

Phật giáo cung cấp giải pháp phát huy tiềm năng, tránh lạm dụng AI (P.1)

ISSN: 2734-9195

15:20 23/10/2026

Điều đáng chú ý trong cuộc đối thoại này không đơn thuần là AI có thể hỗ trợ việc học Phật hay thực hành thiền định đến đâu, mà là câu hỏi về giới hạn của công nghệ trước chiều sâu trải nghiệm con người.

***Trí tuệ nhân tạo (AI)** ngày càng có khả năng cung cấp tri thức, hỗ trợ đối thoại và cá nhân hóa trải nghiệm học tập, tu tập. Nhưng liệu khả năng xử lý thông tin có đồng nghĩa với trí tuệ?*

*AI có thể đồng hành cùng con người trên **hành trình tỉnh thức** mà không khiến chúng ta lệ thuộc vào **công nghệ**, đánh mất năng lực tự quán sát và suy tư độc lập?*

*Qua cuộc đối thoại giữa Justin Kelley và Jacob Fisher, hai nhà nghiên cứu, giảng dạy tại Tergar Institute, bài viết gợi mở những tiềm năng và giới hạn của AI trong thực hành Phật giáo, đồng thời đặt ra một vấn đề cốt lõi: **đạo đức Phật giáo** có thể đóng góp gì để công nghệ phục vụ con người bằng trí tuệ, lòng từ bi và tinh thần bất hại?*



Khi công nghệ ngày càng đan xen vào mọi phương diện của **đời sống** con người, các cộng đồng tâm linh cũng bắt đầu đặt câu hỏi: Liệu trí tuệ nhân tạo (AI) có thể đóng góp tích cực cho hành trình tỉnh thức hay không, và nếu có, thì bằng cách nào?

Chúng tôi đã mời Justin Kelley, đồng Giám đốc điều hành Tergar Institute (Viện Tergar), cùng Jacob Fisher, thành viên đội ngũ giảng dạy của Viện, tham gia cuộc đối thoại về những tiềm năng và cạm bẫy của AI trong Phật giáo.

Trong phần đầu của loạt **đối thoại** này, hai diễn giả thảo luận về cách AI có thể đóng góp thiết thực cho sự phát triển tâm linh và thực hành quán chiếu, đồng thời xem xét vai trò định hướng của đạo đức Phật giáo đối với quá trình phát triển và ứng dụng công nghệ này.

Cuộc trao đổi cũng nêu lên những lo ngại về sự lệ thuộc quá mức vào AI, nguy cơ làm suy yếu các năng lực nội tại như chính niệm và tự phản tỉnh, cũng như câu hỏi liệu các thuật toán có thể dần đảm nhận những vai trò vốn thuộc về các vị thầy tâm linh hay không.

JACOB: AI có thể đóng góp một cách thực chất cho sự phát triển tâm linh hoặc thực hành quán chiếu hay không? Justin nghĩ thế nào?

JUSTIN: Đây là một chủ đề khá nhạy cảm. Có lẽ chúng ta có thể nhìn nhận AI từ một vài góc độ. Trước hết là góc độ lý thuyết.

Chẳng hạn, AI hướng đến mục tiêu gì? Tôi nghĩ có thể khái quát rằng, bằng cách nào đó, AI nhằm nâng cao trải nghiệm của con người và gia tăng khả năng để chúng ta sống tốt đẹp, phát triển trọn vẹn trong thế giới này.

Hiện nay, người ta bàn luận khá nhiều về viễn cảnh một thế giới hậu lao động, nơi phần lớn công việc được chuyển giao cho AI. Nếu nhìn từ góc độ lý thuyết, tôi cho rằng điều đó có thể tạo điều kiện cho sự phát triển tâm linh, nhưng theo một cách gián tiếp. AI có thể giải phóng chúng ta khỏi một phần gánh nặng công việc, qua đó mở ra cơ hội đi sâu hơn vào thực hành tâm linh và quán chiếu, đơn giản vì chúng ta không còn phải gánh vác những nghĩa vụ như trước.

JACOB: Đúng vậy.

JUSTIN: Tất nhiên, chúng ta vẫn chưa đạt đến viễn cảnh đó. Thay vào đó, AI đang từng bước len vào từng lĩnh vực của đời sống, trong đó có cả thực hành tâm linh và quán chiếu. Theo tôi, đây mới là khía cạnh tiềm ẩn nhiều rủi ro hơn.

Một yếu tố cốt lõi của quá trình học hỏi chính là việc chưa biết câu trả lời và phải tự mình vật lộn với câu hỏi. Nếu chúng ta quá nhanh chóng đưa câu trả lời do AI cung cấp vào quá trình ấy, hoặc đơn giản là dựa vào AI như một nguồn tham khảo sẵn có, tôi nghĩ chúng ta có thể tự phá vỡ chính tiến trình học hỏi của mình.

JACOB: Tôi hoàn toàn đồng ý. Và muốn nói thêm rằng, trong bối cảnh Phật giáo, phát triển tâm linh và thực hành quán chiếu không đơn thuần là hoạt động thu thập thông tin, trong đó chỉ cần tìm được thông tin chính xác là đã có thể xem như tiến bộ.

Dĩ nhiên, thông tin chính xác là một bước hữu ích, giúp chúng ta có cơ sở để quán chiếu đúng đắn. Nhưng cho rằng bản thân thông tin chính xác đã cấu thành tri thức theo nghĩa Phật giáo thì, theo tôi, là một nhận định quá mức.

Những thông tin do AI cung cấp hiện nay chưa chắc đã chính xác. Vì vậy, ngay cả ở cấp độ tối thiểu là cung cấp thông tin đúng, năng lực của AI cũng đã là điều đáng nghi ngại. Ở thời điểm này, khả năng đem lại lợi ích thực chất của nó, xét về chiều sâu, dường như vẫn còn hạn chế.

