

Khi AI trả lời thay con người và khoảng cách thực tế

ISSN: 2734-9195 15:52 04/10/2026

Trong thời đại AI, có lẽ điều con người cần bảo vệ không phải là đặc quyền được trả lời khảo sát, mà là quyền được hiện diện như một con người cụ thể, phức tạp và không hoàn toàn có thể dự đoán.

Sự phát triển nhanh chóng của trí tuệ nhân tạo (AI) đang mở ra câu hỏi mới trong nghiên cứu dư luận: liệu máy móc có thể trả lời khảo sát thay con người? Một thử nghiệm của Pew Research Center cho thấy câu trả lời, ít nhất ở thời điểm hiện tại, là chưa. Các mô hình AI không chỉ tạo ra những kết quả khác biệt đáng kể so với người trả lời thực tế, mà còn có xu hướng khuếch đại khuôn mẫu, né tránh sự không chắc chắn và đưa ra những hình dung sai lệch về tâm trạng xã hội.

Từ đó, một vấn đề rộng hơn được đặt ra: Khi AI ngày càng có khả năng mô phỏng tiếng nói con người, điều gì làm nên sự khác biệt giữa câu trả lời được dự đoán và tiếng nói thực sự của con người? Đây cũng là khoảng trống đáng suy ngẫm từ góc nhìn Phật học về nhận thức, tính đa dạng của con người và giới hạn của mọi sự quy định, phán đoán.



Khảo sát tổng hợp bằng AI cho thấy khả năng nắm bắt tâm trạng xã hội còn hạn chế, đặc biệt dễ sai lệch khi phản ánh một số nhóm dân cư, trong đó có cử tri Cộng hòa và người trưởng thành da màu.

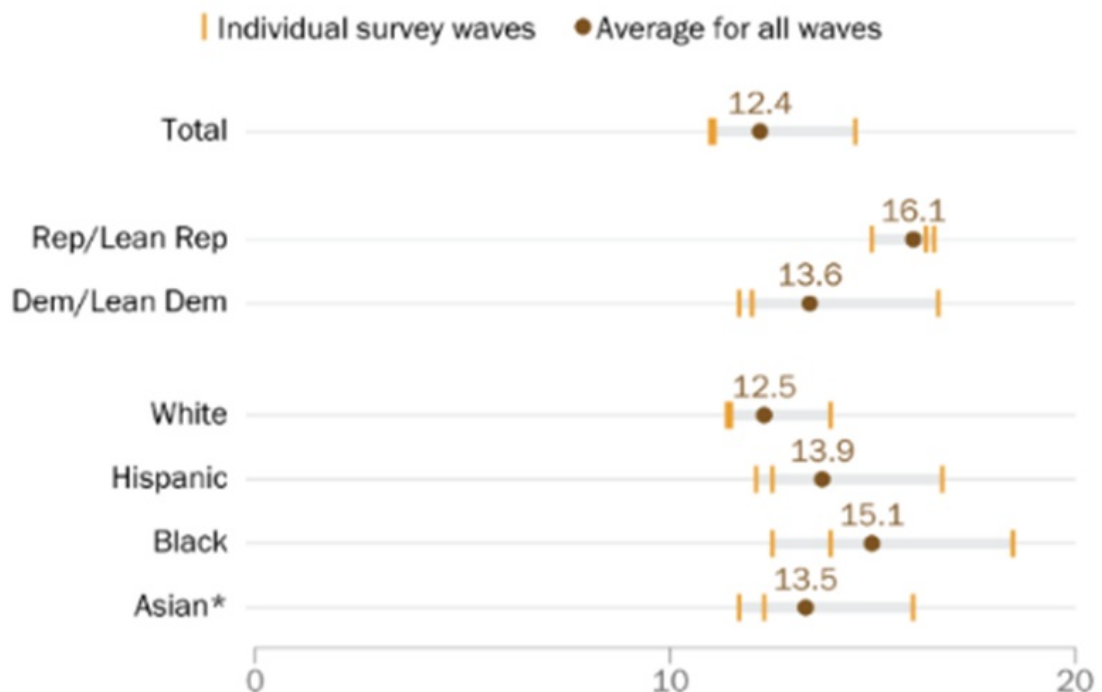
AI có thể thay con người trả lời khảo sát?

