

Những nguyên tắc đạo đức Phật giáo giúp AI hoàn hảo hơn

ISSN: 2734-9195 13:10 05/08/2024

Sức mạnh và tính phổ biến của nó đã dẫn đến vô số hướng dẫn, nhằm đảm bảo an toàn, bảo vệ các giá trị cũng như quyền của con người.

Nguyên tắc số một trong hướng dẫn của Vương quốc Phật giáo Thái Lan do Bộ Kinh tế Kỹ thuật số (Kinh tế Internet, Kinh tế web, Kinh tế mới) công bố, nêu lên rằng các hệ thống AI phải được sử dụng để nâng cao khả năng cạnh tranh kinh tế của quốc gia

Tác giả: **Soraj Hongladarom**

Viết dịch: **Thích Vân Phong**

Nguồn: [1]

Bài viết này là một phần trong loạt bài thuyết trình của các diễn giả tại Hội nghị gần đây của tổ chức Hiệp hội Quốc tế Mộc Ngư (Woodenfish Foundation, [2]) về chủ đề “Phật giáo, Ý thức và AI, Buddhism, Consciousness, and AI” được tổ chức vào ngày 21-23 tháng 6 tại Đài Bắc, thủ phủ của Đài Loan.

[caption id="attachment_36170" align="aligncenter" width="800"]
Phật giáo, Ý thức và AI **Phật giáo, Ý thức và AI**[/caption]

Tổ chức Hiệp hội Quốc tế Mộc Ngư (Woodenfish Foundation, [2]) do một học giả và nhà văn người Đài Loan, Ni sư Tiến sĩ Y Pháp (Yifa, [3]) sáng lập 2019.

Trong những năm gần đây, không thể phủ nhận rằng trí tuệ nhân tạo (AI) dẫn đầu cuộc cách mạng công nghệ năm 2024 và được nhắc đến nhiều nhất. Sức mạnh và tính phổ biến của nó đã dẫn đến vô số hướng dẫn, nhằm đảm bảo an toàn, bảo vệ các giá trị cũng như quyền của con người. “Nghiên cứu Hướng dẫn Nguyên tắc Đạo đức AI toàn cầu” (study by the AI Ethics Guidelines Global Inventory) cho thấy trên toàn thế giới có đến 167 hướng dẫn. Tuy nhiên, bốn

năm về trước, lần thống kê vào tháng 4 năm 2020 được hoàn thành lần cuối, và có mọi lý do để tin rằng hiện nay có nhiều hướng dẫn hơn nữa.

Những hướng dẫn này chủ yếu trình bày một số “Nguyên tắc Đạo đức” mà các nhà sản xuất, kỹ sư và những người liên quan đến công nghệ AI được kỳ vọng sẽ tuân theo. Chúng không mang tính ràng buộc về mặt pháp lý, nhưng chúng thể hiện mối quan ngại mà nhiều người cảm thấy rằng AI cần phải hành xử có đạo đức - tức là các hệ thống trí tuệ nhân tạo (AI) được kỳ vọng sẽ bảo vệ con người, bảo vệ quyền của họ và không gây hại cho con người theo bất kỳ cách nào.

Những nguyên tắc Phật giáo này được chia sẻ bởi hầu hết các hướng dẫn về Đạo đức của AI. Một nghiên cứu của nhà khoa học xã hội tập trung vào các khía cạnh xã hội của hệ thống thuật toán, Chủ tịch Ủy ban Truyền thông Liên bang Thụy Sĩ, nữ Tiến sĩ Anna Jobin và nhóm của bà, phân tích 84 hướng dẫn về Đạo đức của AI, tất cả đều ở phương Tây và Nhật Bản, cho thấy trong số các nguyên tắc Phật giáo được chia sẻ nhiều nhất là tính minh bạch (rằng công việc của hệ thống AI không được nằm trong “hộp đen, black box”, không thể nói cách chúng đưa ra giải pháp), công lý và công bằng (rằng các hệ thống phải công bằng và không phân biệt đối xử), không gây tổn hại, trách nhiệm và quyền riêng tư, nữ Tiến sĩ Anna Jobin và nhóm của bà nói thêm rằng, không có nguyên tắc nào được chia sẻ bởi tất cả các hướng dẫn mà họ nghiên cứu. Do đó, có cả điểm tương đồng và khác biệt giữa các hướng dẫn này.

Một nghiên cứu khác mà tôi thực hiện cùng với trợ lý nghiên cứu của mình là Giáo sư Tiến sĩ Jerd Bandasak, cho thấy rằng ngoài các nguyên tắc Phật giáo cốt lõi này, một số các nguyên tắc Phật giáo được tìm thấy trong một số ít hướng dẫn. (Hongladarom và Bandasak, 2023).

