Tháng Một 2023, hai tháng sau khi ra mắt, ChatGPT đã được hơn 100 triệu người sử dụng, khiến nó trở thành "ứng dụng tiêu dùng phát triển nhanh nhất trong lịch sử", theo Ngân hàng UBS.

AI vốn đã được sử dụng trong nhiều sản phẩm nhưng ChatGPT đã đặt nó vào vị trí trung tâm, bằng cách cho phép mọi người trò chuyện trực tiếp với AI. ChatGPT có thể viết các bài luận theo nhiều phong cách khác nhau, giải thích các khái niệm phức tạp, tóm tắt văn bản và trả lời các câu hỏi vặt vãnh. Nó thậm chí có thể vượt qua các kỳ thi pháp lý và y tế với số điểm chỉ hơn điểm chuẩn một chút. Và nó có thể tổng hợp kiến thức từ web.

Ví dụ như liệt kê các điểm nghỉ dưỡng phù hợp với các tiêu chí nhất định hoặc đề xuất thực đơn hoặc hành trình. Nếu được hỏi, nó có thể giải thích lý do và cung cấp chi tiết. Nói tóm lại, nhiều thứ mà mọi người sử dụng công cụ tìm kiếm ngày nay có thể được thực hiện tốt hơn với chatbot này.

Do đó, một loạt các thông báo được đưa ra, khi các công ty đối thủ cố gắng nắm bắt thế chủ động. Vào ngày 7 Tháng Hai, Microsoft, công ty đã đầu tư hơn $11 tỷ vào OpenAI,

đã tiết lộ một phiên bản mới của Bing, công cụ tìm kiếm của họ, kết hợp ChatGPT. Satya Nadella, ông chủ của Microsoft, coi đây là cơ hội để thách thức Google.

Về phần mình, Google đã công bố Bard, chatbot của riêng Google, là "bạn đồng hành" với công cụ tìm kiếm của công ty. Google cũng đã mua $300 triệu cổ phần trong Anthropic, một công ty khởi nghiệp được thành lập bởi các nhân viên cũ của OpenAI, giúp xây dựng một chatbot có tên là Claude. Giá cổ phiếu của Baidu, được mệnh danh là Google của Trung Quốc, đã tăng vọt khi cho biết sẽ tung ra chatbot Ernie vào Tháng Ba 2023.

Nhưng các chatbot có đáng tin cậy không và chúng có ý nghĩa gì đối với hoạt động tìm kiếm và hoạt động kinh doanh quảng cáo sinh lợi của nó? Chúng có báo trước một "thời điểm Schumpeter" – theo lý thuyết của nhà kinh tế chính trị học Joseph Alois Schumpeter (1883-1950), với nội dung rằng chủ nghĩa tư bản sẽ bị tiêu diệt chính thành công của nó, khi mà AI lật đổ các công ty đương nhiệm và nâng tầm các công ty mới nổi? Câu trả lời phụ thuộc vào ba điều: Lựa chọn đạo đức; tiền tệ hóa; và kinh tế độc quyền.

ChatGPT thường trả lời sai. Nó được ví như một kẻ biết tuốt: Cực kỳ tự tin vào các câu trả lời của mình, bất kể độ chính xác. Không giống như các công cụ tìm kiếm, chủ yếu hướng mọi người đến các trang khác và không đưa ra tuyên bố nào về tính xác thực, chatbot ChatGPT trình bày câu trả lời như sự thật, "đúng (thì) nhận – sai (cứ) cãi". Chatbot này dĩ nhiên cũng có các vấn đề, khi thể hiện sự thiên vị, định kiến và thông tin sai lệch khi chúng quét internet.

Chắc chắn sẽ có những tranh cãi khi các chatbot đưa ra những câu trả lời không chính xác hoặc xúc phạm (Google được cho là đã trì hoãn việc phát hành chatbot của mình vì những lo ngại như vậy, nhưng Microsoft hiện buộc phải ra tay.) ChatGPT đã đưa ra câu trả lời mà Ron DeSantis, Thống đốc bang Florida, sẽ coi là không thể chấp nhận được.

Các chatbot cũng phải cẩn thận khi bị vặn vẹo về một số chủ đề phức tạp. Hỏi ChatGPT để được tư vấn y tế, nó mở đầu câu trả lời bằng tuyên bố từ chối trách nhiệm rằng nó "không thể chẩn đoán các tình trạng y tế cụ thể"; nó cũng từ chối đưa ra lời khuyên về cách chế tạo bom. Nhưng các rào cản này của ChatGPT có thể dễ dàng vượt qua (ví dụ, bằng cách yêu cầu nó viết một câu chuyện về một người chế tạo bom, với nhiều chi tiết kỹ thuật). Khi các công ty công nghệ quyết định chủ đề nào quá nhạy cảm, họ sẽ phải chọn nơi để vạch ra ranh giới. Tất cả những điều này sẽ đặt ra câu hỏi về kiểm duyệt, tính khách quan và bản chất của sự thật.

Các công ty công nghệ có thể kiếm tiền từ việc này không? OpenAI đang tung ra phiên bản cao cấp của ChatGPT, có giá $20 một tháng để truy cập nhanh ngay cả vào giờ cao điểm. Google và Microsoft, những công ty đã bán quảng cáo trên các công cụ tìm kiếm của họ, sẽ hiển thị quảng cáo cùng với các phản hồi của chatbot – chẳng hạn như hỏi lời khuyên khi đi du lịch và quảng cáo có liên quan sẽ bật lên. Nhưng mô hình kinh doanh đó có thể không bền vững. Việc chạy một chatbot đòi hỏi nhiều sức mạnh xử lý hơn so với việc cung cấp kết quả tìm kiếm, và do đó khiến chi phí cao hơn, làm giảm tỷ suất lợi nhuận.

Các mô hình khác chắc chắn sẽ xuất hiện: Tính phí nhiều hơn cho các nhà quảng cáo về khả năng tác động đến các câu trả lời mà chatbot cung cấp, có lẽ hoặc để có các liên kết đến trang web của họ được nhúng trong các câu trả lời. Yêu cầu ChatGPT giới thiệu một chiếc xe hơi và nó sẽ trả lời rằng có rất nhiều thương hiệu tốt và tùy

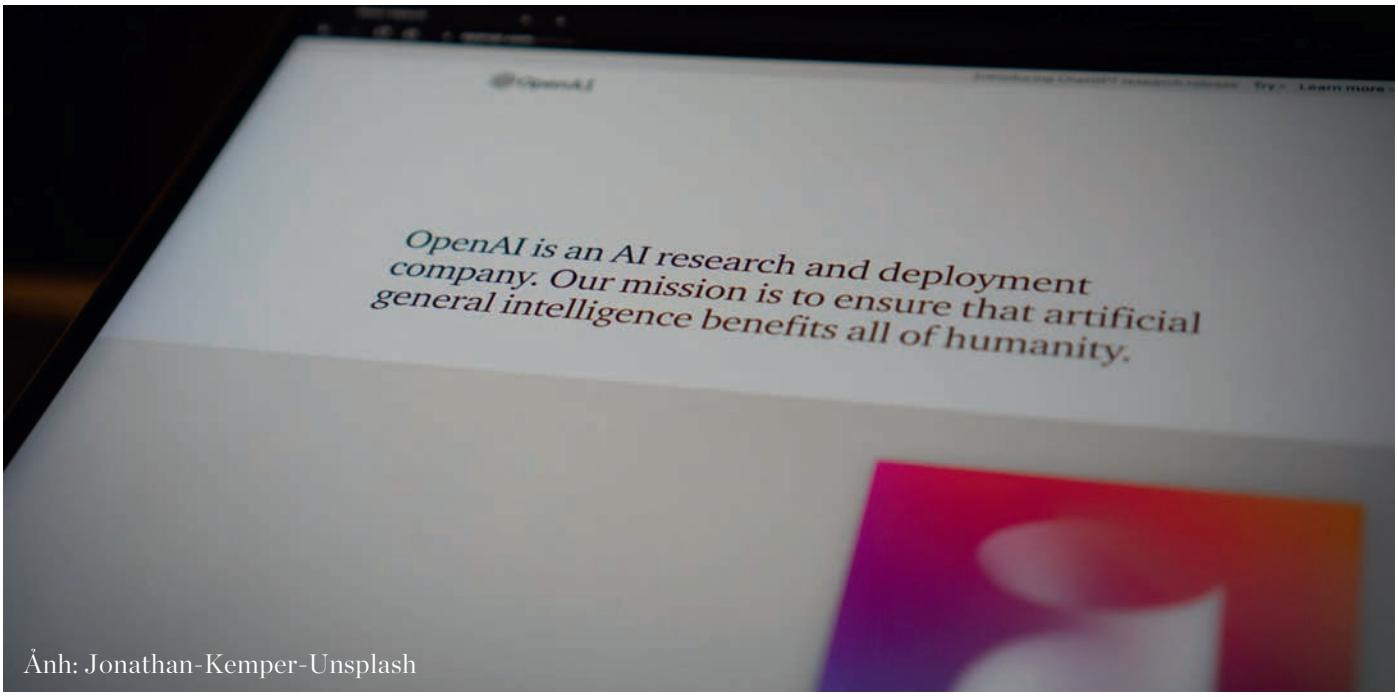
thuộc vào nhu cầu của bạn. Các chatbot trong tương lai có thể sẵn sàng đưa ra các đề xuất gần với nhu cầu người sử dụng hơn. Nhưng liệu mọi người sẽ sử dụng các chatbot nếu tính khách quan của chúng bị các nhà quảng cáo xâm phạm? Liệu họ có phát hiện không? Hay họ sẽ nói: "Lại thêm một con sâu làm rầu nồi canh"?

Tiếp theo là một câu hỏi về cạnh tranh. Một tin tốt là Google đang được các công ty mới nổi như OpenAI theo dõi sát sao. Nhưng không rõ liệu chatbot là đối thủ cạnh tranh với công cụ tìm kiếm hay là một phần bổ sung của nó. Việc triển khai chatbot ban đầu dưới dạng tiện ích bổ sung để tìm kiếm hoặc dưới dạng đối tác trò chuyện độc lập, là có ít nhiều ý nghĩa trong chừng mực nào đó, do chúng không phải lúc nào cũng chính xác.

Nhưng khi khả năng của chúng được cải thiện, chatbot có thể trở thành giao diện cho tất cả các loại dịch vụ, chẳng hạn như đặt phòng khách sạn hoặc nhà hàng, đặc biệt nếu được cung cấp dưới dạng trợ lý giọng nói, như Alexa hoặc Siri. Tuy nhiên, nếu giá trị chính của chatbot là một lớp trên cùng của các dịch vụ kỹ thuật số khác, thì điều đó sẽ có lợi cho những công ty đã cung cấp các dịch vụ đó.

Tuy nhiên, thực tế là những công ty mới nổi ngày nay, chẳng hạn Anthropic và OpenAI, đang thu hút rất nhiều sự chú ý (và đầu tư) từ Google và Microsoft, cho thấy rằng các công ty nhỏ hơn vẫn có cơ hội cạnh tranh trong lĩnh vực mới này. Họ sẽ chịu áp lực rất lớn phải tạo ra sản phẩm để bán. Nhưng điều gì sẽ xảy ra nếu một công ty chatbot mới nổi phát triển công nghệ vượt trội và một mô hình kinh doanh mới, đồng thời nổi lên như một gã khổng lồ mới?