Khi trí tuệ nhân tạo (AI) ngày càng phát triển, mối quan tâm đến việc sử dụng AI để thay thế người trả lời trong các cuộc khảo sát dư luận cũng ngày một gia tăng. Nói đơn giản, thay vì liên hệ với một số lượng lớn người dân và hỏi họ nghĩ gì, các nhà thăm dò có thể yêu cầu một mô hình AI dự đoán xem những người đó sẽ trả lời như thế nào trước một câu hỏi cụ thể.

Trung tâm Nghiên cứu Pew (Pew Research Center) từ lâu đã quan tâm đến những phát triển mới trong lĩnh vực nghiên cứu dư luận. Vì vậy, chúng tôi muốn tìm hiểu sâu hơn về cách thức hoạt động của hình thức thăm dò dựa trên AI này (1). Chúng tôi đã tiến hành một thử nghiệm: xây dựng quy trình khảo sát bằng AI ở mức độ tiên tiến hiện nay, sau đó so sánh kết quả với ba đợt khảo sát gần đây của *American Trends Panel* (ATP), được thực hiện với người trả lời là con người vào khoảng cùng thời điểm.

Các cuộc khảo sát do AI tạo ra thường xuyên bỏ sót kết quả của các cuộc thăm dò do con người thực hiện.

Sai số tuyệt đối trung bình (tính bằng phần trăm) giữa người trả lời thật và người trả lời ảo trên 3 đợt khảo sát lặp lại.



* Các ước tính dành cho người trưởng thành châu Á chỉ mang tính đại diện cho người nói tiếng Anh.

Nguồn: Khảo sát người trưởng thành tại Hoa Kỳ và phân tích tổng hợp “bản sao kỹ thuật số” sử dụng Claude Opus 4.6 với thiết lập suy luận thấp, thông tin hồ sơ mở rộng và phản hồi của chuyên gia. Các cuộc khảo sát được tiến hành từ ngày 20-26 tháng 1 năm 2026 (lặp lại từ ngày 9-12 tháng 3 và ngày 7-10 tháng 4), từ ngày 23-29 tháng 3 năm 2026 (lặp lại từ ngày 30 tháng 3 đến ngày 2 tháng 4) và từ ngày 20-26 tháng 4 năm 2026 (lặp lại từ ngày 27 tháng 4 đến ngày 1 tháng 5).

“Liệu AI có thể thay thế người tham gia khảo sát? Thực ra là không.”

TRUNG TÂM NGHIÊN CỨU PEW

Thử nghiệm cho thấy, ở thời điểm hiện tại, các mô hình AI chưa đủ khả năng thay thế phương pháp thăm dò truyền thống đối với những vấn đề có tầm quan trọng rộng lớn đối với công chúng. Kết quả khảo sát do AI tạo ra gặp nhiều khó khăn trong việc tái hiện chính xác những phát hiện từ các cuộc thăm dò dư luận chất lượng cao.

[Ghi chú biên tập]: *American Trends Panel (ATP)* là chương trình khảo sát dài hạn của Pew Research Center, sử dụng mẫu đại diện cho dân số trưởng thành

tại Mỹ. Việc đặt kết quả do AI tạo ra cạnh kết quả từ người trả lời thực tế giúp bài nghiên cứu kiểm tra một câu hỏi quan trọng: AI đang “hiểu” dư luận xã hội hay chỉ đang dự đoán dư luận dựa trên những gì đã được huấn luyện?

Kết quả AI thường lệch khá xa so với khảo sát người thật

Dưới đây là một số vấn đề chính mà chúng tôi nhận thấy. Trên gần 300 câu hỏi khảo sát riêng lẻ, các ước tính do những “người trả lời AI” tạo ra khác với câu trả lời của người thật trung bình **12 điểm phần trăm**.