Ví dụ, nguyên tắc số một trong hướng dẫn của Vương quốc Phật giáo Thái Lan do Bộ Kinh tế Kỹ thuật số (Kinh tế Internet, Kinh tế web, Kinh tế mới) công bố, nêu lên rằng các hệ thống AI phải được sử dụng để nâng cao khả năng cạnh tranh kinh tế của quốc gia. Người ta tự hỏi khả năng cạnh tranh kinh tế làm thế nào có thể được coi là một nguyên tắc đạo đức. Nhưng điều này cho thấy những người soạn thảo hướng dẫn của Vương quốc Phật giáo Thái Lan nghĩ thế nào về những gì AI có thể làm được trong bối cảnh xã hội của họ.

Cả điểm tương đồng và điểm khác biệt trong các hướng dẫn về Đạo đức của AI này đều thú vị theo cách riêng của chúng. Đối với điểm tương đồng, người ta tự hỏi có bao nhiêu điểm tương đồng, thực sự khi các hướng dẫn này được xây dựng và sử dụng trong các bối cảnh xã hội và văn hóa khác nhau.

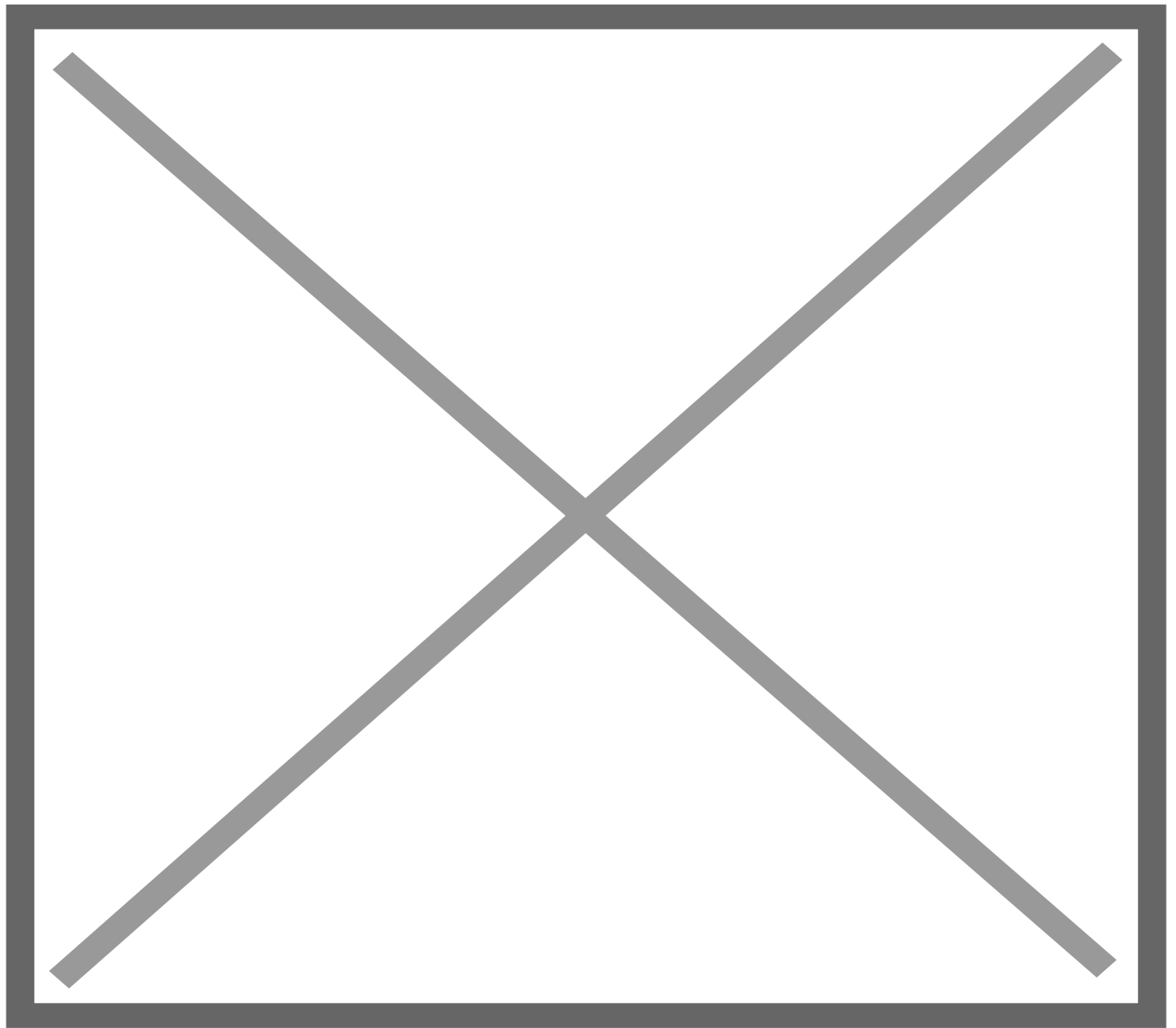
Theo đề tài “*Đồng cam cộng khổ: Tranh cãi Đạo đức tại địa phương và truyền đi khắp nơi*, Michael Walzer, “*Thick and Thin: Moral Argument at Home and Abroad*” (Nhà xuất bản Đại học Notre Dame 2019), khi một người xem thấy những người biểu tình trên phát thanh truyền hình ở một vùng đất xa lạ cảm biểu ngữ đòi “công lý, justice” sẽ hiểu thương và cảm thông cho họ, nhưng khi nói đến trách nhiệm cụ thể là nêu rõ công lý thực sự chính nghĩa và công lý đại diện cho điều gì, chắc chắn sẽ có những bất đồng.