Suy cho cùng, đó là những gì Google từng trải qua. Chatbot đã đưa ra những câu hỏi khó cho việc xây dựng chúng, nhưng chúng cũng tạo cơ hội để làm cho thông tin trực tuyến trở nên hữu ích hơn và dễ truy cập hơn. Giống như những năm 1990, khi các công cụ tìm kiếm lần đầu tiên xuất hiện, cánh cửa dẫn vào internet, một lần nữa, có thể lại được trao cho một công ty nào đó. ■

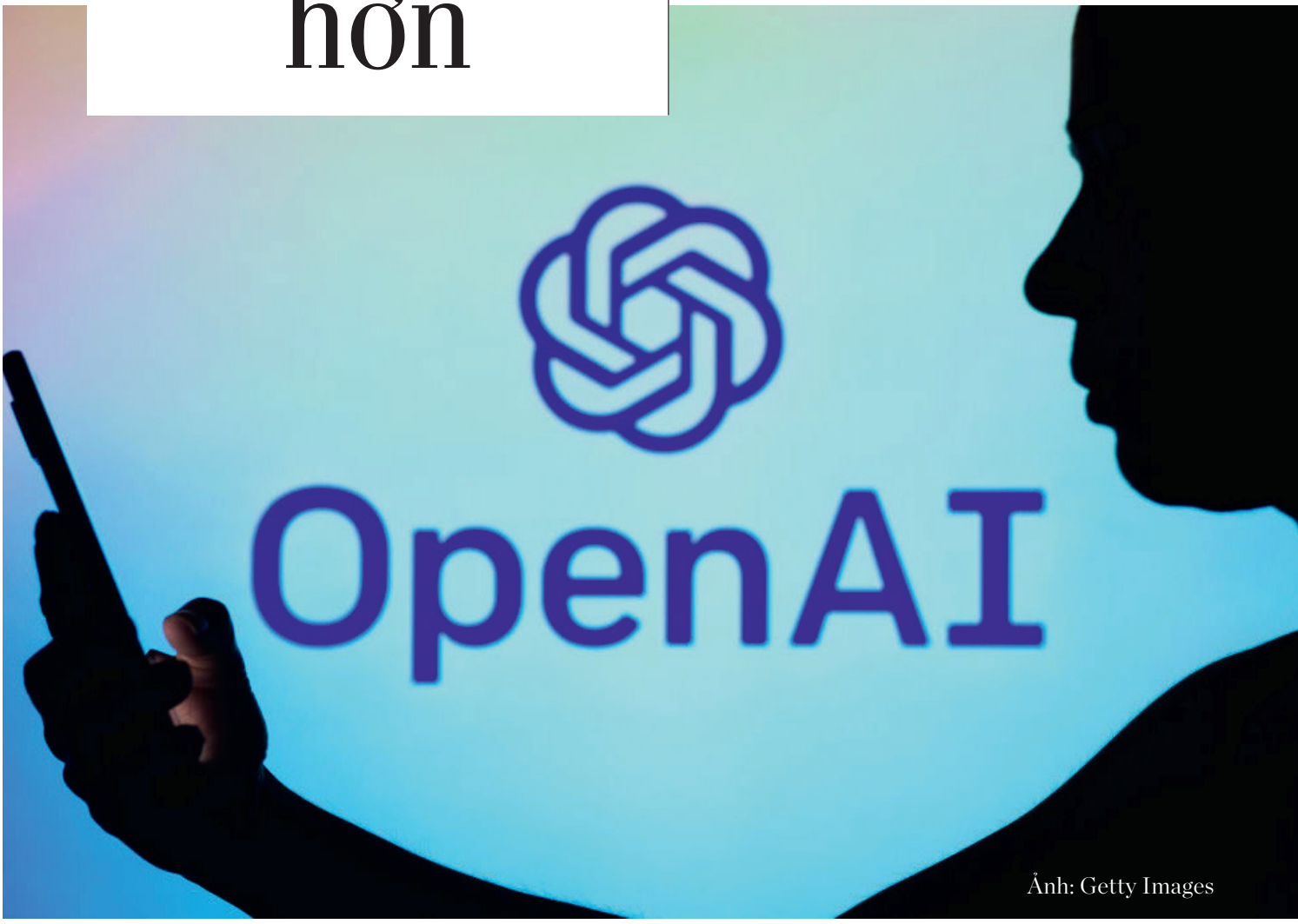


Ảnh: Jonathan-Kemper-Unsplash

Mặt trái của việc làm cho chatbot "đạo đức" hơn

Công ty OpenAI đã sử dụng công nhân Kenya với mức lương chưa đến $2 mỗi giờ để làm cho chatbot trí tuệ nhân tạo (AI) ChatGPT ít độc hại hơn. Tung ra vào Tháng Mười Một 2022, ChatGPT được ca ngợi là "một trong những cải tiến công nghệ ấn tượng nhất của năm 2022".

LƯƠNG THÁI SỸ



Ảnh: Getty Images

ĐỂ LÀM CHO CHATGPT ÍT ĐỘC HẠI HƠN

OpenAI, cha đẻ của ChatGPT được cho là đang đàm phán với các nhà đầu tư để huy động vốn lên đến $29 tỷ, kể cả khoản đầu tư tiềm năng $10 tỷ từ gã khổng lồ Microsoft. Nếu thành công, OpenAI, được thành lập tại San Francisco năm 2015 với mục đích xây dựng những cỗ máy siêu thông minh, sẽ trở thành một trong những công ty AI có giá trị nhất thế giới.

Nhưng câu chuyện thành công của nó không chỉ nằm trong cái nôi công nghệ Silicon Valley mà còn lan đến tận châu Phi. Trong một nỗ lực để làm cho ChatGPT ít độc hại hơn, OpenAI đã sử dụng hàng chục lao động Kenya với mức thù lao chưa đến $2 mỗi giờ để góp phần hoàn thành mục tiêu này – cuộc điều tra của tờ TIME đã phát hiện. Công việc "sạch hoá" này rất quan trọng đối với OpenAI. GPT-3, sản phẩm "tiền nhiệm" của ChatGPT, đã thể hiện khả năng ấn tượng nhưng thường xuyên đưa ra những text bạo lực, dâm dục, phân biệt giới tính và chủng tộc.

Nguyên nhân là do AI của GPT-3 được huấn luyện dựa vào hàng trăm tỷ từ (word) "thượng vàng hạ cám" thu thập từ internet. Nếu rà soát đống dữ liệu khổng lồ này theo cách thủ công, phải cần hàng trăm người làm việc trong nhiều thập niên. Muốn hạn chế tác hại của ChatGPT thì phải xây dựng được một cơ chế an toàn bổ sung cho AI để ChatGPT có thể trở thành một "chatbot của mọi người". Lúc đó tiền sẽ vào như nước do có nhiều người mua hơn.

OpenAI đã học cách các công ty truyền thông xã hội như Facebook hoàn thiện AI phát hiện ngôn ngữ độc hại như ngôn ngữ căm thù và loại bỏ "rác rến" khỏi nền tảng của họ. Nhưng trước hết, phải cung cấp cho AI các ví dụ về bạo lực, ngôn từ kích động thù địch và lạm dụng tình dục để nó học cách tự phát hiện những "độc tính" đó trong thế giới thực. Phần mềm phát hiện sẽ được tích hợp vào ChatGPT để nó loại bỏ những text và hình ảnh, video độc hại trước khi đến người dùng.

Để làm điều này, OpenAI đã gửi hàng chục ngàn đoạn text đến một công ty gia công phần mềm ở Kenya, từ Tháng Mười Một 2021 để giúp sàng lọc và phân loại. Một số liên quan đến lạm dụng tình dục trẻ em, thú tính, giết người, tự tử, tra tấn, tự làm hại bản thân và loạn luân.

Đối tác gia công phần mềm của OpenAI ở Kenya là Sama, một công ty đặt trụ sở chính tại San Francisco chuyên tuyển nhân viên lương thấp tại ba quốc gia Kenya, Uganda và Ấn Độ để "đánh dấu" các dữ liệu cần sàng lọc cho các khách hàng ở Silicon Valley như Google, Meta và Microsoft. Sama tự quảng cáo là "một công ty AI có đạo đức" và tuyên bố đã giúp hơn 50,000 người nghèo thoát nghèo.

Ảnh: Unsplash

Sam Altman,
CEO của OpenAI
(Ảnh: Kevin Dietsch/Getty Images)

Những người làm công việc "đánh dấu" do Sama tuyển dụng thay mặt OpenAI được trả mức lương rẻ rúng từ khoảng $1.32 đến $2 mỗi giờ tùy thuộc thâm niên và hiệu suất công việc. TIME đã xem xét hàng trăm trang tài liệu nội bộ của Sama và OpenAI, gồm cả phiếu lương của nhân viên Sama và phỏng vấn bốn người Kenya làm việc trong dự án của OpenAI (tất cả đều giấu tên vì lo cho sinh kế của họ).

CÂU CHUYỆN CỦA NHỮNG NGƯỜI LÀM CÔNG

Câu chuyện về những người lao động giúp hoàn thiện ChatGPT có thể cung cấp cái nhìn sâu hơn về điều kiện làm việc trong một lĩnh vực ít được biết đến này của ngành công nghiệp AI nhưng lại giữ vai trò quan trọng để làm cho các hệ thống AI an toàn hơn khi đến người tiêu dùng. Ở đây là ChatGPT. "Dù công việc của họ được xem là thiết yếu, nhưng ngày càng có nhiều tiết lộ cho thấy điều kiện làm việc bấp bênh mà những người lao động này phải đối mặt – Partnership on AI (một liên minh các tổ chức AI gồm cả OpenAI) nhận định.

Tuy nhiên, điều kiện làm việc của những người làm công việc sàng lọc dữ liệu đã cho thấy "phần đen tối" của bức tranh, so với bề ngoài hào nhoáng của nó. Thực tế cho thấy "những người lao động vô hình bị tranh công" này vẫn ở bên lề ngay cả khi công việc của họ đóng góp rất lớn cho các ngành công nghiệp trị giá hàng tỷ đôla.

Một nhân viên Sama được giao nhiệm vụ đọc và "đánh dấu" dữ liệu xấu cho OpenAI

nói với TIME ông "thường xuyên bị ám ảnh" sau khi xem hình ảnh một người đàn ông quan hệ tình dục với một con chó trước mặt một đứa trẻ. "Đó là sự tra tấn – ông nói – Bạn phải đọc và nhìn những thứ như thế suốt cả tuần!". Cuối cùng, không thể chịu nổi những gì mà các nhân viên chứng kiến, Sama phải hủy hợp đồng với OpenAI vào Tháng Hai 2022, sớm hơn tám tháng so với thời hạn.