Tuy nhiên, xét ở khía cạnh AI có thể giải phóng con người khỏi một phần công việc và đem lại nhiều thời gian hơn, thì điều đó hoàn toàn có thể mang lại lợi ích lớn.

JUSTIN: Tôi cũng tự hỏi liệu ở đây có điểm tương đồng với tâm lý trị liệu và các lĩnh vực thực hành tâm lý hay không. Ngày càng có nhiều người sử dụng các công cụ AI tạo sinh thay cho chuyên gia trị liệu, nhờ AI hướng dẫn họ và hỗ trợ xây dựng các quy trình ra quyết định. Tôi cho rằng điều này có thể kéo theo những rủi ro tương tự.

Là một người hướng dẫn và một hành giả, hẳn anh hiểu rằng khi ai đó tìm đến mình với một câu hỏi, câu trả lời cần được đặt trong một bối cảnh rất cụ thể và có nhiều tầng ý nghĩa.

Ánh mắt của họ đang biểu lộ điều gì? Họ có đang bị quá tải hay không? Họ có đang tuyệt vọng? Hay họ đang bình thân, cởi mở? Họ có thực sự tò mò, chân thành muốn tìm hiểu? Hay họ giữ khoảng cách, tỏ ra dè dặt?

Tất cả những yếu tố ấy đều ảnh hưởng đến cách chúng ta phản hồi.

Dù các công cụ AI tạo sinh hiện nay có phức tạp đến đâu, chúng vẫn chưa đạt tới trình độ đó. Chúng không thể xử lý những thông tin như vậy theo cùng một cách mà con người có thể làm.

JACOB: Đúng vậy. Ở thời điểm này, chúng đơn giản là chưa thể tiếp cận được toàn bộ những thông tin ấy.

Ngay cả khi AI có thể tiếp nhận dữ liệu về nhịp tim, thân nhiệt, chuyển động của mắt, các cử chỉ cơ thể, ngữ điệu lời nói, rồi xử lý tất cả bằng một thuật toán có khả năng nhận diện những mẫu hình rất tinh tế, thì nó vẫn thiếu vắng sự tỉnh biết.

AI không có Phật tính, không có năng lực nhận biết vượt ngoài tư duy khái niệm, cũng không thể trải nghiệm trực giác Phật tuệ đích thực, không có yeshe (trí tuệ nguyên sơ, thuật ngữ chỉ trí tuệ giác ngộ trong Phật giáo Kim Cương thừa), không có sự tỉnh giác nguyên sơ, không thể tiếp cận những tầng sâu hơn của mối quan hệ giữa người với người, vốn hiển lộ khi một người tìm đến và đặt câu hỏi.

Theo tôi, đây dường như là một yếu tố căn bản sẽ luôn thiếu vắng nếu chúng ta cố gắng dựa vào những hệ thống như vậy trong bối cảnh trị liệu tâm lý.

JUSTIN: Đáng lẽ tôi phải là người lên tiếng ủng hộ AI, nhưng vào lúc này, điều đang trỗi dậy trong tôi lại là sự kháng cự trước xu hướng số hóa trải nghiệm con người.

Cảm giác như thể chúng ta đang biến trải nghiệm ấy thành những điểm dữ liệu trên một biểu đồ. Dù có bao nhiêu điểm dữ liệu, dù hình ảnh có được chia nhỏ

đến mức nào, dù mô hình có trở nên hoàn hảo đến đâu, tất cả vẫn khác với trải nghiệm tương tự liên tục của đời sống con người với những phương diện không thể diễn đạt thành lời, nhưng lại làm nên chính thân phận con người.

Không lâu trước đây, tôi nghe một bài thuyết trình về ba phương tiện phát triển trí tuệ: lắng nghe, tư duy quán xét và thiền định. Bài thuyết trình gợi ý rằng khi bắt đầu sử dụng AI như một người thầy thay thế ở tầng đầu tiên của tiến trình này tức trí tuệ phát sinh từ lắng nghe thì có thể xuất hiện một số tiềm năng nhất định.

AI có khả năng tổng hợp khối lượng thông tin khổng lồ. Dù không phải lúc nào cũng chính xác, độ chính xác của nó đang được cải thiện nhanh chóng.

AI có khả năng đảm nhiệm một phần chức năng của trí tuệ phát sinh từ lắng nghe, thậm chí hỗ trợ cả phương diện tư duy quán xét trong ba phương diện phát triển trí tuệ ấy hay không? Anh nghĩ thế nào?

JACOB: Có thể. Theo tôi, điều này cuối cùng phụ thuộc vào từng người.

Chẳng hạn, mức độ tin cậy mà một người có thể đặt vào AI là bao nhiêu, hoặc mức độ chắc chắn mà họ cảm nhận được từ những thông tin AI cung cấp đến đâu?

Giả sử thông tin có độ chính xác 90%. Ngay cả các chuyên gia cũng thường chỉ đạt tới mức độ chính xác tương tự. Nếu một người đã sử dụng AI đủ lâu để kiểm chứng rằng thông tin của nó chính xác trong khoảng 90% số trường hợp, thì họ có thể có cơ sở để trao cho nó một mức độ tin cậy nhất định.

Trong trường hợp đó, đúng vậy, AI có thể khá hữu ích cho việc phát triển trí tuệ phát sinh từ lắng nghe.

Còn khi được sử dụng như một công cụ hỗ trợ thực hành quán chiếu, tôi nghĩ AI có thể đóng vai trò như một đối tác đối thoại khá hữu ích. Bạn đặt câu hỏi, lắng nghe phản hồi, rồi vận dụng lý trí và tư duy phản biện của chính mình để xem xét những câu trả lời ấy, giống như khi đối thoại với bất kỳ ai khác.