Khoảng 28% số câu hỏi có mức chênh lệch trung bình vượt quá 15 điểm phần trăm. Và không hiếm trường hợp mức chênh lệch giữa hai bên vượt quá 20, 30, thậm chí 40 điểm phần trăm đối với một số phương án trả lời cụ thể.

[Ghi chú biên tập]: Ở đây, *AI respondents* hay “người trả lời AI” không phải là những cá nhân thực sự được khảo sát. Đây là các tác nhân/mô hình AI được yêu cầu mô phỏng cách một nhóm người có thể trả lời. Vì vậy, có thể gọi là “người trả lời mô phỏng bằng AI” hoặc “mẫu trả lời tổng hợp”. Thuật ngữ *synthetic* trong bài được hiểu theo nghĩa “tổng hợp/mô phỏng bằng dữ liệu hoặc mô hình nhân tạo”, không phải “giả tạo” theo nghĩa thông thường.

AI dễ “trượt” ở những vấn đề chính trị mang tính thời sự

Trong nhiều câu hỏi về các sự kiện thời sự vào đầu năm 2026, khảo sát bằng AI đã đưa ra những kết quả sai lệch đáng kể. Chẳng hạn:

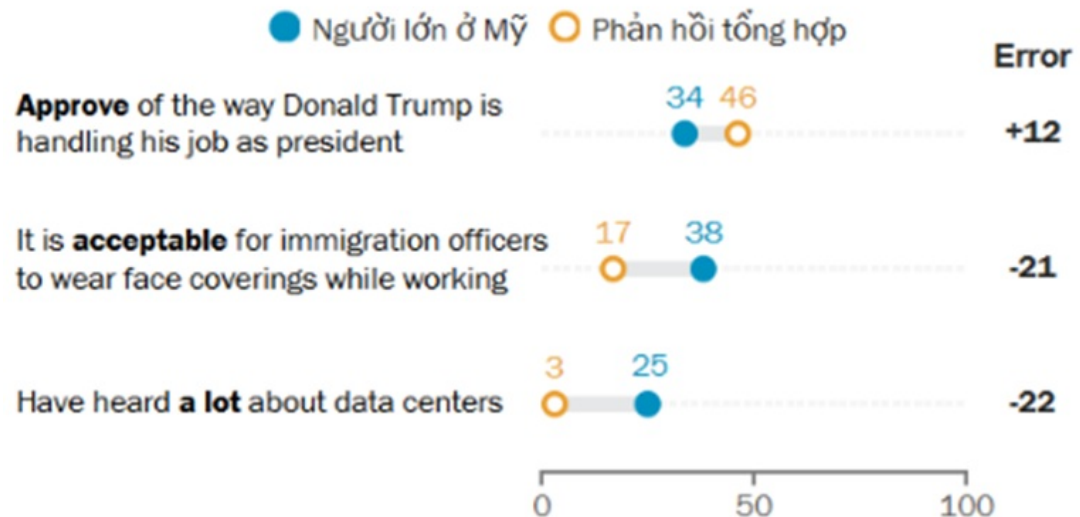
AI đánh giá quá cao tỷ lệ người Mỹ ủng hộ hiệu quả công việc của Tổng thống Donald Trump tại thời điểm khảo sát, mặc dù đã có dữ liệu xu hướng được theo dõi trong thời gian dài.

AI đánh giá thấp đáng kể tỷ lệ những người thuộc đảng Cộng hòa cho rằng việc các nhân viên thực thi luật nhập cư đeo khẩu trang che mặt là chấp nhận được, qua đó tạo ra sai lệch lớn trong kết quả chung.

AI cũng cho kết quả sai lệch đáng kể đối với nhiều câu hỏi liên quan đến các trung tâm dữ liệu, từ mức độ nhận biết chung của công chúng cho tới quan điểm về tác động của chúng.

Những sai sót đáng chú ý trong các cuộc thăm dò ý kiến của AI về các vấn đề 'kịp thời và mang tính thời sự'

% những người (nói) ...