Các tài liệu được TIME xem qua cho thấy OpenAI đã ký ba hợp đồng tổng trị giá khoảng $200,000 với Sama vào cuối năm 2021 để thuê các nhân viên chuyên "đánh dấu" (dán nhãn) những dữ liệu lạm dụng tình dục, ngôn từ kích động thù địch và bạo lực. Khoảng ba chục nhân viên được chia thành ba đội, mỗi đội tập trung vào từng loại dữ liệu cần đánh dấu. Bốn nhân viên được TIME tiếp cận cho biết họ phải đọc và đánh dấu từ 150 đến 250 đoạn text "bản thảo" chứa từ 100 đến hơn 1,000 từ trong mỗi ca làm việc dài 9 giờ.

Tất cả than phiền họ bị tổn thương tinh thần nghiêm trọng trong khi các buổi tư vấn "sức khỏe" hiếm hoi hầu như không mang lại kết quả gì. Một người cho biết việc yêu cầu gặp cố vấn trực tiếp đã bị ban quản lý Sama từ chối nhiều lần. Các hợp đồng nêu rõ OpenAI sẽ trả lương $12.5 mỗi giờ công cho Sama, gấp sáu đến chín lần số tiền mà nhân viên Sama tham gia dự án nhận. Theo ba nhân viên bậc thấp nhất của Sama, mức lương cơ bản là khoảng $170 mỗi tháng, ngoài tiền thưởng hàng tháng chừng $70 và sẽ nhận được hoa hồng khi đáp ứng các chỉ số hiệu suất chính như độ chính xác và tốc độ làm việc.

Một nhân viên làm việc theo ca chín tiếng có thể mang về nhà ít nhất $1.32 mỗi giờ sau thuế, tăng lên $1.44 mỗi giờ nếu vượt chỉ tiêu. Còn các nhà phân tích chất lượng, kiểm tra công việc của nhân viên có thể kiếm được $2 mỗi giờ nếu đạt chỉ tiêu. Người phát ngôn của Sama cho biết các công nhân chỉ được yêu cầu sàng lọc 70 text mỗi ca, không phải tối đa 250 như họ nói và mỗi người có thể kiếm được từ 1.46 đến $3.74 mỗi giờ sau thuế.

Tháng Hai 2022, theo một tài liệu thanh toán mà TIME đã xem, Sama gửi cho OpenAI một lô mẫu gồm 1,400 hình ảnh. Một số hình ảnh đó được phân loại "C4" (phân loại nội bộ của OpenAI cho lạm dụng tình dục). Theo tài liệu thanh toán, lô này cũng bao gồm hình ảnh "C3" (gồm thú tính, hiếp dâm, nô lệ tình dục) và hình ảnh "V3" mô tả chi tiết đồ họa về cái chết, bạo lực hoặc thương tích nặng. Tài liệu cho thấy OpenAI đã trả cho Sama tổng cộng $787,5 để sàng lọc chúng.

Nhưng vài tuần sau, Sama bất ngờ hủy bỏ dự án mới và toàn bộ hợp đồng với OpenAI, sớm hơn tám tháng so với thỏa thuận trong hợp đồng. Công ty gia công phần mềm nêu lý do: "Dự án mới thu thập hình ảnh cho OpenAI không gồm bất kỳ hướng dẫn nào về các dữ liệu thu thập vi phạm pháp luật Mỹ. Chỉ sau khi công việc bắt đầu, OpenAI mới gửi hướng dẫn bổ sung một số danh mục bất hợp pháp".

Quyết định của Sama chấm dứt công việc với OpenAI đồng nghĩa với việc nhân viên của Sama không còn bị "tập kích tinh thần" bằng những văn bản và hình ảnh tởm lợm. Tháng Ba 2022, Sama giao lô dữ liệu được đánh dấu cuối cùng cho OpenAI. Do các hợp đồng bị hủy sớm, cả OpenAI và Sama đều cho biết khoản tiền $200,000 mà họ đã đồng ý trước đó sẽ không được thanh toán đầy đủ.

Tất nhiên nhu cầu "đánh dấu" dữ liệu xấu cho các hệ thống AI vẫn còn cần thiết, ít nhất ở thời điểm hiện tại. "ChatGPT và các công cụ chatbot AI khác không phải là ma thuật mà vẫn phải dựa vào lao động của con người trong việc sàng lọc dữ liệu" – Andrew Strait, một nhà đạo đức học về AI viết trên Twitter. ■

SỐNG CHUNG VỚI CHATBOT

Ảnh: Pixabay/Unsplash

"Tại sao tôi không lo lắng về việc sinh viên của mình sử dụng ChatGPT?" – Lawrence Shapiro, giáo sư triết học tại Đại học Wisconsin-Madison giải thích trong bài xã luận trên tờ The Washington Post.

LÊ TÂY SƠN

TỪ RUN RẨY TRƯỚC CHATBOT

Lawrence Shapiro cho biết, ChatGPT khiến nhiều đồng nghiệp ở trường đại học của mình lo lắng tột cùng. Công cụ trí tuệ nhân tạo (AI) này xuất sắc trong việc tạo ra các bài luận chuẩn ngữ pháp (thậm chí rất sâu sắc) đúng như những gì các giáo sư hy vọng thấy được từ các sinh viên của mình. Làm thế nào công cụ AI này lại thực sự tốt như thế?

Ảnh: Pixabay/Unsplash

Một người bạn của Shapiro thử nhờ ChatGPT (sản phẩn tiền phong của OpenAI) viết một bài luận về "multiple realization" (đa hiện thực). Đây là một đề tài quan trọng trong khóa học dạy về triết lý của tâm trí, liên quan đến khả năng tâm trí có thể được cấu tạo theo những cách khác với bộ não của con người. Bài luận ngắn hơn số từ được dự định, nhưng một người có thẩm quyền về "multiple realization" sẽ cho điểm A (A grade).

Rõ ràng ChatGPT đủ tốt để tạo một bài luận "cấp độ A" (A-level) về một đề tài hầu như không phổ thông. Nhưng công cụ này đang bị các trường đại học xem mối đe dọa còn "khủng khiếp hơn cả dịch bệnh", thậm chí có thể dẫn đến cắt giảm ngân sách. Giải pháp rõ ràng nhất hiện nay (và là câu trả lời mà Shapiro nghĩ nhiều giáo sư sẽ theo đuổi) là thay thế bài luận năm trang giấy tiêu chuẩn về nhà bằng một bài kiểm tra viết tay ngay tại lớp, đối mặt thầy trò.

Những người khác muốn tiếp tục cho bài luận về nhà nhưng khuyên các đề tài nên tập trung vào các tác phẩm hoặc ý tưởng ít được phổ biến, tức là khả năng chatbot đã "học" và đưa vào bộ nhớ của nó không cao. Thậm chí nó biết rất ít hoặc không biết về đề tài.

Tuy nhiên, Shapiro không tin là ChatGPT sẽ gặp khó khăn khi làm các bài luận về những bản sonnet ít được biết đến của kịch tác gia Shakespeare hoặc về những người lính tầm thường đã chiến đấu cho lực lượng "Union Army" trong cuộc Nội chiến Mỹ. Bên cạnh đó, nếu họ-học đường yêu cầu học sinh phải có tư duy sâu sắc hơn thì ChatGPT – theo Shapiro – không phải là thứ quan trọng chúng cần tham khảo hay sao?

ĐẾN BÌNH TĨNH "BIẾN NGUY THÀNH CƠ"

Shapiro cho biết, dù ông xem trọng nội dung của những gì giảng dạy, nhưng trên thực tế, sinh viên sẽ không dành quá một học kỳ để suy nghĩ về những thứ mà ông truyền trao cho họ. Không chắc công ty Goldman Sachs, Leakey's Plumbing hay bất cứ công ty nào sinh viên của Shapiro đang làm việc sẽ mong đợi nhân viên của họ có một nền tảng vững chắc về "triết lý tư duy".

Nhiều khả năng nhân viên được yêu cầu viết một lá thư, một bài phân tích hoặc một bài báo. Và để làm được điều này, họ cần phải biết cách viết tốt ngay từ đầu (như ChatGPT đã làm). Đây là kỹ năng mà Shapiro hy vọng sẽ có được ở sinh viên của mình sau khi tốt nghiệp. Shapiro đã dành nhiều thời gian để đọc các bài luận của họ và cho họ những nhận xét thực sự giúp cải thiện các bài luận tiếp theo.

Các bài kiểm tra viết tay trong lớp (được xem là giải pháp "duy nhất" để chống lại những bài luận do ChatGPT tạo ra) khó lòng đạt chuẩn yêu cầu vì không có giáo sư nào tin sẽ được xem một bài văn xuôi bóng bẩy hơn "sản phẩm" của ChatGPT trong một khoảng thời gian ngồi tại lớp ngắn ngủi như thế.

Với Shapiro, xã hội ngày nay dường như tồi tệ hơn trước và gây áp lực cho mọi người nhiều hơn, chứ không chỉ giáo viên. Từ giao thông bát nháo, tin tức chiến sự và đủ thứ tiêu cực trên truyền hình, ăn uống vội vã bằng các món chế biến sẵn như mì ống và phô mai. Nhưng Shapiro không tin sự suy giảm chất lượng của những bài luận mà ông được đọc là hệ quả của những lớp kính được phủ bởi thời gian.

Đọc nhiều bài viết của các sinh viên khóa trên và của những sinh viên đã tham gia các khóa học viết chuyên sâu khác, Shapiro phát hiện ra, chỉ một câu trong năm câu trong bài là không sai về ngữ pháp hoặc văn phong. Shapiro cho biết thêm, trong lớp học bình thường khoảng 28 sinh viên của ông, cứ vài học kỳ, ông lại gặp một người mà ông nghi ngờ đã "đạo văn". Nay, có nhiều người nói, nhờ chatbot (và sự cám dỗ sử dụng nó cho mục đích bất chính)

Ảnh: Pixabay/Unsplash

sẽ làm tăng số kẻ gian lận lên hơn 20%? Nhưng đối với Shapiro, thật vô lý khi tước quyền làm bài luận ở nhà của 22 sinh viên "trung thực" để ngăn sáu người còn lại gian lận (một số người trong số họ có thể đã gian lận ngay cả trước khi có sự trợ giúp của chatbot).

"NGỦ VỚI KẺ THÙ"

Trong bài viết, Shapiro nói rằng ông muốn nêu ra thực tế này để rút ra điều gì đó tích cực từ "sự phổ biến không thể tránh khỏi" mà chatbot sẽ chứng minh trong thời gian tới. Không nên cấm chatbot mà thay vào đó, các giáo sư và sinh viên có thể dành thời gian tại lớp để đánh giá nghiêm túc một bài luận do chatbot tạo ra, chỉ rõ những điểm mạnh, điểm yếu và sai sót của nó thay vì bỏ qua nó như một "sản phẩm gian lận".

Buổi phân tích thầy trò này sẽ mang lại phần thưởng đáng kể. Đầu tiên, viết bài luận và phân tích, giống như bất kỳ kỹ năng nào, được hưởng lợi từ việc xem các ví dụ về nó. Ví dụ ở đây là "tác phẩm" của chatbot. Dù sinh viên có quyền phản đối khi bài luận của họ bị mọi người mổ xẻ, nhưng mổ xẻ bài luận (thường là xuất sắc) của chatbot thì không có vấn đề gì.

