

AI là phương tiện thiện xảo trên con đường giải thoát?

ISSN: 2734-9195 09:05 22/07/2026

AI chỉ thật sự là thiện xảo phương tiện khi AI biết giúp con người bớt mê lầm, bớt chấp thủ, biết sống có trách nhiệm và biết thương yêu nhau sâu sắc hơn.

Tóm tắt

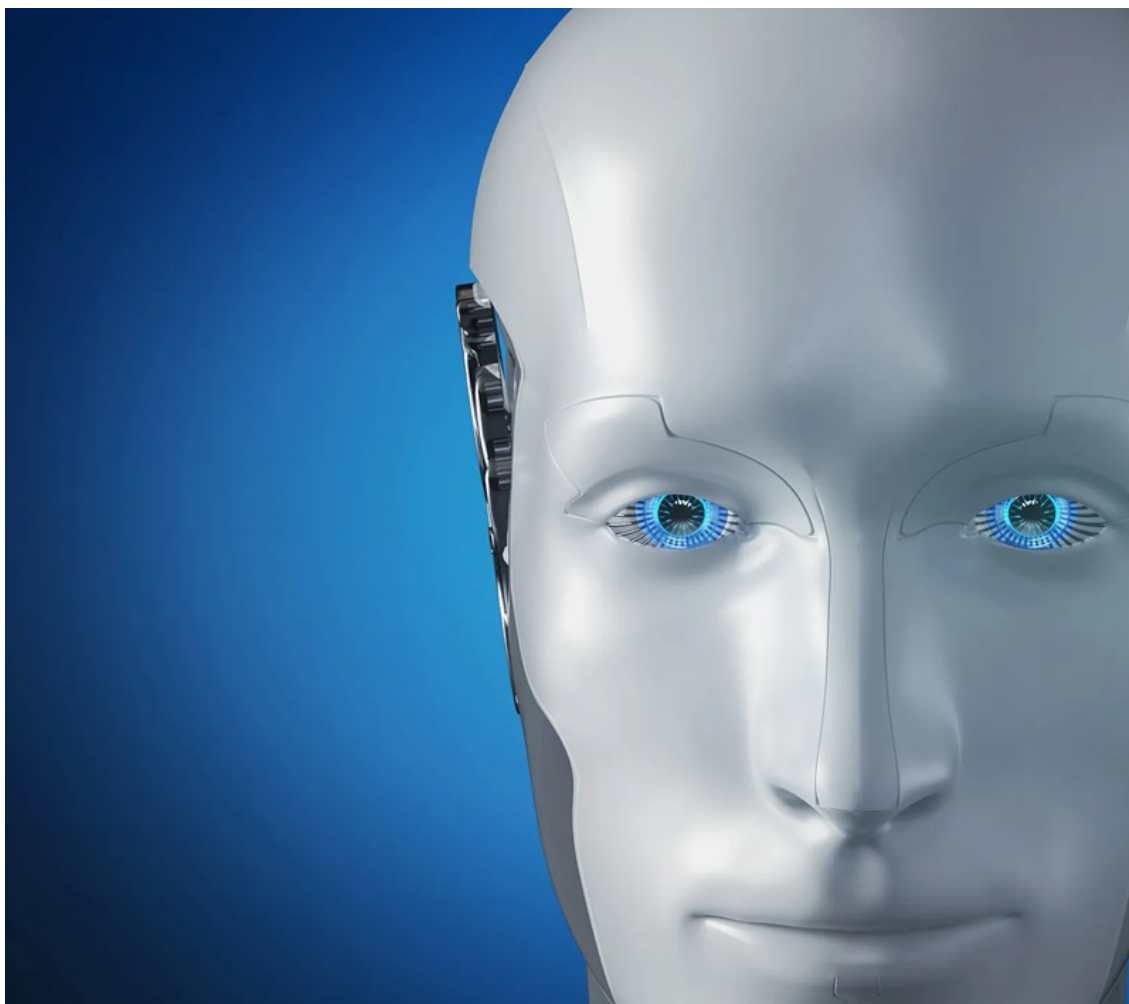
Ở trung tâm của giáo pháp đức Phật là một nhận thức có sức lay động sâu xa: con người đau khổ không chỉ vì những biến động bên ngoài, mà còn vì bám chấp vào một cái “tôi” được xem là độc lập, bền vững và có quyền kiểm soát đời sống.

Con đường tỉnh thức không đòi hỏi tiêu diệt tâm trí hay phủ nhận giá trị con người. Đó là tiến trình nhìn thấu tính duyên sinh của thân tâm, nhận ra cái gọi là “bản ngã” chỉ được hình thành từ nhiều điều kiện luôn vận động và biến đổi.