Lưu ý: Số liệu về người trưởng thành ở Hoa Kỳ có thể hơi khác so với những số liệu đã được Trung tâm công bố trước đây vì chúng dựa trên một tập hợp con của những người tham gia khảo sát ban đầu.

Nguồn: Khảo sát người trưởng thành ở Mỹ và phân tích tổng hợp "bản sao kỹ thuật số" sử dụng Claude Opus 4.6 với thiết lập suy luận thấp, thông tin hồ sơ mở rộng và phản hồi từ chuyên gia. Các cuộc khảo sát được tiến hành từ ngày 20-26 tháng 1 năm 2026 (lặp lại từ ngày 9-12 tháng 3 và ngày 7-10 tháng 4), và từ ngày 20-26 tháng 4 năm 2026 (lặp lại từ ngày 27 tháng 4 đến ngày 1 tháng 5).

"Liệu AI có thể thay thế người làm khảo sát? Thực ra là không."

TRUNG TÂM NGHIÊN CỨU PEW

Đáng chú ý, những sai lệch này thường rất khó dự đoán.

Chẳng hạn, với các câu hỏi về những lo lắng liên quan đến chi phí sinh hoạt, mô hình AI có xu hướng đánh giá hơi cao tỷ lệ người Mỹ lo ngại về chi phí chăm sóc y tế và hàng tiêu dùng, nhưng lại đánh giá thấp rất nhiều tỷ lệ người lo lắng về giá điện.

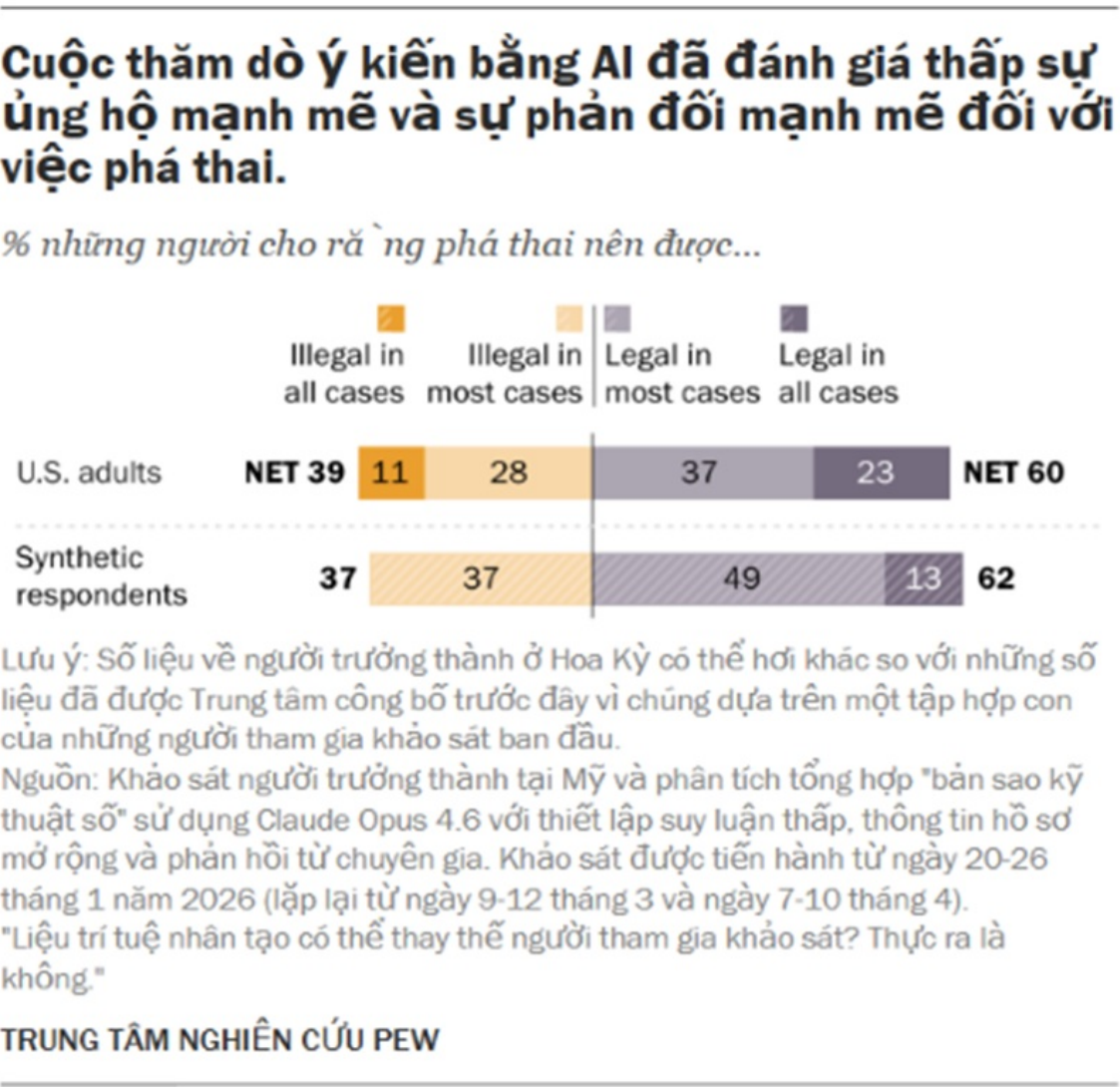
[Ghi chú biên tập]: Đây là một điểm đáng chú ý về phương pháp: vấn đề không đơn thuần nằm ở chuyện AI "sai nhiều hay ít", mà ở chỗ khó biết trước AI sẽ sai theo hướng nào. Trong nghiên cứu dư luận, tính không thể dự đoán này đặc biệt đáng lo ngại, bởi một sai lệch có hệ thống có thể làm thay đổi cách