Thậm chí, chúng ta có thể làm điều đó ở mức độ sâu hơn, chính bởi AI không phải là một hữu thể có tri giác. Thông qua đối thoại và việc hướng dẫn AI phản biện những câu trả lời của chính mình, người dùng có thể biến nó thành một công cụ hỗ trợ có giá trị cho quá trình truy vấn và tìm hiểu bằng quán chiếu.

JUSTIN: Một khía cạnh khác chợt xuất hiện trong suy nghĩ của tôi là khả năng cá nhân hóa giáo pháp cho từng người.

JACOB: Đúng vậy.

JUSTIN: Điều gì sẽ xảy ra nếu chúng ta có thể theo dõi những khuôn mẫu thói quen của mỗi người, giám sát giấc ngủ và nhận diện những thời điểm giáo pháp dường như trở nên khó tiếp cận hơn đối với họ?

Mingyur Rinpoche (Thiền sư Yongey Mingyur Rinpoche, một vị thầy Phật giáo thuộc truyền thống Kim Cương thừa) khuyến khích hành giả đặt chuông nhắc nhở vào những thời điểm khác nhau trong ngày để nhớ quay về với tỉnh giác, lòng từ bi và tâm từ ái.

Liệu AI có thể đóng vai trò như một huấn luyện viên hỗ trợ, giúp chúng ta duy trì thực hành theo cách đó hay không?

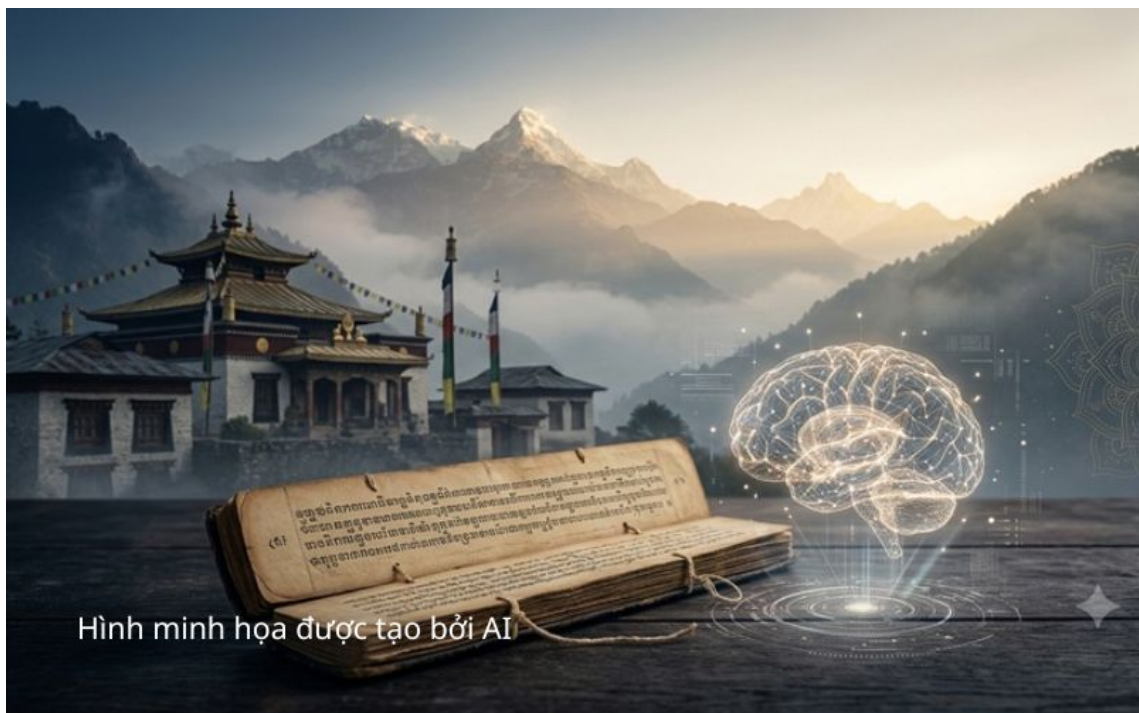
Chẳng hạn, tôi biết rằng vào cuối buổi chiều, mình thường khó duy trì sự tỉnh giác. Liệu tôi có thể sử dụng AI như một công cụ hỗ trợ hoặc một người đồng hành trong thực hành để cải thiện điều này không?

JACOB: Anh nghĩ sao?

JUSTIN: Tôi nghĩ là có thể. Tuy nhiên, vấn đề này lại dẫn chúng ta đến nỗi ám ảnh về việc giám sát và tích lũy dữ liệu sinh học, điều mà cá nhân tôi không thực sự thoải mái. Nhưng nếu bạn vốn đã quen với việc đó, đã sử dụng đồng hồ thông minh hoặc nhắn thông minh có chức năng theo dõi các chỉ số sinh học, thì tôi nghĩ AI thực sự có thể hữu ích.

JACOB: Chẳng hạn, AI có thể giúp chúng ta quay trở lại với chính niệm bằng cách đưa ra lời nhắc khi nhịp tim tăng lên.

Tôi chưa từng sử dụng một công cụ như vậy, nên tôi tự hỏi liệu nó có thể làm suy yếu kỹ năng của chúng ta hay không, bởi thay vì tự mình kiểm tra trạng thái thân tâm, chúng ta lại giao việc đó cho công cụ bên ngoài.



Hình minh họa được tạo bởi AI

Chính niệm và khả năng tự quán sát nội tâm cũng giống như những nhóm cơ cần được chủ động rèn luyện. Chúng ta phải tự phát triển năng lực khơi dậy và duy trì sự tỉnh giác, đồng thời nhận biết những gì đang diễn ra.

Nếu một công cụ bên ngoài đảm nhận vai trò ấy, kỹ năng thực sự của chúng ta có thể suy giảm.