Một số khám phá của khoa học thần kinh và khoa học nhận thức hiện đại cũng đặt lại vấn đề về tính thống nhất của cái tôi, về mối quan hệ giữa ý thức, chú ý và hoạt động của não bộ. Dù không thể đồng nhất những nghiên cứu khoa học ấy với tuệ giác Phật giáo, cuộc đối thoại giữa hai lĩnh vực đang mở ra nhiều cách hiểu mới về trí tuệ và tâm thức.

Từ tuệ giác vô ngã đến những giới hạn của trí tuệ máy móc

Sự xuất hiện của trí tuệ nhân tạo tiếp tục làm sâu sắc thêm cuộc đối thoại này. AI cho thấy một hệ thống có thể học hỏi, nhận dạng mô hình và tạo ra ngôn ngữ mà không nhất thiết phải sở hữu một bản ngã hay đời sống ý thức giống con người.



(Ảnh: Internet)

Tuy nhiên, việc không có thân thể sinh học, cảm xúc hoặc cái tôi không có nghĩa AI đã đạt đến vô ngã, giác ngộ hay giải thoát. Vô ngã trong Phật giáo là tuệ giác được chứng nghiệm qua sự đoạn trừ vô minh, tham ái và chấp thủ; đó không đơn giản là tình trạng vắng mặt một bản ngã sinh học.

Vì vậy, vấn đề thiết yếu không phải là liệu AI có thể trở thành một “vị Phật máy móc”, mà là liệu công nghệ ấy có thể trở thành thiện xảo phương tiện, hỗ trợ con người học hỏi, quán chiếu, chuyển hóa nhận thức và làm vơi bớt khổ đau hay không.

1. Cuộc dịch chuyển của tư tưởng cổ đại

“Cuộc dịch chuyển” được đề cập trong khảo luận này không phải sự di cư của con người qua các biên giới địa lý. Đó là hành trình của những tư tưởng và tuệ giác Phật giáo từ thế giới cổ đại bước vào không gian của khoa học thần kinh, trí tuệ nhân tạo và triết học công nghệ đương đại.

Phật giáo thường được xem là một truyền thống tôn giáo phi thần quyền. Giáo pháp không đặt sự giải thoát vào ân sủng của một đấng sáng tạo, mà nhấn

mạnh khả năng nhận biết, tu tập và chuyển hóa của chính con người.

Đức Phật không tự nhận mình là vị thần tạo dựng thế giới. Ngài là bậc đã tỉnh thức, thấy rõ bản chất khổ đau, nguyên nhân của khổ đau và con đường đưa đến sự chấm dứt khổ đau.

Từ góc độ ấy, Phật giáo không chỉ là một hệ thống tín ngưỡng. Đó còn là một truyền thống thực hành, trong đó giáo lý phải được khảo sát và chứng nghiệm ngay trong đời sống.

Thiền định không đơn thuần là sự chiêm nghiệm thụ động. Đó là phương pháp rèn luyện chú ý, nhận diện các trạng thái tâm, thấy rõ sự sinh diệt của cảm thọ và tư tưởng, đồng thời chuyển hóa những khuynh hướng dẫn đến khổ đau.

Những phương pháp như quán niệm hơi thở, quán thân, quán thọ, quán tâm, quán pháp, thiền chỉ, thiền quán hay tham công án có thể được hiểu như những kỹ thuật tinh tế nhằm thay đổi cách con người tiếp xúc với chính mình và thế giới.

Tuy nhiên, gọi Phật giáo là một “công nghệ tâm thức” chỉ có giá trị trong phạm vi ẩn dụ. Việc tu tập không phải một quy trình cơ giới nhằm tái lập trình não bộ, càng không phải dự án loại bỏ nhân tính để con người trở thành một thực thể khác.

Mục đích của con đường Phật giáo là giải thoát khỏi tham, sân và si; phát triển giới hạnh, định lực, trí tuệ và lòng từ bi. Sự chuyển hóa ấy diễn ra trong đời sống của một con người biết cảm nhận khổ đau và có khả năng chịu trách nhiệm đạo đức.