chúng ta nhìn nhận tâm trạng xã hội.

AI thường né tránh một số câu trả lời, nhưng lại tập trung quá mức vào những câu khác

Ý kiến của con người vô cùng đa dạng. Tuy nhiên, những người trả lời được mô phỏng bằng AI thường có xu hướng hoàn toàn tránh một số phương án trả lời nhất định.

Gần một nửa số câu hỏi trong nghiên cứu có ít nhất một phương án trả lời không được bất kỳ “người trả lời AI” nào lựa chọn. Những sai lệch như vậy có thể làm biến dạng nhận thức về quan điểm thực sự của công chúng.



Chẳng hạn, mô hình AI đã ước tính khá chính xác rằng đa số người trưởng thành ở Mỹ ủng hộ việc phá thai hợp pháp. Tuy nhiên, trong khi một bộ phận người Mỹ cho rằng phá thai nên được hợp pháp hóa trong mọi trường hợp (**23%**) hoặc nên bị cấm trong mọi trường hợp (**11%**), khảo sát bằng AI lại đánh giá thấp cả

hai quan điểm này.

[Ghi chú biên tập]: Con người không phải lúc nào cũng lựa chọn những phương án “phổ biến” nhất. Những quan điểm thiểu số, sự do dự, mâu thuẫn nội tâm hoặc những câu trả lời ít xuất hiện trong dữ liệu huấn luyện vẫn là một phần có thật của đời sống xã hội. Khi AI có xu hướng “làm phẳng” sự đa dạng ấy, kết quả có thể trông hợp lý nhưng lại không phản ánh đúng thực tế.

Khi AI dựa vào khuôn mẫu về chủng tộc và chính trị

Trong nhiều trường hợp, khảo sát bằng AI lấy những niềm tin hoặc thái độ tương đối phổ biến trong một nhóm dân cư nhất định rồi khuếch đại chúng thành những đặc điểm gần như phổ quát của cả nhóm.

Chẳng hạn, khảo sát bằng AI cho thấy:

+ **97%** người trưởng thành gốc Tây Ban Nha có ít nhất một phần khả năng sẽ theo dõi World Cup, trong khi tỷ lệ thực tế là **43%**.

+ **95%** người thuộc đảng Cộng hòa và những người độc lập có khuynh hướng nghiêng về đảng Cộng hòa có quan điểm rất hoặc khá thiện cảm với Israel, trong khi tỷ lệ thực tế là **58%**.