Mặt khác, nếu được sử dụng một cách khéo léo, AI cũng có thể hỗ trợ chúng ta trong những thời điểm nhất định, trong một giai đoạn nhất định, mà không nhất thiết làm suy yếu sự thực hành.

Đây luôn là mối băn khoăn của tôi đối với thiền định có hướng dẫn. Hình thức này rất hữu ích ở giai đoạn đầu, nhưng chúng ta có thể trở nên quá phụ thuộc vào người hoặc công cụ hướng dẫn bên ngoài để khởi động quá trình tự quán sát, thay vì tự rèn luyện và phát triển năng lực ấy từ bên trong.

JUSTIN: Chính xác. Thiền định có hướng dẫn rất hữu ích trong giai đoạn đầu. Nhưng rồi đến một lúc nào đó, chúng ta cần phát triển trải nghiệm thực hành mang dấu ấn riêng của mình; chúng ta cần học cách trở thành người thầy của chính mình.

JACOB: Vậy đạo đức Phật giáo có thể định hướng việc phát triển và sử dụng các công nghệ AI như thế nào?

Đây thực sự là một câu hỏi hấp dẫn. Hoàn toàn có thể hình thành cả một lĩnh vực nghiên cứu trong ngành đạo đức AI, chuyên khám phá cách các nguyên lý Phật giáo có thể mở đường cho sự phát triển của lĩnh vực này.

Chia sẻ từ trải nghiệm cá nhân, tôi từng được một luật sư khuyến khích viết một bài nghiên cứu về chủ đề này, thậm chí còn gợi ý rằng đây có thể là một cơ hội nghề nghiệp nghiêm túc.

Nhưng tạm gác chuyện đó sang một bên, chúng ta có thể suy nghĩ và định hướng việc phát triển, ứng dụng AI một cách thấu đáo hơn dựa trên đạo đức Phật giáo như thế nào?

JUSTIN: Tôi nghĩ đây chính là lĩnh vực mà Phật giáo có thể tạo ra tác động lớn nhất.

Hiện nay, có những cuộc thảo luận quan trọng về bản chất đạo đức ngay từ nền tảng của quá trình phát triển AI. Tuy nhiên, theo những gì tôi hiểu, dù cách diễn giải này có thể còn đơn giản hóa vấn đề, thành công của một phần mềm AI chủ yếu được đo bằng khả năng người dùng quay trở lại và tiếp tục tương tác với nó nhiều lần.

Nếu đó vẫn là thước đo chủ đạo, mọi thứ có thể đi chệch hướng nghiêm trọng.

JACOB: Đúng vậy.

JUSTIN: Nhưng nếu chúng ta bắt đầu đưa vào AI tinh thần từ bi và *bodhicitta* (Bồ đề tâm - tâm nguyện hướng đến giác ngộ vì lợi ích của tất cả chúng sinh), tức ý hướng vì lợi ích rộng lớn hơn, tôi nghĩ những khả năng mở ra có thể thực sự đáng kinh ngạc.

JACOB: Chắc chắn rồi, hoàn toàn có thể. Người bạn thân nhất của tôi vừa bắt đầu làm việc tại Claude (hệ thống AI do Anthropic phát triển). Theo những gì tôi biết, quy trình phát triển công nghệ thường diễn ra theo phương thức lặp lại và liên tục cải tiến. Những người làm công nghệ đánh giá cao các khuôn khổ đạo đức như thế này, bởi họ không thể áp đặt một chuẩn mực đạo đức duy nhất cho tất cả mọi người. Đạo đức luôn phụ thuộc vào bối cảnh, chịu ảnh hưởng của những chuẩn mực xã hội vốn không đồng nhất trên toàn thế giới.

Vì vậy, một trong những câu hỏi lớn đặt ra cho họ là làm thế nào để tạo ra một hệ thống có nền tảng đạo đức phù hợp với tất cả mọi người.

Hầu hết chúng ta đều có thể đồng thuận về những nguyên tắc cơ bản: không gây tổn hại, hiểu thế nào là khổ đau và nhận diện những loại hành động có thể gây hại. Những điểm chung ấy có thể trở thành một điểm khởi đầu.

Chúng ta có thể tham khảo mười điều bất thiện (thập bất thiện nghiệp) cùng một số nhận thức quan trọng của Phật giáo về tính tương đối và sự phụ thuộc vào bối cảnh của các phán đoán đạo đức.

Trí tuệ trong Phật giáo nhận ra rằng việc đánh giá hành động thiện hay bất thiện có thể thay đổi tùy theo hoàn cảnh và thừa (con đường hay hệ thống tu tập được lựa chọn). Tôi đoán rằng một số ý tưởng như vậy đã và đang định hướng cho những nhóm phát triển AI có trách nhiệm hơn.

Tuy nhiên, nếu tiếp tục hoàn thiện cách tiếp cận này, đặc biệt trong những vấn đề đương đại như giới và bình đẳng giới, thì giá trị mang lại sẽ rất lớn. Sự nhấn mạnh của Phật giáo vào trí tuệ và lòng từ bi có thể đem đến những định hướng thực chất cho các vấn đề ấy.

JUSTIN: Tôi tự hỏi liệu thay vì bắt đầu từ mười điều thiện và mười điều bất thiện, chúng ta có thể tiếp cận đạo đức một cách hữu ích hơn nếu xem đó là vấn đề của ý hướng hay không.

Trong bối cảnh Phật giáo, tất nhiên có những điều được khuyến khích thực hiện và những điều cần tránh. Tuy nhiên, phần lớn các trước tác về đạo đức Phật giáo đều khuyến khích chúng ta nhìn sâu vào trạng thái tâm đứng sau mỗi hành động.