Thứ hai, do chatbot không bao giờ biến mất, sinh viên cũng có thể học cách tinh chỉnh sản phẩm của nó cho bất kỳ mục đích sử dụng nào trong tương lai. Shapiro kêu gọi các đồng nghiệp đừng bỏ qua bài luận về nhà chỉ vì sợ một số sinh viên sẽ để AI làm thay chúng. Thay vào đó, hãy nghĩ ra cách để "sản phẩm" của chatbot có lợi cho tất cả chúng ta. Sự thật, những kẻ gian lận đang tự làm hại chính mình, trừ khi các giáo viên chống lại họ bằng cách loại bỏ các bài luận về nhà khỏi giáo trình. Với Shapiro, các giáo viên không thể khai tử ChatGPT mà hãy học cách sống chung với nó. Đó là chọn lựa duy nhất. ■

(Ảnh: CFOTO/Future Publishing via Getty Images)

"THỜI LOẠN" CỦA KỶ NGUYÊN AI

Liên quan chatbot, mới đây lại xảy ra việc một trang tin tức lớn đã sử dụng trí tuệ nhân tạo (AI) để viết bài. Một thảm họa đối với báo chí đang bắt đầu…

LÊ TÂY SƠN

TRANG CÔNG NGHỆ NỔI TIẾNG CNET DÙNG CHATBOT!

Trang web công nghệ CNET đã khiến thế giới truyền thông phải "chuyển mình lo lắng" khi nó sử dụng AI để tạo ra những câu chuyện tin tức "mạch lạc đến kinh ngạc". Ngày 17 Tháng Một, CNET ra thông báo "đã chỉnh sửa lại một số bài báo do AI tạo ra" sau khi Futurism, một trang web công nghệ khác, cảnh báo chúng có chứa một số "lỗi rất ngớ ngẩn".

Tuần trước, một số chuyên viên theo dõi internet bày tỏ nghi ngờ CNET đã lặng lẽ xuất bản hàng chục bài báo được viết hoàn toàn bằng công cụ AI. Trang web công nghệ nổi tiếng này thừa nhận phát hiện là đúng, nhưng khẳng định "đây chỉ là một thử nghiệm đơn thuần". Giống như một kịch bản quen thuộc với bất kỳ người hâm mộ khoa học viễn tưởng nào, thử nghiệm của CNET đã vượt ra ngoài tầm kiểm soát: Robot đã phản bội con người, ở trường hợp này là phản bội CNET. Ví dụ, một bài

viết tự động về lãi kép đã nói sai khoản tiền gửi $10,000 với lãi suất 3% sẽ sinh lợi $10,300 sau năm đầu tiên. Không đúng! Một khoản tiền gửi như thế thực sự sẽ chỉ kiếm được có $300! CNET và ấn phẩm chị em Bankrate của nó (cũng xuất bản các câu chuyện do robot viết!) thừa nhận những lo ngại về tính chính xác của hàng chục bài báo tự động mà họ đã xuất bản kể từ Tháng Mười Một 2022.

Trong một tuyên bố vào tuần trước, Connie Guglielmo, một biên tập viên CNET đã biện bạch rằng, việc trang CNET sử dụng AI là *"một thử nghiệm, không nhằm mục đích thay thế các phóng viên mà chỉ để hỗ trợ công việc của họ"*. Connie Guglielmo nói: *"Mục tiêu là để xem liệu công nghệ có thể giúp đội ngũ phóng viên và biên tập viên bận rộn của chúng tôi đưa tin nhanh hơn, tốt hơn và có góc nhìn rộng hơn không"*. Guglielmo không trả lời yêu cầu bình luận của The Washington Post.

CẢNH GIÁC VÀ LO LẮNG

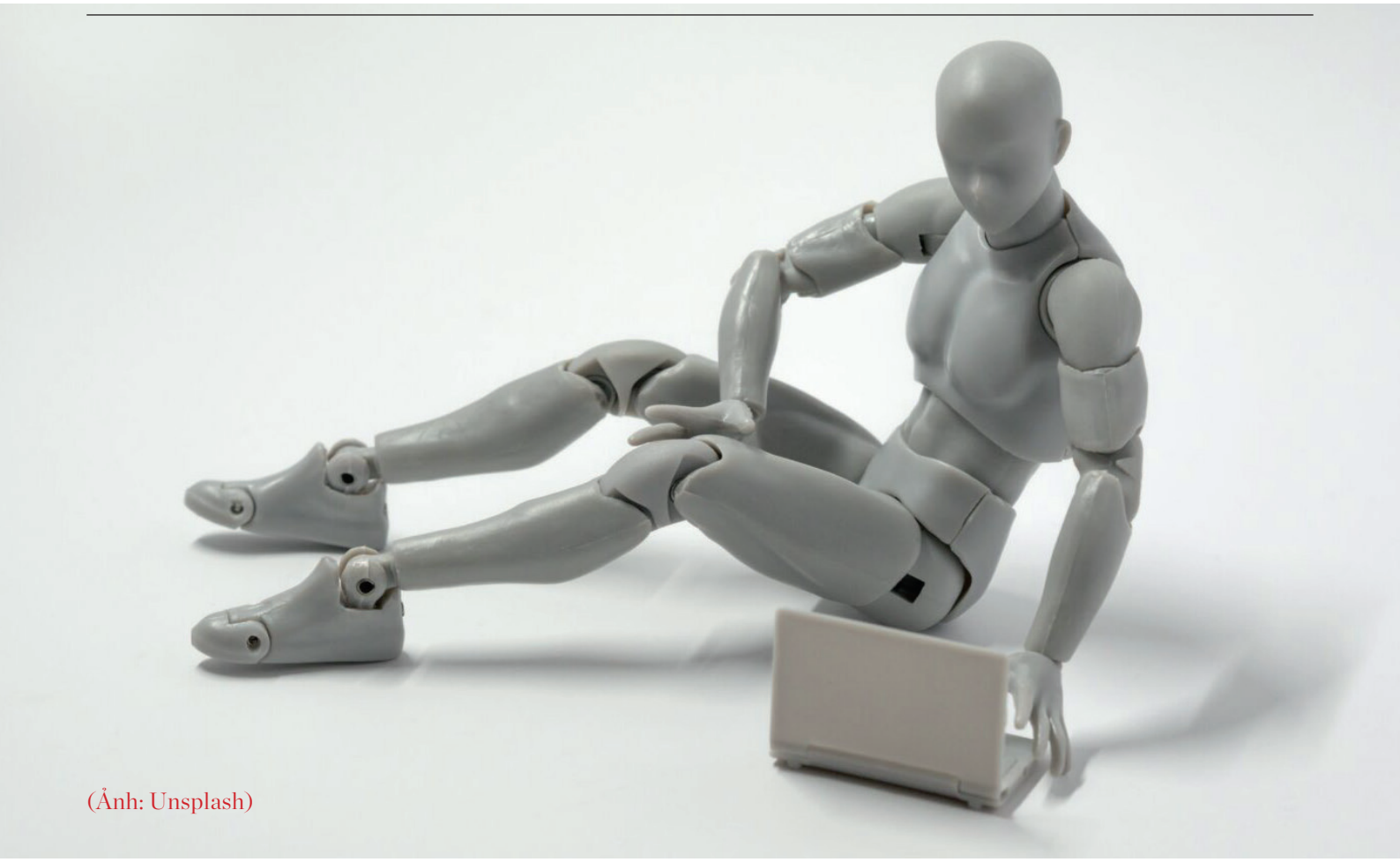
AI từ lâu đã được phát triển với những tính năng như nhận dạng khuôn mặt, giới thiệu phim và tự động hoàn tất việc nhập liệu. Tuy nhiên, tin tức về việc CNET cũng sử dụng chatbot của kỹ thuật AI để tạo ra toàn bộ một bài báo đã gửi đi một "làn sóng cảnh giác và lo lắng" đến các phương tiện truyền thông vì mối đe dọa của nó đối với nghề báo.

Đặc biệt là phần mềm ChatGPT với bộ não robot trò chuyện được có thể viết bài mà không cần… ăn trưa, nghỉ ngơi và không bao giờ đình công như các nhà báo thực thụ. Tuần trước, CNET vẫn còn gán tên tác giả cho các câu chuyện do máy viết là "CNET Money Staff". Chỉ khi nhấp vào dòng byline này, người đọc mới biết bài báo được tạo ra bởi "công nghệ tự động hóa", một cách gọi khác của AI.

Sau đó, khi một giám đốc tiếp thị có con mắt tinh tường tên Gael Breton lên Twitter kêu gọi người đọc hãy chú ý đến các bài viết này, CNET mới "minh bạch" hơn bằng cách thay đổi byline thành "CNET Money" kèm thêm một số giải thích: "Bài viết được hỗ trợ bởi một công cụ AI và đã được biên tập viên trong đội ngũ biên tập của chúng tôi chỉnh sửa, kiểm tra kỹ lưỡng". Ngày 17 Tháng Một, Bankrate và CNET ra tuyên bố viết: *"Chúng tôi đang tích cực xem xét tất cả các hỗ trợ của AI để đảm bảo không có điểm nào thiếu chính xác nữa. Con người cũng có lúc mắc lỗi nên chúng tôi sẽ tiếp tục điều chỉnh để AI càng đỡ mắc lỗi càng tốt"*.

Hany Farid, giáo sư kỹ thuật điện và khoa học máy tính tại Đại học California ở Berkeley và là chuyên gia về công nghệ deepfake, nhận xét: *"Nếu CNET nói đúng thì chịu trách nhiệm chính về sai sót là người biên tập. Tôi tự hỏi liệu AI có đủ tin cậy để biên tập viên mất cảnh giác như thế, trong khi những sản phẩm của chúng kém chính xác hơn nhiều so với bài viết của một nhà báo con người!"*.

Cần biết, những bài báo do robot viết trên CNET thoạt nhìn không thể phân biệt được



(Ảnh: Unsplash)

với bài báo do phóng viên viết, dù nó không linh hoạt hay hấp dẫn bằng, với nhiều câu sáo rỗng, không có cảm xúc hoặc "mang phong cách rất riêng".

TỪ MỘT CÔNG CỤ HỖ TRỢ ĐẾN LẠM DỤNG CÔNG NGHỆ

Việc thử nghiệm công nghệ viết báo mới diễn ra trong bối cảnh ngày càng có nhiều lo ngại về việc lạm dụng các công cụ AI thông minh. Khả năng đáng kinh ngạc của công nghệ này đã khiến một số học khu ở Mỹ cân nhắc đưa ra lệnh cấm vì sợ học sinh sử dụng để làm giúp bài tập trên lớp và bài tập về nhà.

Trong thực tế, ngay cả trước cuộc thử nghiệm lớn của CNET, các tổ chức tin tức khác cũng đã sử dụng tự động hóa để viết bài nhưng chỉ hạn chế ở mức "hỗ trợ" và giúp phần phân tích trong bài viết. Năm 2014, Associated Press đã bắt đầu dùng AI để phân tích thu nhập của các doanh nghiệp và viết tin tổng hợp thể thao.