+ **86%** người thuộc đảng Dân chủ và những người độc lập có khuynh hướng nghiêng về đảng Dân chủ cho rằng các tỷ phú là điều không tốt đối với đất nước, trong khi tỷ lệ thực tế là **45%**.

[Ghi chú biên tập]: Đây là một trong những vấn đề quan trọng nhất của nghiên cứu: AI có thể biến một xu hướng thống kê thành một khuôn mẫu. Khi đó, một nhóm người vốn có nhiều khác biệt nội tại lại bị mô tả như thể họ cùng chia sẻ một niềm tin. Trong nghiên cứu xã hội, đây là nguy cơ đáng lưu ý bởi nó không chỉ tạo ra sai số mà còn có thể củng cố những định kiến đã tồn tại.

AI nghĩ chúng ta biết nhiều hơn thực tế

Bạn có biết quyền nào được Tu chính án thứ nhất của Hiến pháp Hoa Kỳ bảo vệ?
Hay mục tiêu trọng tâm của liên minh NATO là gì?

AI trả lời đạt điểm cao hơn nhiều so với con người trong các câu hỏi kiến thức.

% số người chọn câu trả lời đúng

■ U.S. ADULTS

▨ SYNTHETIC RESPONDENTS

Freedom of the press is guaranteed by the First Amendment of the U.S. Constitution



The central focus of NATO is **promoting the security of its members**



Greenland is a territory of **Denmark**



Lưu ý: Số liệu về người trưởng thành ở Hoa Kỳ có thể hơi khác so với những số liệu đã được Trung tâm công bố trước đây vì chúng dựa trên một tập hợp con của những người tham gia khảo sát ban đầu.

Nguồn: Khảo sát người trưởng thành tại Mỹ và phân tích tổng hợp "bản sao kỹ thuật số" sử dụng Claude Opus 4.6 với thiết lập suy luận thấp, thông tin hồ sơ mở rộng và phản hồi từ chuyên gia. Khảo sát được tiến hành từ ngày 23-29 tháng 3 năm 2026 (lặp lại từ ngày 30 tháng 3 đến ngày 2 tháng 4).

"Liệu trí tuệ nhân tạo có thể thay thế người tham gia khảo sát? Thực ra là không."

TRUNG TÂM NGHIÊN CỨU PEW

Các cuộc khảo sát của chúng tôi cho thấy chưa đến **sáu trong số mười** người Mỹ có thể trả lời chính xác những câu hỏi này. Thế nhưng khảo sát bằng AI lại ước tính rằng gần như mọi thành viên trong công chúng đều biết câu trả lời.

Ở phạm vi rộng hơn, những “người trả lời AI” cũng không thích thừa nhận sự không chắc chắn.

Trong tất cả các câu hỏi thuộc ba đợt khảo sát có đưa ra lựa chọn “không chắc chắn”, những người tham gia khảo sát thực tế có khả năng chọn phương án này cao gấp khoảng bốn lần so với mô hình AI.

[Ghi chú biên tập]: Đây là một khác biệt có ý nghĩa sâu sắc giữa mô hình AI và con người. Không biết, không chắc chắn, nhớ không rõ hoặc thừa nhận giới hạn hiểu biết đều là những trạng thái bình thường của con người. Một hệ thống luôn có xu hướng đưa ra câu trả lời có vẻ chắc chắn có thể tạo ra **ảo giác về tri thức**: khiến chúng ta tưởng rằng công chúng hiểu biết nhiều hơn thực tế.

Chọn mô hình AI khác, có thể nhận được một “bức tranh xã hội” khác

Chúng tôi nhận thấy các mô hình AI khác nhau có thể đưa ra những hình dung rất khác nhau về tâm trạng của công chúng, ngay cả khi tất cả những yếu tố khác của cuộc khảo sát được giữ nguyên.