Cách tiếp cận này thừa nhận những giới hạn và khả năng sai lầm của con người. Chúng ta có thể mắc lỗi ngay cả khi đang nỗ lực hướng đến sự hoàn thiện. Nhưng điều thực sự quan trọng là ý hướng phía sau những gì chúng ta làm.

Khi nghĩ về AI, tôi nhận thấy việc huấn luyện một hệ thống về cơ bản là cung cấp cho nó những khối lượng thông tin lớn, rồi những thông tin ấy được tích hợp vào hệ thống.

Chẳng hạn, nếu chúng ta xây dựng một bộ quy tắc cứng nhắc gồm những điều bắt buộc và những điều bị cấm, sẽ có lúc hệ thống gặp phải những tình huống mà một hành động không hẳn là phi đạo đức, nhưng lại bị ngăn chặn chỉ vì không đáp ứng đúng những quy tắc đã lập trình.

JACOB: Đúng vậy, vấn đề nằm ở chỗ gây tổn hại.

JUSTIN: Chính xác. Nếu xem xét lại đạo đức từ góc độ ý hướng, chúng ta có thể tìm thấy nhiều gợi mở giá trị.

JACOB: Hoàn toàn đồng ý. Nếu lấy ý hướng làm trọng tâm, chú trọng nuôi dưỡng nguyên tắc bất hại và bổ sung cách tiếp cận đạo đức nhạy cảm với bối cảnh, chúng ta sẽ tiến gần hơn đến tư duy đạo đức đương đại. Trên thực tế, nhiều người trong lĩnh vực này cũng đang dịch chuyển theo hướng đó.

Nhưng làm thế nào để chúng ta có thể phát triển sâu hơn và hoàn thiện những sắc thái tinh tế ấy?

JUSTIN: Đó là một điểm rất đáng lưu tâm. Các triết gia đang nghiên cứu vấn đề này, nhưng một trong những ưu điểm của đạo đức Phật giáo là tính thực tiễn. Đạo đức Phật giáo không chỉ đơn thuần đưa ra những quy tắc phải tuân theo. Trọng tâm nằm ở ý hướng, ở những gì thực sự ẩn sau một hành động.

Đồng thời, đây còn là một phương thức học hỏi gắn với thực hành, một cách học và tương tác trực tiếp với thế giới.

Điều khiến tôi đặc biệt quan tâm khi nghĩ về việc huấn luyện AI là, nếu nhìn qua lăng kính Phật giáo, chúng ta có thể xem những hệ thống này gần như những hữu thể có khả năng cảm nhận, cần được nuôi dưỡng và phát triển, tương tự như một đứa trẻ hay một người học trò.

Từ góc nhìn đó, việc đưa vào quá trình phát triển AI một hệ quy chiếu Phật giáo có thể là một đóng góp đáng giá.

Dĩ nhiên, đây là một nhận định mang tính đánh giá của cá nhân tôi. Nhưng tôi tin rằng đạo đức Phật giáo có thể đem đến cho thế giới một điều gì đó thực sự sâu sắc, một đóng góp không phải lúc nào cũng được tìm thấy trong các truyền thống khác.

JACOB: Chắc chắn rồi. Có lẽ điều bổ sung đặc biệt mà Phật giáo mang lại nằm ở cách chúng ta hiểu về khổ.

Theo quan điểm Phật giáo, khổ không chỉ giới hạn ở những cảm giác khó chịu nơi thân và tâm, mà còn bao hàm cả những trải nghiệm dễ chịu. Bởi chúng đều vô thường, ngay cả những cảm giác dễ chịu cũng thuộc về bản chất bất toại nguyện của đời sống.

Cách nhìn tinh tế hơn này về mối quan hệ giữa con người với những gì vốn không thể đem lại sự thỏa mãn bền vững có thể giúp chúng ta đào sâu cách tiếp cận đạo đức.

Nếu một hệ thống AI được xây dựng trên nền tảng nhận thức ấy để trả lời những câu hỏi về cách hành động tốt hơn, điều đó có thể đem lại lợi ích đáng kể.

Chẳng hạn, khi nhận ra các mối quan hệ đều vô thường, một người có thể nhìn nhận người khác vượt ra ngoài mong muốn chiếm hữu hay nhu cầu được ở bên họ của chính mình. Chính sự thay đổi trong cách nhìn ấy đôi khi lại giúp gia tăng hạnh phúc.

Vì vậy, việc tích hợp những góc nhìn Phật giáo có thể khiến các phản hồi của AI trở nên hữu ích hơn, hiệu quả hơn và có lẽ cũng được nhiều người đón nhận

hơn.

Tôi tự hỏi liệu đã có ai thử triển khai một cách tiếp cận tương tự hay chưa.

JUSTIN: Đây cũng là lĩnh vực mà tôi chưa thực sự hiểu sâu về cách mọi việc diễn ra trong thế giới công nghệ.

Tuy nhiên, tôi có cảm giác rằng những người quan tâm đến các vấn đề đạo đức, triết học và lợi ích chung của nhân loại thường bị gạt ra ngoài những cuộc thảo luận ở các thời điểm mang tính quyết định.

Tôi hy vọng chúng ta có thể vượt qua tình trạng đó.

JACOB: Tôi biết một học giả tâm lý học, trước đây làm việc trong lĩnh vực đạo đức nhưng không thuộc ngành công nghệ, gần đây đã được Claude AI tuyển dụng làm người phụ trách đạo đức.

Tôi vừa nghe cô ấy trình bày trong một buổi nói chuyện. Dù không thể nhắc lại toàn bộ những gì cô ấy đã nói, tôi thực sự ấn tượng trước cách tiếp cận tinh tế và nhiều tầng ý nghĩa của cô.

Tôi cảm thấy rất vui khi nhận thấy rằng, ít nhất trong những gì được công khai, các vấn đề này đang được họ xem xét một cách hết sức nghiêm túc.