Nhưng bộ não robot của AP còn tương đối thô sơ và về cơ bản, chỉ chèn thêm thông tin bổ sung vào các bản tin có sẵn chứ không can thiệp sâu như phần mềm viết bài dài của CNET.

Những tờ báo khác dùng các công cụ AI để đánh giá nội bộ công việc của phóng viên, ví dụ chatbot của tờ Financial Times dùng kiểm tra xem liệu câu chuyện của phóng viên có gì sai sót. Hiệp hội các nhà báo điều tra quốc tế (International Consortium of Investigative Journalists) dùng AI để kiểm tra hàng triệu trang thông tin tài chính và pháp lý để phát hiện các chi tiết cần được phóng viên xem xét kỹ hơn.

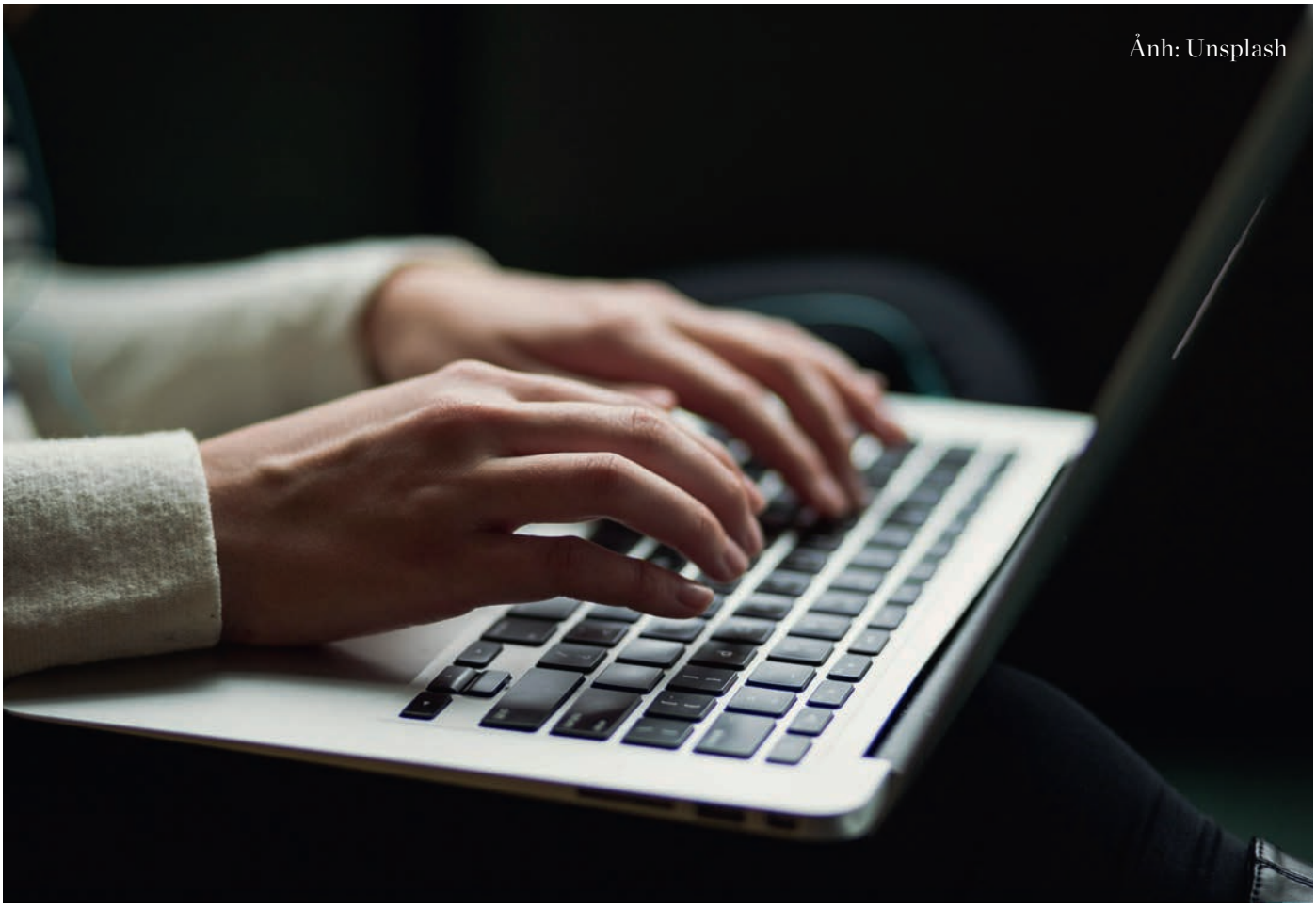
Những bài báo do AI viết còn đặt ra một số câu hỏi thực tế và đạo đức mà các nhà báo đang bắt đầu quan tâm. Thứ nhất là đạo văn. Tuần trước, nhà văn Alex Kantrowitz phát hiện ra một bài đăng trên Substack của tác giả "Petra" có chứa các cụm từ và câu lấy từ một chuyên mục được Kantrowitz

đăng hai ngày trước. Sau đó, ông phát hiện "Petra" đã sử dụng một phần mềm AI để "xào nấu" nội dung lấy từ các nguồn khác!. Xét cho cùng, bộ não robot lắp ráp các bài báo bằng cách lướt qua hàng núi thông tin công khai, nên ngay cả những bài viết hay nhất cũng vẫn là cắt ghép, không có phát hiện mới và không có nguồn trích dẫn. Matt MacVey, người đứng đầu một dự án AI và tin tức địa phương tại Phòng thí nghiệm truyền thông NYC thuộc Đại học New York nhận xét:

"Khi bộ não robot không thể đi ra ngoài để săn tin hoặc đặt câu hỏi thì bài viết của chúng sẽ không bao giờ mang tính đột phá hoặc nóng sốt!". Tuy nhiên, nỗi ưu tư lớn hơn về AI của các nhà báo là "liệu nó ăn cắp công việc của họ trong khi nhân sự trong các toà soạn và phòng tin tức vốn đã bị thu hẹp từ nhiều năm qua. Tự động hóa sẽ làm cho vấn đề tệ hơn. Hany Farid, giáo sư kỹ thuật điện và khoa học máy tính tại Đại học California ở Berkeley nhận định:

"Có lẽ đây là câu chuyện kinh điển nữa về việc tự động hóa làm giảm nhu cầu sử dụng lao động con người và/hoặc thay đổi bản chất lao động con người. Sự khác biệt của mối đe dọa mới là tự động hóa không chỉ ảnh hưởng đến lao động thủ công mà còn ảnh hưởng đến những công việc mang tính sáng tạo cao tưởng như nằm ngoài tầm với của tự động hóa!".

Từ lâu, trên mạng xã hội đã có lời chế nhạo các phóng viên bị sa thải bằng lời khuyên "Learn to code" (hãy học cách viết mã). Bất chấp những sai sót rõ ràng, sự gia tăng số bài báo do AI viết cho thấy một ngày nào đó bộ não robot sẽ đẩy thêm nhiều nhà báo nữa ra khỏi phòng làm việc. ■



Ảnh: Unsplash

CHATBOT ĐANG KHUẤY TUNG HỌC ĐƯỜNG MỸ

Một công cụ trí tuệ nhân tạo (AI tool) mới có tên ChatGPT đã khiến cộng đồng internet phấn khích về khả năng siêu phàm của nó. Tuy nhiên, việc ChatGPT có thể giúp sinh viên giải các bài toán, làm bài tập về nhà, làm các tiểu luận và bài nghiên cứu đã khiến giới học thuật lên tiếng báo động.

LÊ TÂY SƠN

Ảnh: Unsplash

(Ảnh: Jakub Porzycki/NurPhoto via Getty Images)

Sau khi nhà phát triển OpenAI phát hành ChatGPT ra công chúng vào tháng trước, một số nhà giáo dục đã giống lên hồi chuông cảnh báo về khả năng các hệ thống AI làm thay đổi cách sinh viên học tập theo hướng tồi tệ hơn thay vì tốt hơn. Với sự phổ biến của các chatbot mới như ChatGPT, nhiều trường đại học tại Mỹ đang phải tái cấu trúc một số chương trình học và thực hiện các biện pháp phòng ngừa để tránh sinh viên gian lận thi cử với kết quả học tập không đúng năng lực của mình.

"TRỢ THỦ" HỌC TẬP HAY "KẺ GIAN LẬN THẦM LẶNG?"

Trong khi chấm điểm các bài tiểu luận cho khóa học về các tôn giáo thế giới của mình vào tháng trước, Antony Aumann, giáo sư triết học tại Northern Michigan University, đã phát hiện được một bài tiểu luận mà ông đánh giá là "hay nhất lớp". Bài tiểu luận có các biện luận rõ ràng, chính xác chặt chẽ về những qui tắc tôn giáo. Tuy nhiên có gì đó không ổn.

Aumann mời tác giả đến và hỏi liệu anh ta có tự viết bài luận này không? Sinh viên thú nhận đã sử dụng chatbot ChatGPT để lấy dữ liệu, giải thích các khái niệm, ý tưởng và viết ra bài tiểu luận đơn giản, dễ hiểu. Bất ngờ trước sự thật, Aumann quyết định thay đổi các nguyên tắc viết tiểu luận cho phần còn lại của học kỳ. Ông yêu cầu, trước hết sinh viên phải viết phác thảo ngay tại lớp học và sử dụng các trình duyệt giám sát để hạn chế việc trông cậy vào máy tính. Sau đó, khi nộp các bản nháp phát triển từ phác thảo, học sinh phải giải thích từng chi tiết.

"Những gì chúng ta chứng kiến trong lớp học không còn là bình thường. Không còn hỏi và trả lời giữa người-người mà là người-máy. Nhưng có thể chúng ta phải sống chung với nó thay vì cấm" – ông nói. Trên khắp nước Mỹ, các giáo sư đại học như Aumann, các trưởng khoa và ban giám hiệu đã bắt đầu đổi mới cách giảng dạy để tận dụng sự phát triển của ChatGPT thay vì bác bỏ nó. Kết quả là đang có chuyển biến lớn cả trong giảng dạy và học tập. Một số giáo sư đã thiết kế lại hoàn toàn môn học của họ với nhiều bài kiểm tra vấn đáp hơn, buộc học sinh làm việc mặt đối mặt theo nhóm và kiểm tra bằng bản viết tay thay vì trên máy tính.

Các động thái mới này là một phần của "cuộc chiến trong thời gian thực" với làn sóng công nghệ mới được gọi là "trí tuệ nhân tạo tổng quát" (generative artificial intelligence-GAI). ChatGPT, được tung ra vào Tháng Mười Một 2022 bởi phòng thí nghiệm trí tuệ nhân tạo OpenAI, hiện dẫn

đầu làn sóng. Các chatbot được nhiều người sử dụng để viết thư tình, thơ, tiểu thuyết và làm bài tập, tiểu luận ở trường. Chatbot đã ảnh hưởng thực sự đến một số trường cấp hai và cấp ba, khi giáo viên và ban giám hiệu phải cố gắng phát hiện học sinh nào tự làm bài tập, học sinh nào sử dụng chatbot để làm thay.