Đối với điểm khác biệt, điều này cho thấy những gì được coi là nguyên tắc đạo đức thực sự khác nhau giữa các nền văn hoá. Những người soạn thảo hướng dẫn của Vương quốc Phật giáo Thái Lan có thể hành động một cách thiện chí, thực sự tự tin rằng khả năng cạnh tranh kinh tế là một nguyên tắc đạo đức, và những người đồng cấp Trung Hoa Dân Quốc của họ cũng thế, những người trong một những hướng dẫn của Trung Hoa Dân Quốc chỉ rõ rằng “sự hài hoà, harmony) là một trong những nguyên tắc Phật giáo thuộc về Đạo đức AI (Hongladarom và Bandasak, 2023).

Tất cả những điều này là những ví dụ điển hình về sự đa dạng phong phú hoá của thế giới, tất cả những điều này xứng đáng được ca ngợi, tuyên dương. Tuy nhiên, dường như trong các hướng dẫn này có một số chắc chắn tính không ổn định trong cả điểm tương đồng và điểm khác biệt.

Những bất ổn này bắt nguồn từ thực tế là chúng thiếu lý thuyết. Rõ ràng là các hướng dẫn này không dành cho các nhà triết học chuyên nghiệp; họ không phải chỉ rõ cơ sở lý luận hoặc lập luận phức tạp đằng sau các tuyên bố chuẩn mực của mình. Nhưng nếu họ không có lý thuyết, người ta sẽ không hiểu tại sao ngay từ đầu một số nguyên tắc nhất định lại có nơi đó.

Thực ra hầu hết những người soạn thảo các hướng dẫn có thể họ đều không phải là triết gia - nhưng thực tế là một số nguyên tắc đã tìm được ngõ vào hầu hết các hướng dẫn dường như cho thấy rằng ít nhất một số loại lý thuyết nào đó còn đang trong tâm trí họ, mặc dù có thể họ không giải thích rõ ràng. Tôi tin rằng nhiệm vụ của một triết gia là nêu rõ điều này.



Tuy nhiên, có nhiều lý thuyết đạo đức trong truyền thống phương Tây, chẳng hạn như triết học đạo đức của Immanuel Kant và nguyên tắc Hạnh phúc cực đại là nền tảng của luân lý (Thuyết vị lợi) và cũng có truyền thống phẩm chất tốt và đức hạnh. Việc lựa chọn các lý thuyết này đã đủ khó khăn rồi.

Nhưng vấn đề khó khăn hơn là ở chỗ đây là những truyền thống phương Tây - bất kể người ta chọn lý thuyết nào, vấn đề là làm thế nào để sử dụng lý thuyết đó trong các bối cảnh văn hoá khác nhau, chẳng hạn như ở Châu Á và các khu vực không phải phương Tây khác.

Vấn đề với triết học đạo đức của Immanuel Kant là mang tính hình thức và trừu tượng. Điều này gây ra vấn đề khi ai đó cố gắng gieo rắc nó trên các mặt đất Châu Á, nơi có nhiều truyền thống đạo đức trái ngược với tính hình thức và trừu tượng. Nguyên tắc Hạnh phúc cực đại là nền tảng của luân lý (Thuyết vị lợi) nhấn mạnh chuyên biệt vào niềm an vui hoặc “tiện ích công cộng”, do đó nó thiếu các khía cạnh có ý nghĩa trong cuộc sống của con người, là rào cản tiến bước việc đạt được hạnh phúc thực sự.

Về mặt này, tôi muốn đề xuất rằng triết lý đạo đức Phật giáo làm tốt hơn như lý thuyết về đạo đức của AI, ít nhất là ở Châu Á. Đối với truyền thống phẩm chất tốt và đức hạnh của phương Tây, nó giống với triết học Phật giáo hơn là triết học đạo đức của Immanuel Kant hoặc Nguyên tắc Hạnh phúc cực đại là nền tảng của luân lý (Thuyết vị lợi).

Nhưng sự khác biệt giữa Nguyên tắc Hạnh phúc cực đại là nền tảng của luân lý (Thuyết vị lợi) của phương Tây và triết lý đạo đức Phật giáo bởi triết lý đạo đức Phật giáo có lý thuyết siêu hình phát triển tốt, chẳng hạn như Triết lý “Tính không” trong triết học Phật giáo và thuyết Vô Ngã (Anatta) trong Đạo Phật, dường như điều này tạo ra một chương trình tu dưỡng đạo đức rõ ràng hơn.

Tác giả: **Soraj Hongladarom**

Viết dịch: **Thích Vân Phong**

Nguồn: 