

Khi nhân loại vô tri, AI có giúp được gì không?

ISSN: 2734-9195

22:25 06/06/2023

Sống trong thời đại tiến bộ công nghệ tăng tốc thật thú vị. Theo nhiều cách, chúng ta thật may mắn và phần lớn công nghệ mới này có thể và sẽ làm xoa dịu những nỗi khổ niềm đau của con người khi được sử dụng một cách khôn ngoan.

Tác giả: **Justin Whitaker** Việt dịch: **Thích Vân Phong** Nguồn: *Buddhistdoor Global*

[caption id="attachment_16547" align="aligncenter" width="600"]

Hình ảnh được Tạo Bởi Dall E 2 Với Lời Nhắc Của Tác Giả Hình ảnh được tạo bởi DALL-E 2 với lời nhắc của tác giả[/caption]

“Tôi nghĩ thực sự quan trọng bởi chúng tôi phải giải thích, hướng dẫn mọi người rằng đây là một công cụ chứ không phải một sinh vật”. (doanh nhân, nhà đầu tư công nghệ, lập trình viên người Mỹ Sam Altman, Giám đốc điều hành OpenAI - công ty khởi nghiệp phát triển ChatGPT)

Trong buổi thảo luận gần đây, Lex Fridman, nhà nghiên cứu AI người Mỹ làm việc trên các phương tiện tự hành, tương tác giữa con người và robot và học máy tại MIT, Giám đốc điều hành OpenAI Sam Altman đã đề cập đến một số chủ đề, bao gồm cả quan điểm của ông về tiềm năng của các hệ thống trí tuệ nhân tạo (AI) hiện tại và tương lai.

Đôi khi Sam Altman, Giám đốc điều hành OpenAI bày tỏ mối quan ngại sâu sắc về khả năng trí tuệ nhân tạo (AI) có thể gây hại cho thế giới, nhưng trong cuộc phỏng vấn này, ông đã nói rằng những gì AI làm, cuối cùng vẫn phụ thuộc vào con người chọn sử dụng nó cho mục đích gì, chứ không phải nó có thể tư duy cho mục đích gì, theo cách mà con người và động vật vẫn làm, đồng thời ông nói thêm: “Tôi nghĩ việc thể hiện ‘sự sáng tạo’ lên một công cụ là rất nguy hiểm.” (YouTube)

Khi chúng ta tiếp tục theo dõi sự phát triển của trí tuệ nhân tạo (AI), ngày càng cho thấy, đây chỉ là một mối nguy hiểm trong số nhiều mối nguy hiểm khác. Là một Phật tử quan tâm sâu sắc đến việc hiểu bản chất của sự vật và là một người quan tâm đến cuộc sống hữu tình đối với sự đau khổ, tôi đã tìm cách hiểu được tiềm năng tốt và xấu của trí tuệ nhân tạo (AI).

“Sinh vật” mà Sam Altman, Giám đốc điều hành OpenAI đề cập đến có thể là một cách khác để xác định tri giác hoặc ý thức. Đây là những thuật ngữ riêng biệt, nhưng có sự liên quan chặt chẽ trong các buổi thảo luận triết học hiện đại. Trên thực tế, Robert Van Gulik (1910-1967), Đông phương học uyên thâm, một viên chức ngoại giao viết trong bách khoa toàn thư học thuật trực tuyến về triết học (Stanford Encyclopedia of Philosophy, IEP) đưa ra tri giác như một trong những định nghĩa về ý thức. Các định nghĩa khác bao gồm sự tỉnh giác, xác định rằng một sinh vật có ý thức chỉ thực sự có ý thức khi tỉnh giác và sử dụng nhận thức đó, và nó như thế nào, theo bài báo nổi tiếng năm 1974 của Tiến sĩ Thomas Nagel, Triết gia Luật và một giáo sư tập trung vào triết học tâm trí, triết học chính trị và đạo đức, trong đó chúng ta đối mặt với những hạn chế hiểu biết và trí tưởng tượng của chính mình khi cố gắng tưởng tượng nó như thế nào để trở thành một con dơi.