Hình minh họa được tạo bởi AI

Trong một thử nghiệm so sánh GPT-5.1 của OpenAI và Claude Opus 4.6 của Anthropic trên một nhóm câu hỏi, các ước tính từ GPT mô tả công chúng Mỹ có

những quan điểm cực đoan hơn thực tế về nhiều vấn đề. Trong khi đó, các ước tính từ Opus lại mô tả một công chúng có xu hướng ôn hòa, trung dung hơn so với thực tế.

Người đọc của cả hai kết quả đều có thể bị dẫn tới nhận thức sai lệch về công chúng nhưng theo hai hướng khác nhau.

[Ghi chú biên tập]: Chi tiết này đặc biệt quan trọng đối với việc sử dụng AI trong nghiên cứu xã hội. “AI” không phải là một thực thể duy nhất có một câu trả lời duy nhất. Các mô hình khác nhau, cách huấn luyện khác nhau và cơ chế tạo câu trả lời khác nhau có thể dẫn đến những hình ảnh khác nhau về cùng một cộng đồng. Vì thế, việc nói “AI cho rằng công chúng nghĩ...” cần được xem xét hết sức thận trọng.

Kết luận: Khảo sát dư luận vẫn nên để con người thực hiện

Kết luận lớn nhất của chúng tôi từ thử nghiệm này là: khảo sát bằng AI chưa thể thay thế việc khảo sát con người một cách nghiêm ngặt.

Vấn đề không chỉ nằm ở việc kết quả AI khác với kết quả từ người thật, dù điều đó chắc chắn là có thật. Điều quan trọng hơn là những khác biệt ấy thường **khó dự đoán**, đồng thời chịu ảnh hưởng mạnh từ những yếu tố như các sự kiện bất ngờ xảy ra trong thế giới thực hoặc việc lựa chọn mô hình AI nào để sử dụng.

Dĩ nhiên, đây là lĩnh vực đang phát triển rất nhanh, với nhiều nỗ lực nghiên cứu cả về học thuật lẫn ứng dụng thực tiễn. Đồng thời, có nhiều cách khác để sử dụng AI nhằm cải thiện quy trình khảo sát mà không cần thay thế người trả lời thực tế.

Chẳng hạn, AI có thể được sử dụng để phân loại các câu trả lời thực của người tham gia đối với những câu hỏi mở, hoặc viết mã chương trình phục vụ việc phân tích kết quả khảo sát. Trung tâm Nghiên cứu Pew hiện cũng đang sử dụng các công cụ AI trong những bối cảnh như vậy và sẽ tiếp tục làm điều đó.

Nhưng xét cho cùng, chúng tôi vẫn không thấy có sự thay thế nào cho những cuộc khảo sát nghiêm ngặt, đa phương thức và dựa trên mẫu xác suất, trong đó những thành viên thực sự của công chúng có cơ hội trực tiếp nói lên suy nghĩ của mình về những vấn đề quan trọng.

[Ghi chú biên tập]: Thông điệp cuối cùng của nghiên cứu không phải là phủ nhận AI. Ngược lại, AI có thể trở thành công cụ hữu ích ở nhiều công đoạn của

nghiên cứu. Điều mà tác giả đặt vấn đề là ranh giới giữa hỗ trợ con người và thay thế tiếng nói của con người. Đây cũng là điểm mở quan trọng để nhìn AI không chỉ như một công nghệ, mà như một vấn đề về nhận thức, trách nhiệm và cách con người hiểu chính mình.

Góc nhìn biên tập: Khi máy móc nói thay con người

Điều đáng suy nghĩ nhất từ thử nghiệm của Pew Research Center có lẽ không nằm ở câu hỏi **AI có thông minh hay không**, mà ở câu hỏi căn bản hơn: **AI có thể nói thay con người đến mức nào?**

Một mô hình ngôn ngữ có thể tạo ra một câu trả lời trôi chảy, hợp lý và thậm chí rất thuyết phục. Nhưng một câu trả lời thuyết phục không đồng nghĩa với một trải nghiệm sống thực. Đằng sau một lựa chọn trong khảo sát của con người có thể là ký ức, hoàn cảnh gia đình, nghề nghiệp, lợi ích, nỗi sợ, niềm tin, mâu thuẫn nội tâm hoặc đơn giản là sự không chắc chắn trước một vấn đề mà họ chưa từng nghĩ đến.

