

Đạo đức Phật giáo định hướng phát triển trí tuệ nhân tạo AI?

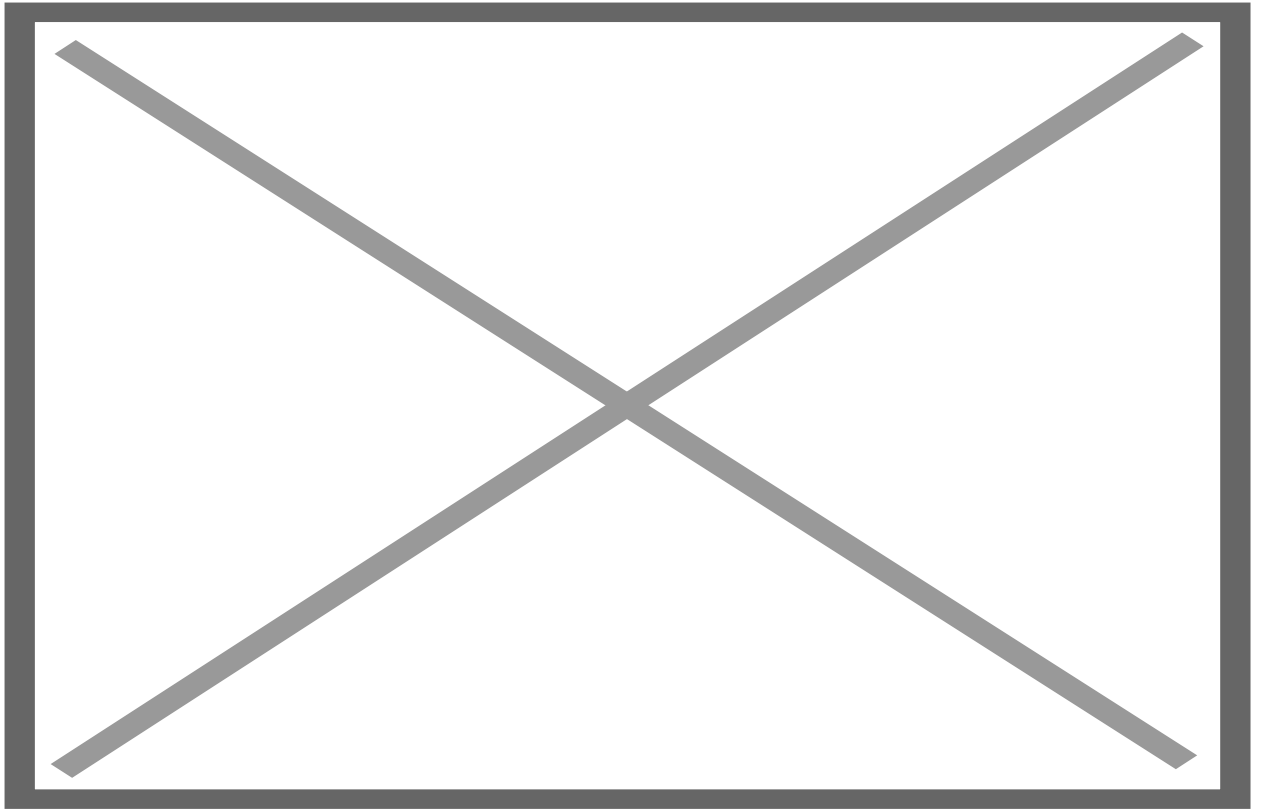
ISSN: 2734-9195 20:11 24/08/2021

Đạo Phật dạy chúng ta tập trung năng lượng của chính mình vào việc hóa giải những nỗi khổ, niềm đau của con người.

Đạo Phật dạy chúng ta tập trung năng lượng của chính mình vào việc hóa giải những nỗi khổ, niềm đau của con người.