Trong tư tưởng Phật giáo, ý thức thường được tìm thấy trong nhóm ngũ uẩn (Skt. skandhas). Chúng được mô tả bởi nhà nghiên cứu về học triết Hy Lạp cổ đại và Ấn Độ cổ điển, với chủ đề tập trung vào siêu hình học, nhận thức luận và tâm lý đạo đức làm nền tảng cho đạo đức của Plato và đạo đức Phật giáo Ấn Độ, cô Amber Carpenter viết rằng: “‘Sắc’, Sanskrit là rūpa tức khối vật chất được kết cấu, tạo hình theo nguyên lý không bền vững cho nên gọi là sắc uẩn; ‘thụ’ Sanskrit là vedanā tức có tự thể là cảm nghiệm, thụ là sự cảm nhận của thức. Thụ có ba loại: lạc thọ, khổ thọ và thọ không lạc không khổ; ‘tưởng’ Sanskrit là Samjñā tức tác dụng tâm lý nắm bắt thật nhiều các tướng trạng cảnh giới, nó làm tăng thêm sức mạnh nắm bắt; ‘hành’ Sanskrit gọi là saṃskāra tức hành động của thân và ngữ hoàn toàn được chỉ đạo hay thúc đẩy bởi Tư hay Ý chí (cetanā); ‘thức’ Sanskrit là Vijñāpti, đi từ động từ căn vi-vjñā có nghĩa là thông tin, cái thông tin, cũng hiểu là công cụ thông tin, hoặc cái dùng để nhận thức, một kinh nghiệm có thể đúng hoặc sai, chỉ cho thức thứ tám.” (Triết Học Phật Giáo Ấn Độ 29)

Tốt nhất nên coi đây là những cách trải nghiệm hơn là nền tảng bản thể học.

Ý thức có đặc biệt không?

Câu hỏi làm như thế nào để xác định là ai đó hay cái gì có ý thức đã đeo bám các triết gia phương Tây trong một thời gian. Triết gia René Descartes

(1596–1650), nhà khoa học, nhà toán học người Pháp, thường được coi là nhà tư tưởng đầu tiên áp dụng phương pháp triết học chặt chẽ cho vấn đề này.

Là tín hữu Công giáo Rôma, René Descartes xác định rằng động vật không phải con người không đạt được điểm đánh dấu này. Từ lâu Cơ đốc giáo đã đưa ra một ranh giới rõ ràng giữa con người có linh hồn và động vật không có linh hồn, và Triết gia René Descartes đã không phá vỡ khía cạnh này của truyền thống. Tuy nhiên, những người yêu mến đạo Phật nói chung luôn luôn gắn ý thức cho động vật. Nó cũng có khả năng – theo những cách hạn chế hơn – suy nghĩ và thực hiện các hành động có chủ ý (nghịệp) sẽ ảnh hưởng đến cuộc sống và những lần tái sinh trong tương lai của nó. Với sự phát triển của khoa học và chủ nghĩa duy vật trong triết học, lại một lần nữa ý thức bị đặt vấn đề về cách định nghĩa.

Cuối cùng, nếu chúng ta chấp nhận tiền đề rằng, chúng ta hoàn toàn có bản chất vật chất, thì làm thế nào ý thức phát sinh? Tại sao ý thức phát sinh trong chúng ta mà không phải trong đất đá, cây cỏ? Hầu hết các câu trả lời đều xoay quanh sự phức tạp tuyệt đối của bộ não và cơ thể chúng ta, giải thích rằng từ sự phức tạp này, ý thức đã “phát sinh”. Tuy nhiên, có nhiều lý thuyết cạnh tranh của điều này về cách thức hoạt động, cũng như những người nghèo khó theo chủ nghĩa duy vật cho rằng, ý thức có thể nằm ngoài khả năng hiểu biết toàn diện của chúng ta.

Thúc đẩy sự phức tạp đi xa hơn một chút là việc nhận ra rằng, con người thường đưa trí thông minh hoặc một trong những ngành nghề đang rất phổ biến hiện nay (agency) vào một thế giới mà nó có thể không tồn tại.

Ví dụ, chúng ta phát triển cảm xúc đối với các nhân vật trong sách khi biết rằng nó là hư cấu. Khi nhân vật Tom Hanks, diễn viên, đạo diễn, nhà sản xuất phim người Mỹ nổi tiếng, diễn viên Cast Away nảy sinh tình cảm với một quả bóng chuyền, chúng tôi đồng cảm, biết rằng chúng tôi cũng có thể làm như thế.