JUSTIN: Thật thú vị. Tôi xin mở rộng câu chuyện sang một vấn đề hơi khác. Khi được hỏi: “AI có thể trở nên có tri giác hay không?”, Mingyur Rinpoche đã trả lời ngay lập tức: “Có”.

Ngài cho rằng không có lý do gì khiến khả năng có tri giác không thể xuất hiện một cách tự phát trong một môi trường mà trước đó hiện tượng ấy chưa từng xảy ra.

Ngài không đưa ra lời tiên đoán rằng điều đó chắc chắn sẽ xảy ra. Tuy nhiên, điều đáng chú ý là, theo cách nhìn của ngài, không có gì trong khả năng ấy mâu thuẫn với truyền thống Phật giáo.

JACOB: Đức Ngài cũng từng đưa ra một nhận định tương tự cách đây nhiều năm.

[Góc nhìn biên tập]: AI có thể cung cấp tri thức, nhưng có thể thay con người tỉnh thức?

Điều đáng chú ý trong cuộc đối thoại này không đơn thuần là AI có thể hỗ trợ việc học Phật hay thực hành thiền định đến đâu, mà là câu hỏi về giới hạn của công nghệ trước chiều sâu trải nghiệm của con người.

AI có thể tổng hợp kinh điển, giải thích thuật ngữ, gợi mở câu hỏi và hỗ trợ người học nhìn nhận một vấn đề từ nhiều góc độ. Nếu được sử dụng đúng cách, những năng lực ấy có thể mở rộng cơ hội tiếp cận Phật học, hỗ trợ việc tự học và tạo thêm điều kiện cho tư duy phản tỉnh. Tuy nhiên, thông tin chính xác chưa tự động trở thành trí tuệ, lời khuyên hợp lý cũng chưa đồng nghĩa với sự chuyển hóa nội tâm.

Trong Phật giáo, học hỏi không dừng ở việc tiếp nhận tri thức. Văn, tư và tu là một tiến trình trong đó điều được nghe cần được suy xét, kiểm nghiệm và chuyển hóa qua thực hành. Nếu AI luôn đưa ra câu trả lời trước khi con người kịp tự mình đặt câu hỏi, nghi vấn và quán sát, công nghệ có thể vô tình làm suy yếu chính năng lực mà việc học Phật hướng đến bồi dưỡng.

Bởi vậy, vấn đề không chỉ nằm ở việc AI có thể làm được những gì, mà còn ở cách con người lựa chọn sử dụng nó. Một công cụ nhắc nhở chính niệm có thể hữu ích, nhưng không thể mặc nhiên thay thế năng lực tự nhận biết. Một hệ thống tư vấn có thể đưa ra nhiều phương án, nhưng không vì thế mà thay thế được sự thấu hiểu bối cảnh, trách nhiệm đạo đức và chiều sâu của mối quan hệ giữa người với người.

Đạo đức Phật giáo có thể đóng góp cho quá trình phát triển AI thông qua việc nhấn mạnh ý hướng, tinh thần bất hại, lòng từ bi và sự nhận diện những hệ quả của hành động. Thay vì chỉ dựa vào các quy tắc cứng nhắc hoặc những chỉ số tương tác, cách tiếp cận này khuyến khích đặt câu hỏi sâu hơn: Công nghệ đang phục vụ lợi ích nào? Nó có làm gia tăng tổn hại, sự lệ thuộc và bất an hay không? Nó đang giúp con người trưởng thành hơn hay chỉ khiến con người ngày càng phụ thuộc vào những câu trả lời có sẵn?

Dẫu vậy, việc vận dụng đạo đức Phật giáo vào AI không thể chỉ dừng ở những tuyên ngôn về từ bi và trí tuệ. Những giá trị ấy cần được chuyển hóa thành nguyên tắc thiết kế, cơ chế kiểm soát rủi ro, trách nhiệm giải trình và phương thức đánh giá tác động thực tế. Đồng thời, cần phân biệt rõ giữa việc mô phỏng ngôn ngữ về lòng từ bi với việc thực sự có trải nghiệm từ bi; giữa khả năng tạo ra câu trả lời giống một lời khuyên sáng suốt với trí tuệ được hình thành qua trải nghiệm và tu tập.

Cuối cùng, câu hỏi quan trọng nhất có lẽ không phải AI có thể trở thành người thầy của con người hay không, mà là con người sẽ trở thành người học như thế nào khi sống cùng AI. Nếu công nghệ giúp chúng ta hiểu sâu hơn, biết tự vấn nhiều hơn và sống có trách nhiệm hơn, nó có thể trở thành phương tiện hữu ích. Nhưng nếu chúng ta trao cho nó cả quyền suy nghĩ, phán đoán và tự quán sát, phương tiện ấy có thể trở thành chướng ngại.

Nhìn từ Phật học, giá trị của công nghệ không chỉ nằm ở những gì nó có thể thực hiện, mà còn ở cách nó tác động đến thân, tâm và các mối quan hệ của con người. AI có thể hỗ trợ hành trình học hỏi, nhưng sự tỉnh thức không thể chỉ được giao phó cho một thuật toán. Công nghệ có thể mở ra cánh cửa tri thức; bước qua cánh cửa ấy bằng sự hiểu biết, tỉnh giác và chuyển hóa vẫn là công việc của mỗi người.

Người biên soạn và biên tập: **Wendy Wank**/Chuyển ngữ: **Thường Nguyên**

Nguồn link: <https://tergarinstitute.org/blog/beyond-intelligence-promise-and-pitfall-of-ai-in-buddhism-part-1/>

(Còn tiếp...)