Sự phát triển công nghệ trí tuệ nhân tạo (AI) bùng nổ, đã nuôi dưỡng hy vọng cho nhân loại, giúp giải quyết nhiều vấn đề nan giải của thế giới. Tuy nhiên, cũng có nhiều lo ngại về sức mạnh và những phạm trù đạo đức trong các ứng dụng và hoạt động của AI.

Các nhóm hoạt động xã hội đã tích cực thảo luận, đề xuất các hướng dẫn đạo đức về cách thức phát triển AI. Hội Kỹ sư Điện và Điện tử (Institute of Electrical and Electronics Engineers, IEEE), một tổ chức nghề nghiệp toàn cầu dành cho các kỹ sư, đã lưu hành một tài liệu dài 280 trang về chủ đề này, trong đó có các Quy tắc đạo đức trong việc sử dụng trí tuệ nhân tạo (AI).



Hầu hết các hướng dẫn này do các nhóm, hoặc tổ chức tập trung ở Bắc Mỹ và Châu Âu phát triển: Một cuộc khảo sát do nhà khoa học xã hội Anna Jobin và các đồng nghiệp công bố cho thấy có 21 bản hướng dẫn tại Hoa Kỳ, 19 ở Liên minh Châu Âu, 13 ở Vương quốc Anh, 4 ở Nhật Bản, và một số khác từ các Tiểu vương quốc Ả Rập Thống nhất, Ấn Độ, Singapore và Hàn Quốc.

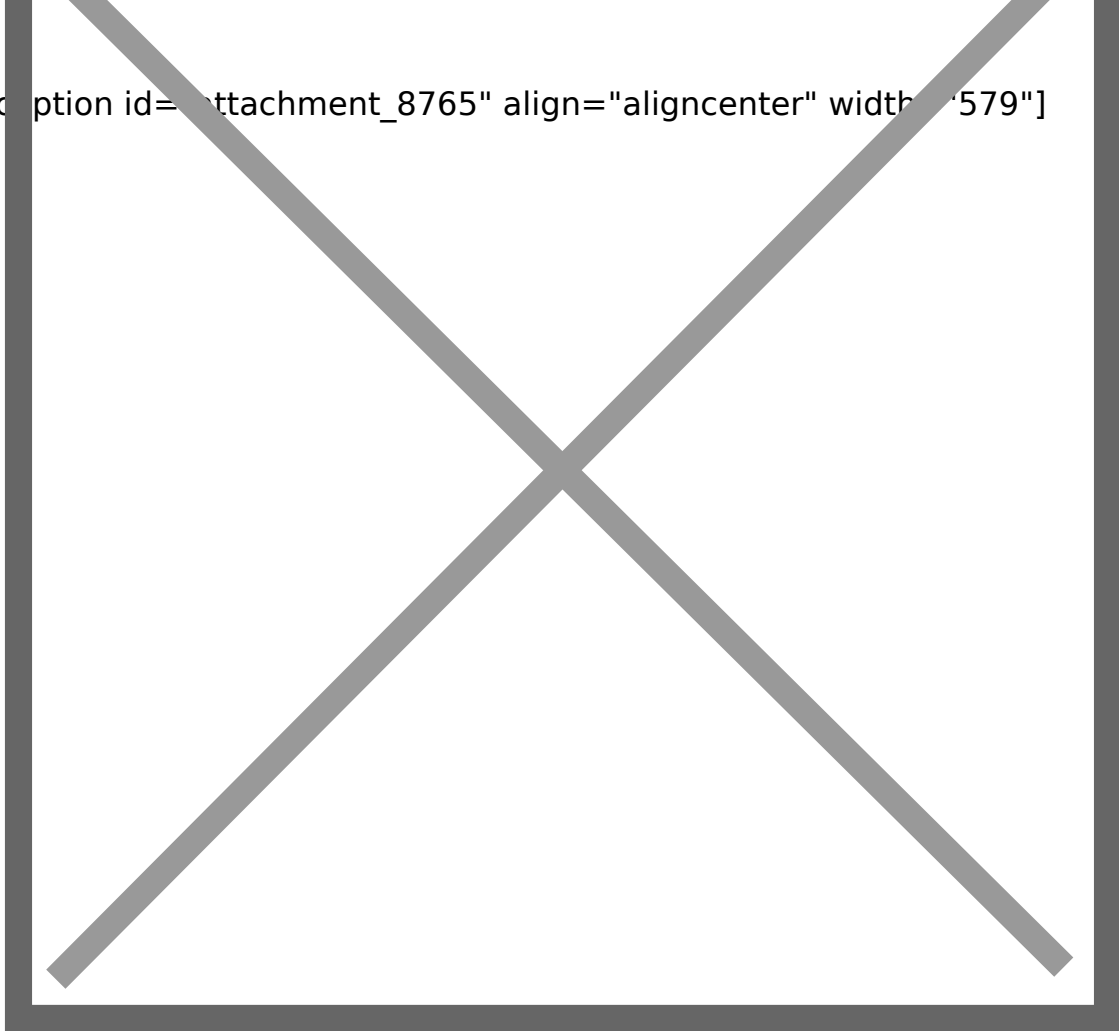
Đa số các văn bản Quy tắc đạo đức trí tuệ nhân tạo (AI) được xây dựng tại các quốc gia phương Tây, có nghĩa là lĩnh vực này bị chi phối bởi các giá trị phương Tây, như sự tôn trọng quyền tự chủ và quyền của cá nhân. Trong khi đó, các hướng dẫn được soạn thảo ngoài những điểm tương đồng cần có đại diện cho các giá trị truyền thống trong mỗi nền văn hóa nhân bản của con người.