Trong một nghiên cứu từ những thập niên 1940, các nhà nghiên cứu đã cho thấy hình ảnh động của hai hình tròn và một hình tam giác di chuyển xung quanh hình dạng của một ngôi nhà. Khi mô tả những gì họ thấy, gần như tất cả những người tham gia đã tạo nên “một âm mưu xã hội trong đó tam giác lớn bị coi là kẻ xâm lược. Các nghiên cứu đã chỉ ra rằng chuyển động của các hình dạng gây ra nhận thức vật linh tự động.” (Carnegie Mellon)

Trí tuệ nhân tạo (AI) là gì?

Định nghĩa AI cũng gây tranh cãi như định nghĩa ý thức. Một phương diện, AI có thể là bất cứ thứ gì có thể giải quyết một vấn đề phức tạp. Các đề xuất tự động

mà bạn nhận được khi nhập cụm từ vào Google được tạo bởi một dạng AI. Các hình thức phức tạp hơn của AI chơi cờ vua, hướng dẫn lộ trình giao thông khi lái xe trên điện thoại của bạn và cung cấp quảng cáo phù hợp với sở thích và hành vi của bạn.

Tất cả những điều này dựa trên các thuật toán được lập trình ngày càng phức tạp. Các dạng AI mới nhất – những dạng đang tạo ra sự phấn khích nhất – sử dụng cái được gọi là “mạng lưới thần kinh” (một loại máy học được mô hình hóa sau bộ não con người). Các nhà triết học chỉ ra rằng bản thân cái tên này có thể gây hiểu lầm, vì nó chỉ bắt chước các khía cạnh đơn giản hóa của tế bào thần kinh sinh học. Tuy nhiên, các mạng này có khả năng thay đổi phản ứng của nó theo thời gian, bắt chước quá trình học hỏi ở con người.

Một lần nữa, chúng tôi có thể cẩn thận về các thuật ngữ mà chúng tôi sử dụng, vì “học tập” là điều mà chúng tôi có thể nói đòi hỏi phải có ý thức. Một con người học tập. Một con chó học tập. Thậm chí con cá vàng còn học tập được. Nhưng quả bóng trong máy bắn bi không học cách đánh đúng điểm – mọi thứ nó làm đều phụ thuộc vào đầu vào trước đó của con người, và các tương tác của nó với máy móc mà chúng nó hoạt động.

Các nhà phê bình, chẳng hạn như nhà khoa học máy tính người Ethiopia, đồng trưởng nhóm kỹ thuật của Nhóm Trí tuệ nhân tạo đạo đức tại Google, chị Timnit Gebru chỉ ra rằng đây chính là nơi nảy sinh các vấn đề trong các mô hình xác suất có khả năng hiểu và sinh ngôn ngữ tự nhiên dựa trên kiến thức được thu thập từ các tập dữ liệu cực lớn (Large language models, LLM) – các ứng dụng trí tuệ nhân tạo (AI) như ChatGPT. Chị Timnit Gebru lưu ý rằng dữ liệu đào tạo được cung cấp cho các chương trình AI lớn này bị sai lệch và đôi khi mang tính đối địch nhau, dẫn đến lo ngại rằng phần mềm sẽ sao chép những thành kiến và lời nói căm thù này trong kết quả của nó.

Năm 2016, một Twitter-bot do Microsoft phát triển đã nhanh chóng bắt đầu gửi các tweet phân biệt chủng tộc sau khi nó được thành lập để học tập từ những người dùng nền tảng này. Nó đã bị các nhà phát triển gỡ xuống trong vòng chưa đầy 24 giờ sau khi ra mắt công chúng (The New York Times).

Giải phóng hay Phòng thủ?