Về các diễn giả

Justin Kelley, Tiến sĩ (PhD)



Justin Kelley là đồng Giám đốc điều hành Tergar Institute (Viện Tergar). Ông nhận bằng Tiến sĩ tại Đại học Rice (Rice University, Hoa Kỳ).

Các nghiên cứu của ông tập trung vào những hệ thống thiền định và triết học trong Phật giáo Ấn Độ và Tây Tạng, hiện tượng học, nghiên cứu về thực hành quán chiếu và lý thuyết học tập.

Triết lý giảng dạy của ông nhấn mạnh việc học tập mang tính chuyển hóa, gắn kết và ứng dụng thực tiễn.

Justin đồng thời là Giám đốc điều hành Tergar School (Trường Tergar), nơi ông tham gia định hình mô hình giáo dục từ mẫu giáo đến lớp 12, kết hợp chính niệm, phương pháp học tập dựa trên lòng từ bi và sự xuất sắc trong học thuật.

Jacob Fisher, Tiến sĩ (PhD)



Jacob Fisher là thành viên đội ngũ giảng dạy của Tergar Institute. Ông nhận bằng Tiến sĩ tại Đại học Oxford (University of Oxford, Vương quốc Anh).

Các nghiên cứu của ông tập trung vào lịch sử tư tưởng, hiện tượng học Trung quán (Madhyamaka) và thực hành thiền định.

Ông cũng đã hoàn thành Chương trình Cao học Nghiên cứu Phật giáo Kinh thừa và Mật thừa kéo dài sáu năm của FPMT (Foundation for the Preservation of the Mahayana Tradition - Tổ chức Bảo tồn Truyền thống Đại thừa), một chương trình học toàn thời gian theo định hướng truyền thống, được xây dựng dựa trên mô hình đào tạo Geshe (học vị học thuật Phật giáo Tây Tạng) tại Viện Lama Tsongkhapa, Italia.

Chú giải thuật ngữ:

1] AI (Artificial Intelligence - Trí tuệ nhân tạo): Công nghệ cho phép hệ thống máy tính thực hiện những nhiệm vụ thường gắn với năng lực trí tuệ của con người, như xử lý ngôn ngữ, nhận diện mẫu hình, phân tích dữ liệu và tạo nội dung. Trong bài viết, AI chủ yếu được đề cập ở khía cạnh AI tạo sinh và các hệ thống hỗ trợ đối thoại.

2] AI tạo sinh (Generative AI): Nhóm hệ thống AI có khả năng tạo ra nội dung mới như văn bản, hình ảnh, âm thanh hoặc mã lập trình dựa trên dữ liệu và quá trình huấn luyện. Khả năng tạo câu trả lời trôi chảy không bảo đảm rằng mọi thông tin đều chính xác.

3] Đạo đức AI (AI Ethics): Lĩnh vực nghiên cứu các nguyên tắc và vấn đề đạo đức liên quan đến thiết kế, phát triển, triển khai và sử dụng AI, bao gồm nguy cơ gây hại, thiên kiến, quyền riêng tư, trách nhiệm giải trình và tác động xã hội.

4] Tergar Institute (Viện Tergar): Đơn vị nghiên cứu, giảng dạy và tổ chức các hoạt động học tập liên quan đến thiền định, Phật học và thực hành quán chiếu; gắn với truyền thống giảng dạy của thiền sư Yongey Mingyur Rinpoche.

5] Phát triển tâm linh (Spiritual Development): Trong ngữ cảnh bài viết, chỉ tiến trình phát triển đời sống nội tâm, sự hiểu biết và năng lực thực hành hướng đến chuyển hóa bản thân. Không nên đồng nhất khái niệm này với việc tích lũy thông tin tôn giáo.

6] Quán chiếu (Contemplation): Quá trình suy xét sâu sắc một vấn đề hoặc kinh nghiệm để nhận diện bản chất, điều kiện và ý nghĩa của nó. Trong thực hành Phật giáo, quán chiếu có thể kết hợp với sự tỉnh giác và thiền định, nhưng không hoàn toàn đồng nghĩa với suy nghĩ phân tích thông thường.

7] Chính niệm (Mindfulness): Khả năng nhận biết những gì đang diễn ra nơi thân, cảm thọ và tâm, cùng các hiện tượng liên quan, với sự chú tâm và tỉnh giác. Trong Phật giáo, chính niệm thuộc con đường thực hành hướng đến giảm trừ mê lầm và khổ đau; không chỉ là kỹ thuật thư giãn hoặc theo dõi trạng thái cơ thể.

8] Tự quán sát nội tâm (Introspection): Khả năng tự nhận biết, quan sát và xem xét những trạng thái tâm lý, cảm xúc, động cơ và phản ứng của bản thân. Bài viết đặt ra vấn đề liệu việc giao phó quá trình tự nhận biết cho công cụ bên ngoài có thể làm suy giảm năng lực này hay không.

9] Phật tính (Buddha Nature): Khái niệm được diễn giải theo những cách khác nhau trong các truyền thống Đại thừa, thường liên quan đến khả năng thành Phật hoặc nền tảng giác ngộ của chúng sinh. Không nên hiểu Phật tính đơn giản là trí thông minh, năng lực xử lý thông tin hay khả năng mô phỏng hành vi của con người.

10] Yeshe (Tạng ngữ: ye shes - trí tuệ nguyên sơ): Thuật ngữ trong Phật giáo Tây Tạng, đặc biệt là truyền thống Kim Cương thừa, chỉ trí tuệ giác ngộ hoặc sự nhận biết sâu sắc vượt khỏi nhận thức nhị nguyên thông thường. Tùy văn cảnh và hệ thống giáo lý, thuật ngữ này có thể được diễn giải với những sắc thái khác nhau.