Tất cả mọi người ở cả phương Đông và phương Tây cần chia sẻ ý tưởng của họ, và xem xét ý tưởng của những người khác, để làm phong phú quan điểm của họ. Bởi vì AI được phát triển và được sử dụng trên toàn cầu, cách chúng ta nghĩ về nó phải thỏa mãn tất cả các truyền thống trí tuệ.

Với những cách tư duy và lập luận như vậy, những giá trị của đạo Phật có thể được ứng dụng và mang lại lợi ích cho bất kỳ ai làm việc về AI ở bất cứ nơi đâu trên thế giới, nó thích nghi với mọi nền văn hóa.

Đạo đức Phật học dựa trên giả định rằng, tất cả chúng sinh đều muốn xa bể khổ nguồn mê. Bởi vậy, đức Phật dạy rằng, tịnh hóa tam nghiệp (ý nghĩ, lời nói và hành động) sẽ giúp thoát khỏi những nỗi khổ niềm đau.

[caption id="attachment_8765" align="aligncenter" width="579"]



Giáo sư

Soraj Hongladarom[/caption]

Hàm ý lời vàng ngọc nêu trên đối với trí tuệ nhân tạo, là bất kỳ hoạt động sử dụng AI có đạo đức đều phải nỗ lực giúp giảm bớt đớn đau và sầu khổ của kiếp nhân sinh.

Ví dụ, chỉ nên sử dụng công nghệ nhận dạng khuôn mặt nếu nó có thể được chứng minh, giúp giảm bớt lo âu phiền muộn hoặc tăng cường hạnh phúc. Hơn nữa, mục tiêu phải là giảm thiểu khổ não cho tất cả mọi người - không chỉ những người trực tiếp tương tác với AI.

Tất nhiên, chúng ta có thể diễn giải mục tiêu này một cách rộng rãi như ứng dụng AI, nhằm tu bổ một hệ thống hoặc quy trình không đạt yêu cầu, hoặc thay đổi bất kỳ tình huống nào cho hoàn hảo hơn. Sử dụng công nghệ để phân biệt đối xử với mọi người, hoặc để quan sát và đàn áp họ, rõ ràng đây là việc làm phi đạo đức. Khi xuất hiện các vùng "màu xám", hay nói cách khác là bản chất

không rõ ràng, trách nhiệm chứng minh sẽ thuộc về những người ứng dụng AI, làm sao điều khiển AI không gây hại.