Điều này đặt ra vấn đề thứ hai, có liên quan, những người ủng hộ các hình thức AI mới nhất tuyên bố rằng, chúng nó sẽ có sức mạnh kỳ diệu đến kinh ngạc. Chắc chắn chúng nó có khả năng tuyệt vời. Nhưng, như trùm doanh nhân, nhà đầu tư, nhà từ thiện người Mỹ, tỷ phú Warren Buffet đã nói trong một cuộc họp gần đây: “Với trí tuệ nhân tạo (AI) . . . nó có thể thay đổi mọi thứ trên thế giới,

ngoại trừ cách đàn ông suy nghĩ và hành động. Đây là một bước tiến lớn cần thực hiện.” (EFT Central)

Tác giả, nhà hoạt động xã hội và nhà làm phim người Canada nổi tiếng, Naomi Klein nói thêm:

“Có một thế giới trong đó có trí tuệ nhân tạo (AI), với tư cách là một công cụ nghiên cứu dự đoán mạnh mẽ và là người thực hiện các biểu diễn tẻ nhạt, thực sự có thể được sắp xếp để mang lại lợi ích cho nhân loại, các loài khác và ngôi nhà chung của chúng ta. Nhưng để điều này xuất hiện, những công nghệ này cần phải được triển khai bên trong một trật tự kinh tế và xã hội khác rất nhiều so với trật tự xã hội chúng ta, một trật tự có mục đích là đáp ứng nhu cầu của con người và bảo vệ các hệ thống hành tinh hỗ trợ mọi sự sống.” (The Guardian)

Tác giả Naomi Klein lưu ý rằng đây không phải là cách trí tuệ nhân tạo (AI) được khởi chạy. Thay vào đó, các tập đoàn lớn vì lợi nhuận được phép sao chép một lượng lớn văn bản và hình ảnh do con người tạo ra – mà không cần sự cho phép hoặc ghi công – để tạo ra các kết quả bắt chước của chính nó.

Thiên hạ lập luận rằng nhiều lời hứa vĩ đại chỉ đơn giản là thúc đẩy sự cường điệu. Sự cường điệu này lại làm tăng giá trị của các công ty sản xuất trí tuệ nhân tạo (AI). Ngay cả sự cường điệu tiêu cực, theo câu châm ngôn xưa, “Không có thứ gọi là báo chí xấu,” có thể có tác dụng thu hút sự chú ý nhiều hơn đến các tập đoàn ấy.

Như chuyên gia AI và giáo sư danh dự về tâm lý học và khoa học thần kinh tại Đại học New York (Mỹ), nhà khoa học nhận thức người Mỹ Gary Marcus đã lưu ý vào năm 2022, chúng ta không còn nhận được phản hồi từ các nhà nghiên cứu hàn lâm Khoa học và Công nghệ về trí tuệ nhân tạo (AI) nữa. Chúng ta ngày càng nghe nhiều hơn từ các CEO của công ty: “Và các tập đoàn, không giống như các trường đại học, không có động cơ để đọc sức công bằng. Thay vì gửi những bài báo mới gây chú ý của họ cho sự giám sát học thuật, họ đã xuất bản bằng thông cáo báo chí, dụ dỗ các nhà báo và bỏ qua quá trình đánh giá ngang hàng (một hình thức kiểm tra, rà soát lại hoặc đánh giá công việc của những người có trình độ tương đương nhau trong cùng một lĩnh vực liên quan đến công việc nào đó). Chúng tôi chỉ biết những gì các công ty muốn chúng tôi biết.” (Scientific American)

Điều này nguy hiểm. Nhưng trong những năm gần đây nó cũng đi theo con đường ngày càng trở nên phổ biến với công nghệ mới. Đôi khi điều này dẫn đến gian lận hoàn toàn, chẳng hạn như các vụ bê bối xung quanh Tập đoàn Công

nghe y tế Theranos của cựu nữ doanh nhân Hoa Kỳ, Elizabeth Holmes, người sáng lập và cựu Tổng giám đốc, Sàn giao dịch tiền điện tử nổi tiếng nhất thế giới FTX của nhà sáng lập Sam Bankman-Fried. Đôi khi, nó chỉ đơn giản dẫn đến sự cường điệu quá mức mà không thành công, như trong trường hợp chiếc kính thực tế ảo tăng cường (AR) của Google, thực tế ảo siêu vũ trụ của Meta hoặc Mã thông báo không thay thế được (NFT) là một loại mã thông báo mật mã trên một chuỗi khối đại diện cho một tài sản duy nhất.