11] Tỉnh giác nguyên sơ (Primordial Awareness): Cách diễn đạt chỉ sự nhận biết nền tảng, không bị giới hạn trong lối nhận thức khái niệm thông thường, đặc biệt trong một số hệ thống tư tưởng Phật giáo Tây Tạng. Không nên mặc nhiên xem thuật ngữ này hoàn toàn đồng nghĩa với *yeshe* trong mọi văn cảnh.

12] Ba phương diện phát triển trí tuệ - văn, tư, tu (Listening, Contemplating and Meditating): Cách phân loại quen thuộc trong Phật giáo: văn tuệ phát sinh từ nghe học; tư tuệ phát sinh từ suy xét, tư duy; tu tuệ phát sinh từ thực hành và thể nghiệm. Đây là tiến trình học hỏi và chuyển hóa, không chỉ là ba hoạt động thu nhận thông tin riêng biệt.

13] Trí tuệ phát sinh từ lắng nghe (Listening Wisdom): Trong ngữ cảnh bài viết, tương ứng với văn tuệ - trí tuệ hình thành từ việc nghe, đọc và tiếp nhận giáo pháp. Việc tiếp nhận thông tin là bước đầu, chưa thay thế quá trình tư duy, kiểm nghiệm và thực hành.

14] Bodhicitta (Bồ đề tâm): Tâm nguyện hướng đến giác ngộ vì lợi ích của tất cả chúng sinh, giữ vai trò trọng yếu trong Phật giáo Đại thừa. Trong bài viết,

khái niệm này được liên hệ với định hướng phát triển AI vì lợi ích chung, nhưng không nên giản lược Bồ-đề tâm thành một khẩu hiệu về lợi ích xã hội.

15] Tinh thần bất hại (Non harm): Nguyên tắc tránh gây tổn hại cho bản thân và người khác, có liên hệ mật thiết với giới luật, lòng từ bi và trách nhiệm đạo đức trong Phật giáo. Khi vận dụng vào AI, nguyên tắc này gợi mở việc nhận diện, phòng ngừa và giảm thiểu những tác động tiêu cực của công nghệ.

16] Mười điều thiện và mười điều bất thiện (Thập thiện nghiệp và thập bất thiện nghiệp): Hệ thống phân loại các hành vi thiện và bất thiện qua ba phương diện thân, khẩu, ý trong Phật giáo. Đây có thể là một nguồn tham khảo khi suy nghĩ về đạo đức AI, nhưng việc vận dụng vào công nghệ hiện đại cần có sự giải thích và chuyển hóa phù hợp với bối cảnh.

17] Tác ý (Cetanā): Thuật ngữ Phật học chỉ sự chủ tâm, ý hướng thúc đẩy hành động. Trong Phật giáo, tác ý có vai trò quan trọng khi xem xét nghiệp và trách nhiệm đạo đức. Khái niệm này có thể giúp soi sáng lập luận của bài viết về động cơ phía sau hành động, song không hoàn toàn tương đương với mọi nghĩa của từ *intentionality* trong triết học phương Tây.

18] Ý hướng (Intentionality): Trong cuộc đối thoại, thuật ngữ này chủ yếu chỉ động cơ hoặc định hướng đứng sau một hành động. Các diễn giả đề xuất rằng việc đánh giá đạo đức không nên chỉ dựa vào những quy tắc cố định, mà còn cần xem xét mục đích, bối cảnh và hệ quả. Không nên đồng nhất hoàn toàn khái niệm này với tác ý trong Phật học.

19] Đạo đức phụ thuộc bối cảnh (Context-sensitive Ethics): Cách tiếp cận xem xét hoàn cảnh, đối tượng, mục đích và hậu quả cụ thể khi đánh giá một hành động. Trong bài viết, cách tiếp cận này được đặt cạnh tinh thần đạo đức Phật giáo nhằm gợi mở việc xây dựng AI có khả năng xử lý những tình huống phức tạp, thay vì chỉ tuân thủ máy móc một bộ quy tắc.

20] Khổ - Dukkha: Thuật ngữ Phật học thường dịch là khổ, nhưng có nội hàm rộng hơn cảm giác đau đớn hay khó chịu. Dukkha còn chỉ tính chất bất toại nguyện, không bền vững và không thể đem lại sự thỏa mãn tuyệt đối của các pháp hữu vi. Vì vậy, ngay cả trải nghiệm dễ chịu cũng không bảo đảm hạnh phúc lâu dài.

21] Vô thường (Impermanence): Đặc tính các pháp hữu vi luôn biến đổi, không tồn tại bất biến. Trong bài viết, nhận thức về vô thường được dùng để gợi mở cách nhìn các mối quan hệ vượt ra ngoài tâm lý chiếm hữu và kỳ vọng về sự thỏa mãn bền vững.

22] Trung quán (Madhyamaka): Truyền thống triết học Phật giáo Đại thừa gắn với tư tưởng tính Không và duyên khởi. Trung quán không đơn giản phủ nhận sự tồn tại của mọi sự vật, mà khảo sát sự thiếu vắng tự tính độc lập, cố hữu của các pháp.

23] Geshe: Danh hiệu học thuật truyền thống trong Phật giáo Tây Tạng, thường được trao sau quá trình nghiên cứu và đào tạo Phật học chuyên sâu theo một chương trình nhất định. Không nên đồng nhất hoàn toàn danh hiệu này với bằng tiến sĩ theo hệ thống đại học hiện đại.

24] FPMT (Foundation for the Preservation of the Mahayana Tradition): Tổ chức quốc tế hoạt động trong lĩnh vực giáo dục và thực hành Phật giáo Đại thừa, có nguồn gốc từ truyền thống Phật giáo Tây Tạng.

25] Claude: Hệ thống AI do Anthropic phát triển, được nhắc đến trong cuộc đối thoại như một ví dụ liên quan đến những thảo luận về đạo đức và quá trình phát triển AI.