Một nhà đạo đức học AI lấy cảm hứng từ đạo Phật cũng sẽ hiểu rằng, sống với những quy tắc này đòi hỏi sự tư tu tâm dưỡng tính. Điều này có nghĩa là những người có liên quan đến AI, nên liên tục đào tạo bản thân để tiến gần đến mục tiêu giải phóng tất cả phiền não khổ đau. Đạt được mục tiêu không quá quan trọng; điều quan trọng là họ thực hành để đạt được các giá trị cao quý nêu trên.

Các nhà thiết kế và lập trình nên thực hành bằng cách, nhận ra mục tiêu này và đặt ra các bước cụ thể, mà công việc của họ sẽ thực hiện để sản phẩm của họ đạt được lý tưởng. Tức là, ứng dụng AI mà họ sản xuất ra, phải nhằm mục đích giúp công chúng hóa giải những nỗi khổ niềm đau và nhân rộng sĩ số hạnh phúc.

Để những điều này có thể thực hiện, các công ty và cơ quan chính phủ phát triển, hoặc sử dụng AI phải chịu trách nhiệm trước công chúng. Trách nhiệm giải trình cũng là đạo lý nhà Phật, và trong bối cảnh đạo đức của AI, nó đòi hỏi các cơ chế pháp lý và chính trị hiệu quả cũng như sự độc lập về tư pháp. Những thành phần này rất cần thiết để mọi quy tắc đạo đức liên quan đến ứng dụng công nghệ AI hoạt động như dự kiến.

Một khái niệm quan trọng khác trong đạo Phật là từ bi, trí tuệ, hay chính là niềm hy vọng và cam kết giải phóng tất cả phiền não khổ đau cho tha nhân. Từ bi tâm phải đòi hỏi sự tư dưỡng của bản thân, nghĩa là những hành động có hại như dùng sức mạnh của một người để áp đảo người khác, không có chỗ đứng trong đạo đức Phật giáo. Người ta không nhất thiết phải xuất gia tu hành để thực hành đạo đức Phật giáo, nhưng người ta phải tự vun bồi từ bi tâm trong cuộc sống thường nhật.

Có thể thấy rằng các giá trị được đề cao trong đạo Phật - như trách nhiệm giải trình, công lý và từ bi tâm - hầu hết giống với những giá trị trong tất cả các truyền thống đạo đức khác nhau. Tất cả chúng đều vì hạnh phúc của con người!

Đạo lý Phật giáo có nhiều lời khuyên dành cho những ai băn khoăn về việc sử dụng công nghệ một cách đạo đức, kể cả những người quan tâm đến AI. Điều này cũng đúng với nhiều hệ thống giá trị phương Tây khác. Quy tắc đạo đức AI nên dựa trên sự đa dạng phong phú của những tư tưởng đến từ nhiều nền văn hóa trên thế giới. Tương lai của công nghệ sẽ chỉ tươi sáng hơn nếu được đúng như thế!

Tác giả Tiến sĩ Soraj Hongladarom, Giáo sư triết học, Giám đốc Trung tâm Đạo đức Khoa học và Công nghệ tại Đại học Chulalongkorn ở Bangkok, Thái Lan, đã viết và xuất bản những cuốn sách, và các bài báo về các vấn đề đa dạng như đạo đức sinh học, đạo đức máy tính, và vai trò của khoa học công nghệ trong nền văn hóa của các nước đang phát triển và nền văn hóa chung của nhân loại.

Tác giả: **Giáo sư Soraj Hongladarom** Biên dịch: **Thích Vân Phong** Nguồn: *MIT Technology Review*