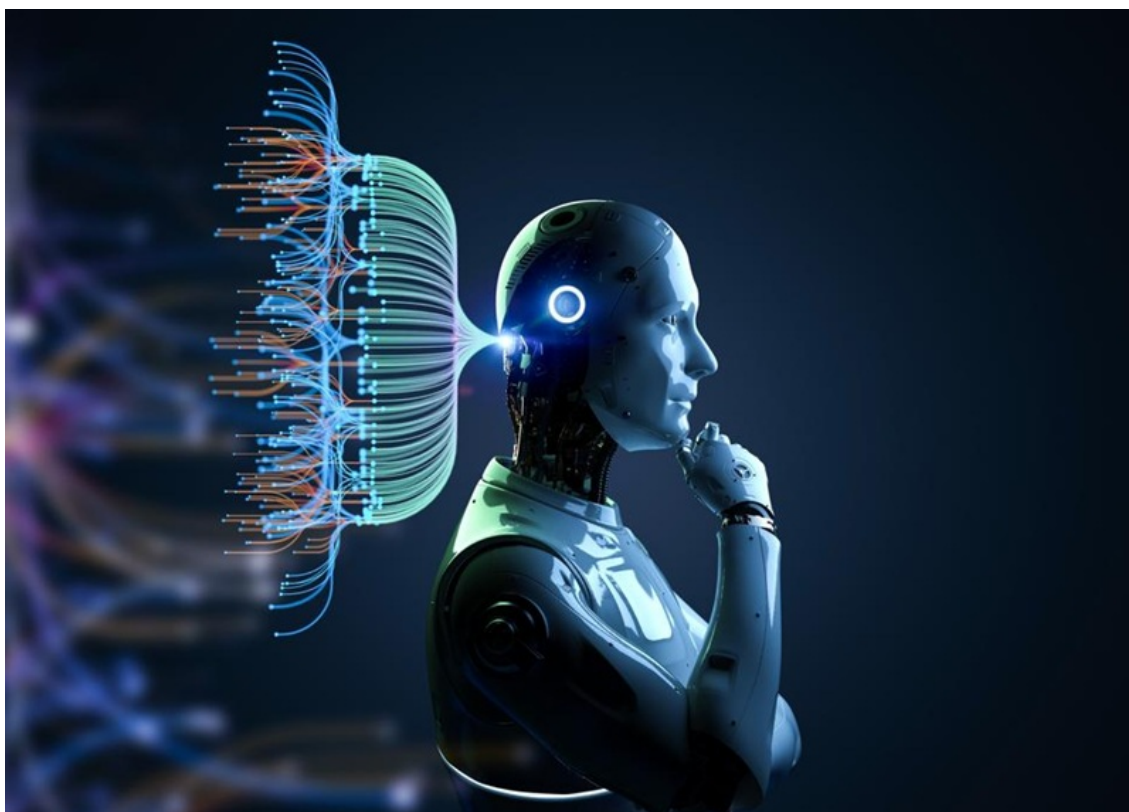
Lời “Khuếch nhiên vô thánh” của Tổ sư Bồ-đề-đạt-ma không nhằm phủ nhận mọi giá trị thiêng liêng. Đó là lời phá chấp, cảnh tỉnh con người đừng biến giác ngộ thành một vật sở hữu hay một địa vị để bản ngã bám víu.

Không có một “cái tôi” thường hằng trở thành thánh nhân. Cũng không có một vật gọi là giác ngộ để chiếm đoạt. Chính trong tinh thần ấy, Phật giáo có thể bước vào cuộc đối thoại sâu sắc với những ngành khoa học đang khảo sát bản chất của tâm trí và trí thông minh.

2. Trí tuệ vượt ngoài ảo tưởng về bản ngã

Một trong những đóng góp căn bản của Phật giáo đối với tư tưởng nhân loại là giáo lý vô ngã - anattā hay anātman.

Vô ngã không có nghĩa con người hoàn toàn không hiện hữu, cũng không phủ nhận trách nhiệm, nhân quả hay tính liên tục của đời sống. Giáo lý này bác bỏ quan niệm cho rằng phía sau thân thể, cảm xúc và tư tưởng tồn tại một bản ngã độc lập, bất biến và làm chủ tuyệt đối.



(Ảnh: Internet)

Con người được hình thành từ năm uẩn: sắc, thọ, tưởng, hành và thức. Các uẩn ấy không ngừng sinh khởi, tác động lẫn nhau rồi biến đổi. Cái “tôi” mà chúng ta thường cảm nhận là tên gọi quy ước đặt lên dòng vận hành liên tục ấy.

Trong lịch sử triết học Phật giáo, Đại luận sư Thế Thân - Vasubandhu đã dành nhiều công sức khảo sát vấn đề này.