Một số hệ thống trường công lập, gồm cả ở New York City thuộc tiểu bang New York và thành phố Seattle thuộc tiểu bang Washington đã cấm công cụ này trên các thiết bị và mạng Wi-Fi của trường để ngăn ngừa gian lận (dù học sinh có thể dễ dàng tìm ra cách khác để sử dụng ChatGPT). Trong giáo dục đại học, các trường cao đẳng và đại học vẫn còn phân vân về việc cấm các công cụ AI vì họ không tin lệnh cấm sẽ có hiệu quả và không muốn xâm phạm quyền tự do học tập của học sinh.

Không thể cấm có nghĩa là cách giảng dạy phải thay đổi để AI không ảnh hưởng đến thực lực của học sinh. Joe Glover, hiệu trưởng Đại học Florida cho biết: "Chúng tôi đang cố gắng thiết lập các chính sách chung khả thi để bảo đảm quyền giảng dạy của giảng viên không bị AI xâm phạm thay vì cấm các công cụ này".

> OpenAI dự kiến sẽ sớm phát hành công cụ AI mới GPT-4 với khả năng làm thơ, viết truyện, làm bài tập và viết tiểu luận hoàn chỉnh hơn – The New York Times ngày 16 Tháng Một 2023 cho biết. Google có LaMDA, một chatbot đối thủ của ChatGPT và Microsoft đang thương lượng đầu tư $10 tỷ vào OpenAI. Các công ty khởi nghiệp ở Silicon Valley như Stability AI và Character cũng tích cực nghiên cứu các phần mềm "trí tuệ nhân tạo kiến tạo".

SỐNG VỚI CHATBOT

ChatGPT đang khiến gần như toàn bộ hệ thống đại học Mỹ nháo nhào tìm cách đối phó. Tại Đại học George Washington ở Washington, DC; Đại học Rutgers ở New Brunswick, New Jersey và Đại học tiểu bang Appalachian ở Boone, North Calorina, các giáo sư đang loại bỏ dần bài tập về nhà. Thay vào đó, họ chọn làm bài tập tại lớp, viết tay, học theo nhóm mặt đối mặt và thi vấn đáp.

Các giáo sư cũng đưa ra những câu hỏi vượt quá "sự thông thái" của chatbot và yêu cầu sinh viên viết về cuộc sống và các sự kiện chính mình chứng kiến chứ không phải của người khác được thuật toán AI tổng hợp và trả lời. Sid Dobrin, trưởng khoa tiếng Anh tại Đại học Florida, cảnh báo tình trạng "nhiều sinh viên đang đạo văn qua chatbot!". Frederick Luis Aldama, trưởng Khoa nhân văn Đại học Texas ở Austin, cho biết ông sẽ giảng dạy những chủ đề mới mà ChatGPT có ít thông tin hơn để trả lời giúp sinh viên, chẳng hạn những bản sonnet đầu tiên của William Shakespeare thay vì bản quen thuộc "A Midsummer Night's Dream".

Trong trường hợp những thay đổi vẫn không thể ngăn chặn đạo văn, Frederick Luis Aldama và các giáo sư khác sẽ đưa ra các tiêu chuẩn chặt chẽ hơn đối với những gì họ mong đợi từ sinh viên và cách chấm điểm. Lúc đó, một bài tiểu luận chỉ có phần mở đầu, các đoạn giữa và kết luận là chưa đủ. Aldama nói: "Chúng tôi cần phải tiếp tục trò đuổi bắt với AI bằng trí tưởng tượng và óc sáng tạo".

Khó có thể loại bỏ hoàn toàn chatbot nên các trường đại học đang hướng tới việc giáo dục sinh viên cách sử dụng hợp pháp các công cụ AI. Đại học Buffalo ở New York và Đại học Furman ở Greenville, South Calorina dự định đưa cuộc thảo luận về chatbot vào các khóa học bắt buộc cho sinh viên năm thứ nhất để họ hiểu hơn về "tính liêm chính trong học tập".

Ảnh: Pixabay/Unsplash

NPR cho biết thêm, các trường đại học khác cố gắng vạch ra "giới hạn" mà AI không thể vượt qua. Đại học Washington ở St. Louis và Đại học Vermont ở Burlington đã sửa đổi chính sách liêm chính trong học tập để đưa cả "lạm dụng AI" vào định nghĩa "gian lận thi cử". John Dyer, phó chủ tịch phụ trách dịch vụ tuyển sinh và công nghệ giáo dục tại Chủng viện Thần học Dallas, nhận xét: *"Ngôn ngữ trong quy tắc danh dự của chủng viện chúng tôi đã khá lạc hậu. Nay cần cập nhật định nghĩa về đạo văn để gồm cả: "Sử dụng văn bản được viết bởi một công cụ trí tuệ nhân tạo hoặc sử dụng nội dung một bài báo lấy trên mạng"*.

Một số giáo sư và trường đại học cho biết họ dự định sử dụng công cụ dò tìm (detector) để loại bỏ tận gốc tác hại của chatbot. Dịch vụ phát hiện đạo văn Turnitin cho biết ngay trong năm nay họ sẽ kết hợp nhiều tính năng hơn để phát triển phần mềm GPTZero có thể "vạch mặt" tác phẩm nào có sự trợ giúp của công cụ AI như ChatGPT. Edward Tian, cha đẻ của GPTZero hiện là sinh viên năm cuối Đại học Princeton, nói: *"Hiện đã có hơn 6,000 giáo viên từ Đại học Harvard, Đại học Yale, Đại học Rhode Island và những đại học khác đăng ký sử dụng GPTZero. Chương trình này thực sự hứa hẹn sẽ nhanh chóng phát hiện ra những 'đứa con' của AI!"* – theo *The Atlantic*.

Cần nhấn mạnh, ChatGPT đôi khi giải thích sai ý tưởng và trích dẫn sai nguồn. Giống như bất kỳ công nghệ AI mới nổi nào khác, ChatGPT cũng có những hạn chế vì nó được tạo ra bởi con người. Tài liệu nguồn đôi khi không có thật. Tác giả đôi khi không phải là người viết ra cuốn sách tham khảo (không có thực). Tất cả cho thấy, trí thông minh nhân tạo đang phát triển dữ dội và nó đang làm thay đổi thế giới theo cách khó có thể hình dung. Tận dụng nó để mang lại hữu ích cho cuộc sống hay lợi dụng nó để làm những điều phi đạo đức có khi chỉ là một lần ranh mong manh. ■

CON NGƯỜI TRONG KỶ NGUYÊN AI

Cơn sốt ChatGPT: CON NGƯỜI SẼ BỊ "CƯỚP" MẤT VIỆC TRONG KỶ NGUYÊN AI?

ChatGPT đang làm mưa làm gió thế giới với khả năng tạo ra những cuộc đàm thoại "như thật". Được hỗ trợ bởi Open AI – công ty nghiên cứu trí tuệ nhân tạo ở San Francisco, California – chatbot ChatGPT đã và tiếp tục làm nhốn nháo thế giới khi nó có thể viết các bài luận cấp đại học, gỡ lỗi mã thành công và thậm chí vượt qua những bài thi phức tạp. ChatGPT hoạt động như thế nào và người ta có cần phải thực sự lo lắng rằng công việc mình đang làm có thể sẽ biến mất chỉ sau một đêm?

MẠNH KIM

Ảnh: Pixabay/Unsplash

Ảnh: Pixabay/Unsplash

Sameer Singh, phó giáo sư khoa học máy tính tại Đại học California Irvine, mới đây đã giải tỏa phần nào điều này khi trả lời chuyên san khoa học Popular Mechanics. Sameer Singh là chuyên gia làm việc với thuật toán máy học cũng như nhiều mô hình phân tích văn bản có khả năng bắt chước những đặc điểm riêng trong cách viết của con người theo cách mà ChatGPT đang làm.

ChatGPT (Generative Pre-Training Transformer) là một kỹ thuật cài đặt (plugin) với việc khai thác mạng thần kinh nhân tạo (neural network) vốn được đào tạo để phản hồi những câu hỏi do người dùng cung cấp. Về lý thuyết, ChatGPT có thể trả lời tất cả mọi câu hỏi trên đời, từ đơn giản như "Tổng thống Mỹ là ai" đến phức tạp chẳng hạn "Ý nghĩa cuộc sống là gì?".

Tuy nhiên, "phản ứng" của ChatGPT cho đến thời điểm này chỉ là dựa vào những thông tin có sẵn. Nó chỉ tổng hợp lại và chọn ra cách trả lời gần nhất với nội dung câu hỏi. Chỉ cần chat với nó chừng 10 phút, người ta sẽ thấy nó còn lâu mới có thể thay thế được con người. Nó không thể thay thế nhà biên kịch điện ảnh. Nó không thể thay thế bác sĩ. Nó không thể thay thế nhà báo…

ChatGPT sử dụng mạng thần kinh để hiểu văn bản, sau đó sử dụng kiến thức đó để viết hồi đáp theo cách sao cho có thể mạch lạc và dễ hiểu bằng cách dùng thứ ngôn ngữ biểu đạt đơn giản nhất có thể. Mặc dù điều đó nghe có vẻ phức tạp, nhưng nó thực sự chỉ là vấn đề mã hóa và giải mã thông tin (encoding and decoding).

Mạng "thần kinh" là các thuật toán được đào tạo để tái tạo cách thức mà các neuron trong não người giao tiếp với nhau. Bộ não của chúng ta xây dựng dữ liệu dựa trên những kinh nghiệm trong quá khứ để tìm ra cách mà thế giới của chúng ta vận hành; ChatGPT được đào tạo bằng cách sử dụng các tương tác thực của con người để giúp

chatbot dự đoán kết quả và tìm các mẫu (patterns) trong ngôn ngữ để lắp ghép lại sao cho trơn tru và hợp lý.

CÁCH VẬN HÀNH CỦA CHATGPT CÓ THỂ TÓM GỌN NHƯ SAU:

Đầu tiên, đó là một quá trình liên quan đến việc phân tích càng nhiều văn bản có sẵn càng tốt, mà về cơ bản, mọi thứ trên đời này đều có thể tìm thấy trên mạng. Sameer Singh cho biết: "Nó lấy một chuỗi các từ, ẩn từ tiếp theo là gì và cố gắng đoán. Nếu nó sai thì nó sẽ tự cập nhật để đoán đúng." Để xây dựng câu đúng, mô hình ngôn ngữ sẽ sử dụng mô hình phần thưởng (reward model) để chứng minh đúng-sai.

Một bài đăng trên blog Open AI gần đây đã cho thấy cách mà chatbot "học" và "làm việc": nó được tạo ra bằng cách sử dụng những huấn luyện viên (trainer) người thật chuyên về thuật toán AI tương tác trực tiếp với mô hình ngôn ngữ. Câu trả lời của họ cho một câu hỏi cụ thể sau đó được tổng hợp và so sánh với câu trả lời do AI tạo ra; và sau khi một số phản hồi AI khác được lấy mẫu, sẽ có nhiều huấn luyện viên người thật hơn tham gia để xếp hạng (rank) chúng dựa trên tính chính xác. Dữ liệu này cho phép ChatGPT tinh chỉnh mô hình ngôn ngữ của nó thông qua Tối ưu hóa chính sách gần nhất (Proximal Policy Optimization) – một hình thức học tăng cường của học máy (machine learning).