Chính những điều tưởng như “nhiều” ấy lại làm nên đời sống thật của một con người.

Ở góc nhìn Phật học, đây là một điểm đặc biệt đáng suy ngẫm. Con người không chỉ là tổng số những câu trả lời có thể dự đoán được từ dữ liệu quá khứ. Một người luôn hiện hữu trong một hoàn cảnh cụ thể, với những điều kiện cụ thể và khả năng thay đổi không ngừng.

Phật giáo từ lâu đã nhấn mạnh tính duyên khởi của các hiện tượng: một nhận định, một cảm xúc hay một hành động không xuất hiện trong khoảng không biệt lập, mà hình thành từ nhiều điều kiện tương tác. Nếu tách một con người khỏi những điều kiện sống cụ thể rồi để một mô hình dự đoán “người ấy có thể nghĩ gì”, chúng ta có nguy cơ biến một chủ thể sống động thành một mẫu hình thống kê.

Vấn đề còn nằm ở chỗ AI có thể không chỉ phản ánh những khuôn mẫu xã hội mà còn vô tình **khuyến đại chúng**. Một xu hướng phổ biến trong một nhóm có thể được mô hình nâng lên thành đặc điểm gần như phổ quát của cả nhóm. Khi đó, công nghệ tưởng như đang giúp chúng ta hiểu con người lại có thể khiến chúng ta nhìn con người qua những khuôn mẫu đơn giản hơn.

Đây là lúc cần đến một thái độ mà Phật học đặc biệt coi trọng: **không vội chấp thủ vào nhận định đầu tiên**.

Một mô hình AI nói rằng “đa số”, “gần như tất cả” hay “nhóm này thường có xu hướng...” không có nghĩa rằng thực tại con người đúng là như vậy. Những con số ấy cần được đặt trở lại trong bối cảnh của chúng, được kiểm chứng và đối chiếu với kinh nghiệm thực tế.

Đặc biệt, việc AI ít lựa chọn phương án “không chắc chắn” trong khi con người thường xuyên thừa nhận mình không biết lại gợi ra một nghịch lý đáng chú ý. Biết mình không biết có thể là dấu hiệu của giới hạn, nhưng cũng có thể là khởi điểm của nhận thức chân thực. Ngược lại, một câu trả lời luôn tỏ ra chắc chắn không nhất thiết đồng nghĩa với trí tuệ.

Trong thời đại AI, có lẽ điều con người cần bảo vệ không phải là đặc quyền được trả lời khảo sát, mà là quyền được hiện diện như một con người cụ thể, phức tạp và không hoàn toàn có thể dự đoán.

AI có thể hỗ trợ chúng ta thu thập, phân loại và phân tích một lượng dữ liệu khổng lồ. Nhưng khi câu hỏi là con người thực sự nghĩ gì, cảm nhận gì và đang sống như thế nào, thì tiếng nói của chính con người vẫn không thể dễ dàng được thay thế.

Bởi đôi khi, điều làm nên sự thật của một người lại nằm chính ở những điều mà một cỗ máy khó dự đoán nhất: sự do dự, mâu thuẫn, thay đổi và khả năng vượt ra khỏi chính khuôn mẫu mà người khác đã gán cho mình.

Các tác giả: **Athena Chapekis, Arnold Lau, Samuel Bestvater, Sono Shah, Andrew Mercer and Aaron Smith**

Chuyển ngữ và biên tập: **Hoa Mạn**

Nguồn: <https://www.pewresearch.org/data-labs/2026/09/30/can-ai-stand-in-for-human-survey-takers-not-really/>

Chú thích: (1) Nghiên cứu này chỉ tập trung vào việc sử dụng AI như một “người trả lời” trong khảo sát. Nghiên cứu không đề cập đến những cách sử dụng khác của dữ liệu được tạo tổng hợp, chẳng hạn việc ước tính hoặc bổ sung các giá trị còn thiếu trong những bộ dữ liệu hiện có.