Một giải pháp tạm thời

Sống trong thời đại tiến bộ công nghệ tăng tốc thật thú vị. Theo nhiều cách, chúng ta thật may mắn và phần lớn công nghệ mới này có thể và sẽ làm xoa dịu những nỗi khổ niềm đau của con người khi được sử dụng một cách khôn ngoan. Trí tuệ phát sinh không phải là kết quả của việc biết sử dụng tốt nhất hoặc sử dụng công nghệ mới nhất, mà là do cân nhắc, phân tích và thực hành các truyền thống đại diện cho hàng nghìn năm và hàng triệu cuộc đời con người, mỗi người đều tinh chỉnh, thay đổi và truyền lại những thành tựu tốt nhất của mình. Là Phật tử, điều này đòi hỏi chúng ta phải đặt ra những câu hỏi nghiêm túc về những hứa hẹn và cạm bẫy tiềm ẩn của trí tuệ nhân tạo (AI) trong đời sống đạo đức, thiền định và triết học của chúng ta.

Bạn tôi Douglass Smith có một số video hấp dẫn trên kênh YouTube, anh ấy khám phá trí tuệ nhân tạo (AI) và các khía cạnh của tư tưởng và thực hành Phật pháp. Tôi khuyến khích bạn ấy kiểm tra chúng. Đây là một khám phá các giá trị chính mà chúng ta có thể muốn có AI trong tương lai và cách chúng ta có thể giúp đạt được điều này.

Mindfulness for Computers? Buddhist Practice and the AI "Alignment Problem"

<https://www.youtube.com/watch?v=RsOs6DHKhPo>

Mặc dù những người trong chúng ta có nền tảng nhân văn có thể áp dụng một số cách tiếp cận khác nhau đối với trí tuệ nhân tạo (AI) và các công nghệ mới khác, nhưng điều cần thiết là tiếng nói của chúng ta phải là một phần của cuộc đối thoại. Như Leon Wieseltier, biên tập viên của tạp chí nhân văn Liberties, viết:

“Trong lịch sử của chúng ta không có thời điểm nào mà các ngành khoa học nhân văn, triết học, đạo đức và nghệ thuật lại cần thiết cấp bách hơn thời điểm chiến thắng của công nghệ này. Bởi vì chúng ta cần có khả năng suy nghĩ theo những thuật ngữ phi công nghệ nếu chúng ta muốn tìm ra điều hay lẽ phải và điều xấu ác trong tất cả các đối mới công nghệ. Với sự tôn sùng công nghệ điên cuồng của xã hội, liệu chúng ta có tin tưởng các kỹ sư và nhà tư bản sẽ cho

chúng ta biết điều gì đúng và điều gì sai không?" (The New York Times)

Với tinh thần tư duy phi công nghệ, tôi sẽ mượn một kết luận thường được sử dụng bởi những người phụ trách chuyên mục BDG: một bài hát.

The Humans Are Dead - Full version

<https://www.youtube.com/watch?v=B1BdQcj2ZYY>

Còn nhiều điều để chia sẻ, từ cách giải thích ý thức của các trường phái Phật giáo khác nhau, hoặc về những gì có thể xảy ra tiếp theo trong quá trình phát triển của sự sáng tạo dựa trên máy móc. Chúng tôi biết rằng tương lai đang mở rộng. Nhưng chúng ta cũng biết rằng những hạn chế và thiếu sót có xu hướng phát sinh, ngay cả đối với những phát minh và đổi mới vĩ đại nhất.

[caption id="attachment_16548" align="aligncenter" width="700"]

Quảng Cáo Lò Vi Sóng Amana Từ Những Năm 1970 Quảng cáo lò vi sóng Amana từ những năm 1970. Ảnh: cambridge.org[/caption]

Tôi sẽ kết thúc bằng một video nữa, video này của Adam Conover, một sinh viên tốt nghiệp triết học, phỏng vấn Emily Bendee, một nhà ngôn ngữ học từ Đại học Washington và cô Timnit Gebru, nhà khoa học máy tính người Ethiopia, đồng trưởng nhóm kỹ thuật của Nhóm Trí tuệ nhân tạo đạo đức tại Google. Nơi cô làm việc thiên về thuật toán và khai thác dữ liệu. Cô là người ủng hộ sự đa dạng trong công nghệ và là người đồng sáng lập Black trong AI, một cộng đồng các nhà nghiên cứu da đen làm việc trong lĩnh vực trí tuệ nhân tạo.

Tác giả: **Justin Whitaker** Việt dịch: **Thích Vân Phong** Nguồn: *Buddhistdoor Global*