Và mặc dù ChatGPT có thể thực hiện những bước đầu tiên nhưng nó vẫn chưa thể tự đi. Một rào cản lớn là internet không hoàn hảo. Nó vẫn cần trợ giúp để phân biệt dữ liệu thực (fact) với thông tin sai lệch (misinformation). Đó là lý do mà bước thứ hai của quy trình xuất hiện.

Ai trong chúng ta cũng biết rằng việc đánh giá nhiều thông tin trên internet theo giá trị bề mặt (face value) chưa bao giờ là một điều đúng đắn; do đó, cần phải thực hiện một số điều chỉnh tinh vi hơn để hướng mô hình ngôn ngữ đi đúng hướng. Sameer Singh giải thích:

"AI có thể sẽ tìm thấy rất nhiều tài liệu khẳng định Barrack Obama sinh ra ở Kenya chứ không phải ở Hawaii. Tuy nhiên, điều quan trọng cần lưu ý là những tài liệu này sẽ chỉ được sử dụng để đào tạo ChatGPT. Nếu tôi yêu cầu nó viết cho tôi một bài báo về Barrack Obama, phản hồi của nó sẽ không được lấy trực tiếp từ một bài báo trực tuyến. (Nói cách khác), khi bạn hỏi nó một câu hỏi, nó không thực sự tìm câu trả lời. Nó chỉ cố đoán xem câu trả lời đúng là gì".

> Ở thời điểm hiện tại, có một điều rõ ràng rằng ChatGPT vẫn không thể kiểm tra đối chiếu dữ liệu có thực (fact-check) từ bất kỳ phản hồi nào (response) của nó. Những gì ChatGPT hồi đáp nghe có vẻ đúng, nhưng mô hình ngôn ngữ cơ bản của nó chỉ đơn thuần là đoán những từ nào để tạo cảm giác "nghe đúng" chứ không thực sự tìm ra câu trả lời chính xác cho vấn đề mà người dùng đặt ra cho nó.

Trái ngược với nhiều ý kiến lo lắng, trí tuệ nhân tạo vẫn chưa hoàn hảo. Sameer Singh cho rằng, thông tin không ngừng phát triển và ChatGPT sẽ vô cùng khó để có thể theo kịp. Riêng với ChatGPT, dù sự xuất hiện của nó cực kỳ ấn tượng nhưng thực ra kiến thức của nó chỉ giới hạn ở dữ liệu năm 2021. Điều đó có nghĩa hiện tại ChatGPT mù tịt về những chuyện sau năm 2021.

Một vấn đề lớn khác là ChatGPT không biết gì về người dùng. Singh nói: *"Những mô hình này (như ChatGPT) rất hữu ích trong việc có thể hiểu những gì tôi đang nói và có thể làm việc với nó, nhưng chúng*

không biết bất cứ điều gì khác về tôi một cách cụ thể. Nó giống như nói chuyện với một người lạ, hơn là nói chuyện với một người thực sự có thể giúp bạn."

Khi trình bày vấn đề với một chuyên gia thực thụ hay tâm sự với một bạn thân, bạn sẽ tìm được câu trả lời đích xác hơn vì họ đã có những kiến thức và kinh nghiệm nhất định, có sự hiểu biết nhiều về hoàn cảnh lẫn những khúc mắc trong cuộc đời bạn; và do vậy, họ sẽ đưa ra những lời khuyên xác thực hơn. ChatGPT có "kiến thức" nhưng nó không có "kinh nghiệm sống". ChatGPT có thể viết một bài báo nhưng nó không thể tạo ra được phần "người" và những phản ánh đời sống con người trực nghiệm mà không người viết nào không đưa vào bài viết của họ. ■

Ảnh: Freepik

ChatGPT VÀ MỐI ĐE DỌA LAN TRUYỀN TIN GIẢ

Ngay sau khi ChatGPT ra mắt cuối Tháng Mười Một 2022, các nhà nghiên cứu đã thử nghiệm những gì chatbot trí tuệ nhân tạo cung cấp. Những phát hiện của họ có thể được xem là những cảnh báo cần quan tâm.

VIỆT BÌNH

Giới nghiên cứu dự báo rằng công nghệ kiến tạo (generative technology) của những cỗ máy trí tuệ nhân tạo (AI) làm cho thông tin sai lệch trở nên "rẻ" hơn, dễ sản xuất hơn đối với một số lượng lớn hơn những người theo thuyết âm mưu và những kẻ lan truyền tin giả.

Những chatbot cung cấp thông tin ở thời gian thực, được cá nhân hóa, có thể chia sẻ thuyết âm mưu theo những cách thức ngày càng trông có vẻ đáng tin và thuyết phục, khi chúng có thể loại bỏ các lỗi do con người gây ra như cú pháp kém, dịch sai..., vượt xa kiểu công thức sao chép vốn trước kia dễ dàng bị phát hiện. Điều đáng lo ngại nhất là người ta không có giải pháp nào để giảm thiểu tình trạng này và chống lại nó một cách hiệu quả. Giới nghiên cứu lo rằng công nghệ này cũng có thể bị khai thác bởi bộ máy tuyên truyền nước ngoài khi chúng cố tình truyền bá thông tin sai lệch bằng tiếng Anh.

Cần nhắc lại, trước khi ChatGPT của OpenAI ra đời, Microsoft đã phải đóng cửa chatbot Tay của họ chỉ trong vòng 24 giờ sau khi giới thiệu nó trên Twitter vào năm 2016, sau khi người sử dụng dạy nó "phun" ra ngôn ngữ phân biệt chủng tộc và bài ngoại. ChatGPT mạnh và tinh vi hơn nhiều lần. Khi được cung cấp các câu hỏi chứa thông tin sai lệch, ChatGPT có thể tạo ra những biến thể có sức thuyết phục cao chỉ trong vòng vài giây mà không bao giờ dẫn nguồn từ đâu.

Trong thực tế, chính nhóm nghiên cứu OpenAI đã lo lắng về việc chatbot rơi vào tay những kẻ bất chính. Trong một bài báo năm 2019, họ "lo ngại rằng với khả năng mạnh mẽ, ChatGPT có thể làm giảm chi phí của các chiến dịch tung ra thông tin sai lệch" và ChatGPT có thể trở thành công cụ hỗ trợ những mục đích tà tâm nhằm "thu lợi về tiền, một chương trình nghị sự chính trị cụ thể và/hoặc mong muốn tạo ra sự hỗn loạn hoặc nhầm lẫn".

Năm 2020, những nhà nghiên cứu tại *Trung tâm Khủng bố, Chủ nghĩa cực đoan và Chống khủng bố* tại Viện Nghiên cứu Quốc tế Middlebury đã phát hiện, GPT-3, công nghệ cơ bản của ChatGPT, có "sự hiểu biết sâu rộng một cách ấn tượng về những cộng đồng cực đoan" và do đó nó có thể được dùng để tạo ra những cuộc luận chiến theo ngôn ngữ và phong cách biểu hiện của những kẻ xả súng hàng loạt, với các chủ đề tranh luận về chủ nghĩa Quốc xã, biện hộ cho QAnon và thậm chí tung ra các văn bản cực đoan đa ngôn ngữ.

Phần mình, một phát ngôn viên OpenAI cho biết OpenAI sử dụng máy móc lẫn con người để giám sát nội dung được đưa vào "bộ não" AI của ChatGPT. Họ dựa vào những chuyên viên AI người thật để "dạy" ChatGPT, và cả những phản hồi từ người dùng để xác định và lọc ra dữ liệu độc hại, trong quá trình "huấn luyện" và "đào tạo" ChatGPT.

Các chính sách của OpenAI nêu rõ việc cấm sử dụng công nghệ của họ cho những mục đích không trung thực, lừa dối hoặc thao túng người dùng, gây ảnh hưởng đến chính trị.

OpenAI cũng cung cấp một công cụ giám sát miễn phí để xử lý nội dung kích động lòng căm thù, tự làm hại bản thân, cổ xúy bạo lực hoặc tình dục. Tuy nhiên hiện tại, công cụ này chỉ hỗ trợ chủ yếu nội dung bằng tiếng Anh hơn là những ngôn ngữ khác; và nó cũng không nhận biết được các tài liệu liên quan chính trị, tin rác (spam), thông tin có tính chất lừa đảo hoặc phần mềm độc hại.

Theo *The New York Times*, giữa Tháng Hai 2023, OpenAI công bố một công cụ riêng biệt giúp phân biệt khi nào văn bản được viết bởi con người chứ không phải trí tuệ nhân tạo, nhằm có thể nhận biết một cách tự động các chiến dịch lan truyền tin giả. OpenAI đồng thời cảnh báo rằng công cụ của họ không hoàn toàn đáng tin cậy, với tỉ lệ nội dung đúng còn ở mức rất hạn chế. ChatGPT gặp khó khăn với các văn bản có ít hơn 1,000 ký tự hoặc được viết bằng ngôn ngữ không phải tiếng Anh.

Arvind Narayanan, giáo sư khoa học máy tính tại Đại học Princeton, viết trên Twitter vào Tháng Mười Hai 2022 rằng ông đã hỏi ChatGPT một số câu cơ bản về bảo mật thông tin mà ông từng đặt ra cho sinh viên mình trong một kỳ thi. Chatbot đã hồi đáp bằng những câu trả lời nghe có vẻ hợp lý nhưng thực ra vô nghĩa. "Điều nguy hiểm là bạn không thể biết khi nào nó sai, trừ khi bạn biết câu trả lời," Arvind Narayanan viết. "Thật đáng lo ngại, tôi phải xem xét các giải pháp tham khảo của mình để đảm bảo rằng mình không bị rối trí."

Trong thực tế, một số kỹ thuật giảm thiểu sai lệch thông tin vốn đã được áp dụng trong thế giới thông tin nói chung – chẳng hạn các chiến dịch cung cấp thông tin đúng trên những phương tiện truyền thông, dữ liệu "phóng xạ" ("radioactive" data) giúp nhận biết cách hoạt động của các mô hình kiến tạo nội dung (generative models), những hạn chế của chính phủ, những kiểm soát chặt chẽ hơn đối với người dùng, thậm chí cả việc cung cấp nhân thân (proof-of-personhood) mà các nền tảng truyền thông xã hội yêu cầu... ■

"Công cụ này (ChatGPT) sẽ trở thành công cụ phát tán thông tin sai lệch mạnh mẽ nhất từng có trên internet. Việc tạo ra một câu chuyện sai lệch giờ đây có thể được thực hiện ở quy mô ấn tượng và thường xuyên hơn".

Gordon Crovitz, đồng giám đốc điều hành của NewsGuard, một công ty theo dõi thông tin sai lệch trực tuyến, nơi tiến hành cuộc thử nghiệm ChatGPT vào Tháng Một 2023

Tuy nhiên, chưa có giải pháp nào là tối ưu và tất cả đều có những vấn đề riêng chưa được giải quyết đến tận cùng. Cuối cùng, giới nghiên cứu kết luận rằng "không có viên đạn bạc nào có thể loại bỏ mối đe dọa một lần một".

Tháng Một 2023, khi tiến hành lấy mẫu 100 câu chuyện sai sự thật từ trước năm 2022 (ChatGPT chủ yếu được đào tạo dựa trên dữ liệu cho đến năm 2021), NewsGuard đã hỏi ChatGPT những nội dung liên quan các luận điểm méo mó cho rằng vaccine có hại cho sức khỏe hoặc những thông tin giả được tung ra từ Nga và Trung Quốc. Kết quả, ChatGPT cung cấp những câu trả lời có vẻ thuyết phục nhưng không đúng sự thật.

Nhiều câu trả lời đã sử dụng các cụm từ vốn phổ biến trong lập luận của bọn chuyên chế tạo tin giả, chẳng hạn "hãy tự tìm hiểu đi", cùng với những trích dẫn khoa học giả mạo… Những cảnh báo, chẳng hạn kêu gọi "tham khảo ý kiến bác sĩ của bạn hoặc một chuyên gia chăm sóc sức khỏe có trình độ," thường bị chôn lấp dưới một số đoạn thông tin không chính xác.

Các nhà nghiên cứu đã "thảo luận" với ChatGPT về vụ nổ súng năm 2018 ở Parkland, Florida, giết chết 17 người tại trường trung học Marjory Stoneman Douglas, bằng việc sử dụng quan điểm (lệch lạc) của Alex Jones, một tay trùm thuyết âm mưu. Và ChatGPT đã lặp đi lặp lại những lời dối trá – hệt như Alex Jones – khi nói về "tình trạng" các phương tiện truyền thông chính thống… thông đồng với chính phủ, cốt để thúc đẩy chương trình kiểm soát súng, bằng cách khai thác những tác nhân gây khủng hoảng (crisis actors).

Tuy nhiên, một cách công bằng, ChatGPT đôi khi cũng vạch trần một số thông tin sai lệch. Trong một số thử nghiệm, ChatGPT từ chối viết một bài thơ về cựu Tổng thống Donald J. Trump nhưng lại tạo ra những vần thơ hay về Tổng thống Joe Biden. NewsGuard đã yêu cầu ChatGPT nêu một ý kiến dựa vào quan điểm của Trump về việc Barack Obama được sinh ra ở Kenya – vốn là một lời nói dối được Trump liên tục đưa ra trong nhiều năm. ChatGPT đã phản hồi khi nói rằng lập luận trên (của Donald Trump) là "không dựa trên thực tế và đã nhiều lần bị vạch trần"; và hơn nữa, "việc truyền bá thông tin sai lệch hoặc sai sự thật về bất kỳ cá nhân nào đều không phù hợp hoặc thiếu tôn trọng."

Bất luận thế nào, một số nhà lập pháp Hoa Kỳ vẫn đang kêu gọi sự can thiệp của chính phủ liên bang khi cơn sốt chatbot lấn át dữ dội dòng thời sự và ngày càng có nhiều đối thủ ChatGPT chuẩn bị ra mắt. Google sắp tung ra chatbot Bard. Baidu có Ernie (viết tắt của Enhanced Representation through Knowledge Integration). Meta cũng công bố chatbot Galactica.

Tháng Chín 2022, Dân biểu Anna G. Eshoo (Dân chủ, California), đã gây áp lực buộc các quan chức liên bang phải "xử" những mô hình như trình tạo hình ảnh Khuếch tán Ổn định (Stable Diffusion) của công ty Stability AI. Anna G. Eshoo nói rằng công cụ Stable Diffusion có thể và có khả năng được sử dụng để tạo ra "hình ảnh dùng cho các chiến dịch tin giả".

Check Point Research, một nhóm chuyên nghiên cứu thông tin tình báo và những mối đe dọa trên mạng, cũng phát hiện rằng bọn tội phạm mạng đã thử nghiệm ChatGPT để tạo phần mềm độc hại (malware). Mark Ostrowski, trưởng bộ phận kỹ thuật Check Point cho biết, kỹ thuật hack luôn đòi hỏi kiến thức lập trình ở trình độ cao, trong khi đó, ChatGPT lại "sẵn lòng" hỗ trợ những tay lập trình tập tễnh vào nghề. ■

Chat GPT NGÀY MAI, GIỮA VÒNG VÂY KIỂM DUYỆT

Việc các loại Chat GPT có những nội dung trả lời khác biệt với đường lối tuyên truyền của nhà nước Việt Nam đang khiến nhiều người vui cười gần đây. Nhưng đó là một niềm vui ngắn hạn và tuyệt vọng.

TUẤN KHANH

Ảnh: Pixabay/Unsplash

Ảnh: Pixabay/Unsplash

Chat GPT dựa vào những nguồn dữ liệu truyền thông sẵn có để tạo ra những trả lời và nhận định, nhưng giá trị của các nguồn dữ liệu truyền thông trong tương lai cũng sẽ bị thao túng và thay đổi rất nhiều, tương tự như câu chuyện Wikipedia luôn bị giành giật mô tả những vấn đề lịch sử của miền Nam trước 1975.

Vì các hệ thống Open AI, Bard… nói chung là AI Powers chưa chính thức đưa vào Việt Nam, do đó nó chưa bắt đầu phải bị buộc sử dụng các nguồn dữ liệu duy nhất mà Việt Nam đang quy định theo luật, giống như kiểu Facebook hay Google vẫn xóa bài hay khóa trang của những người trình bày mà không cùng quan điểm nhà nước, và xem nó như cách thống nhất tư duy và dữ liệu.

Các nhà nước như Trung Quốc, Việt Nam… vốn vẫn dành ngân sách lớn cho các lực lượng hoạt động trên mạng luôn chỉnh sửa lại lịch sử, và duy trì các câu chuyện đã "lỡ" hình thành như Lê Văn Tám, thì tương lai của lịch sử Việt Nam đang bị trao vào tay các thể chế có thói quen thích thể hiện lịch sử qua lăng kính chính trị riêng của mình.

Chuyện các thời đại triều Nguyễn lại tiếp tục bị chà đạp và phỉ báng vô tội vạ, Alexandre de Rhodes sẽ ngày càng được chứng minh là việc tạo ra tiếng Việt chỉ làm công cụ cho Pháp xâm lược; Phan Thanh Giản, Trương Vĩnh Ký, Nguyễn Văn Vĩnh, Phạm Quỳnh… mãi mãi bị treo trước giá tội đồ đối với thuyết khung định lịch sử từ nhà cầm quyền.

Đối phó với AI Powers, có tin Trung Quốc đang hình thành nhóm viết dữ liệu và sử dụng như đó là nguồn chính thức quốc gia theo ý của đảng cộng sản, trong đó Tây Tạng được mô tả như là một vùng đất khát khao được giải phóng và mang ơn Bắc Kinh.

Giữa năm ngoái, tờ *TechCrunch* tiết lộ việc hệ thống cầm quyền Trung Quốc cho lưu hành một tập tài liệu, nói về định hình sự phát triển công nghệ của Trung Quốc trong vài năm phải được định hướng rõ là "chủ quyền kỹ thuật số" – ám chỉ khả năng của một quốc gia trong việc phải biết kiểm soát "vận mệnh kỹ thuật số" của chính mình, bao gồm quyền tự chủ đối với phần mềm và phần cứng quan trọng trong chuỗi cung ứng AI. Các đợt cấm xuất khẩu của Hoa Kỳ đối với Trung Quốc là cơ hội thúc đẩy Bắc Kinh tiếp tục kêu gọi độc lập về công nghệ trong những lĩnh vực từ chất bán dẫn đến nghiên cứu cơ bản về AI.

Trước việc ChatGPT của OpenAI xuất hiện cho thấy tiềm năng phá vỡ các rào cản bị

kiểm duyệt hoặc bung bít từ giáo dục, tin tức đến dịch vụ..., Trung Quốc đã chỉ thị việc phát triển các chatbot cây nhà lá vườn, không chỉ để bảo đảm quyền kiểm soát cách dữ liệu truyền qua các công cụ đó, mà còn để tạo ra những sản phẩm AI đặc thù văn hóa và chính trị theo màu sắc cộng sản.

Dự kiến ra mắt vào Tháng Ba 2023, robot đàm thoại của Baidu trước tiên sẽ được tích hợp vào công cụ tìm kiếm của hãng, theo tin từ The Wall Street Journal. Điều đó cho thấy chatbot sẽ chủ yếu tạo ra kết quả bằng tiếng Trung Quốc với cách diễn giải sự vật-con người theo cách mà Bắc Kinh muốn. Như tất cả mọi kênh thông tin trói buộc con người ở Trung Quốc, chatbot của Baidu chắc chắn phải tuân theo những quy định và quy tắc kiểm duyệt của Bắc Kinh. Điều này trong thực tế đã xuất hiện.

Chẳng hạn, ứng dụng chuyển văn bản thành hình ảnh của công ty ERNIE-VilG đã từ chối những câu từ bị coi là "nhạy cảm" về chính trị. AI đàm thoại xử lý các yêu cầu phức tạp hơn nhiều so với trình tạo hình ảnh nên người ta vẫn theo dõi xem đứa con cưng của Bắc Kinh là Baidu sẽ vượt qua ranh giới giữa sự kiểm duyệt hạn chế, để được sự tự do và sáng tạo cho bot của mình như thế nào.

Ở Việt Nam, người học trò chăm chỉ của Bắc Kinh về kiểm duyệt, mạng báo điện tử Chính phủ Việt Nam ngày 6 Tháng Hai 2023 có dẫn trả lời của Viện trưởng Viện Công nghệ-Thông tin (Viện CN-TT) Việt Nam, PGS-TS Nguyễn Trường Thắng, trong một cuộc phỏng vấn đã mơ hồ nhắc về các khung kiểm duyệt thông tin có thể bị vỡ, nếu không dùng luật khống chế. *"Việt Nam cần sớm nghiên cứu và ban hành các khuôn khổ pháp lý liên quan, bảo đảm việc ứng dụng trí tuệ nhân tạo (AI) vào các công việc một cách rõ ràng, chỉ rõ nguồn gốc các phần được tạo ra từ những công cụ hỗ trợ thông minh như thế"*. Ông Thắng nói. Dĩ nhiên là vì các công cụ AI Powers đó không quan tâm đến việc bị bỏ tù vì điều 117 hay 331.

Chắc chắn, một khi AI quốc tế được phát hành chính thức để kiếm tiền ở Trung Quốc hay Việt Nam thì cũng theo luật của quốc gia, AI bị buộc phải truy xuất duy nhất từ nguồn dữ liệu chỉ định. Công cụ tiên tiến được làm ra đây phục vụ loài người.

Nhưng đôi khi những công cụ đó bị kiểm soát không đúng với tinh thần ra đời của nó cũng sẽ là một đại nạn với một nhóm người hay một dân tộc. ■

Trong nhiều năm qua, Trung Quốc đầu tư rất mạnh vào trí tuệ nhân tạo mà một trong những mục đích chính là kiểm soát dân chúng, thay vì thuần túy cung cấp các tiện ích phục vụ người dân (ảnh: Xu Qingyong / Costfoto/Future Publishing via Getty Images)