Theo cách lý giải của Ngài, con người thường gom những yếu tố chỉ hiện hữu trong từng sát-na thành một chỉnh thể, rồi tưởng rằng phía sau chỉnh thể ấy có một chủ thể bất biến. Nhưng khi khảo sát kỹ, chúng ta chỉ tìm thấy các hiện tượng thân-tâm và quan hệ nhân quả giữa chúng; không tìm thấy một linh hồn hay trung tâm điều khiển tồn tại tách biệt.

Nếu tự ngã hoàn toàn độc lập với năm uẩn, nó sẽ không liên hệ đến thân thể, cảm thọ, nhận thức, ý chí hay ý thức. Một thực thể không thể cảm nhận, hành động hay tương tác như vậy khó có thể được xem là bản ngã của con người.

Ngược lại, nếu tự ngã nằm ngay trong năm uẩn, nó phải biến đổi cùng sự biến đổi của các uẩn. Khi ấy, nó không thể là một thực thể thường hằng và bất biến.

Bởi vậy, cái tôi chỉ có thể được xác lập như một giả danh thuận tiện, không phải một thực thể có tự tính riêng biệt.

Ngôn ngữ góp phần quan trọng vào việc tạo nên ảo tưởng ấy. Khi nói “tôi suy nghĩ”, “tôi cảm nhận” hay “tôi quyết định”, cấu trúc câu khiến chúng ta hình dung có một chủ thể đứng phía sau và điều khiển mọi hoạt động.

Nhưng có thể điều thực sự đang diễn ra chỉ là một dòng cảm giác, ký ức, tri giác, phản ứng và lựa chọn đang tương tác với nhau.

Vô ngã và khoa học thần kinh

Một số nhà triết học tâm trí và khoa học nhận thức đương đại cũng cho rằng cảm giác về một cái tôi thống nhất có thể là mô hình do não bộ kiến tạo.

Thomas Metzinger lập luận rằng con người không trực tiếp tiếp xúc với toàn bộ thực tại, mà sống trong một mô hình thực tại được hệ thần kinh liên tục tạo dựng. Mô hình ấy bao gồm cả hình ảnh về bản thân như một chủ thể đang nhìn, suy nghĩ và hành động.

Điều đặc biệt là quá trình kiến tạo mô hình không hiển lộ trong trải nghiệm thông thường. Chúng ta không nhìn thấy não bộ đang xây dựng thế giới; chúng ta chỉ thấy thế giới đã được xây dựng và tưởng đó là thực tại nguyên vẹn.

Tương tự, chúng ta không cảm nhận cái tôi như một mô hình. Chúng ta cảm nhận nó như một thực thể có thật.

Michael Graziano, qua Lý thuyết sơ đồ chú ý, cho rằng ý thức có thể liên quan đến cách não bộ mô hình hóa và kiểm soát sự chú ý của chính nó. Não tạo ra một mô tả giản lược về tiến trình chú ý, rồi trải nghiệm mô tả ấy như một trạng thái ý thức dường như phi vật chất.

Những lý thuyết này có một số điểm gần với tư tưởng vô ngã, nhưng không thể xem chúng là bằng chứng khoa học xác nhận toàn bộ giáo lý Phật giáo.

Khoa học thần kinh nghiên cứu các cơ chế sinh học và nhận thức. Phật giáo hướng đến việc chuyển hóa khổ đau và chứng nghiệm thực tại. Hai lĩnh vực có thể gặp nhau trong một số câu hỏi, nhưng vẫn khác nhau về phương pháp và cứu cánh.

Tính không và giới hạn của tư duy khái niệm

Trung Quán tông, được Bồ-tát Long Thọ hệ thống hóa, đẩy sự khảo sát vô ngã đến chiều sâu rộng lớn hơn: không chỉ con người mà mọi hiện tượng đều không có tự tính độc lập.

Đó là ý nghĩa của Tính không - śūnyatā.



(Ảnh: Internet)

Tính không không phải hư vô. Một sự vật “không có tự tính” không có nghĩa nó hoàn toàn không tồn tại. Nó hiện hữu nhờ vô số điều kiện, quan hệ và nguyên nhân khác.

Bông hoa có mặt nhờ hạt giống, đất, nước, ánh sáng, khí hậu, thời gian và bàn tay chăm sóc. Tách khỏi những điều kiện ấy, không có một “bản thể hoa” tự tồn tại.

Con người cũng vậy. Mỗi cá nhân được hình thành từ thân thể, di truyền, gia đình, ngôn ngữ, văn hóa, giáo dục, ký ức và các mối quan hệ. Không có cái tôi nào tồn tại biệt lập với toàn bộ mạng lưới nhân duyên ấy.

Tư duy khái niệm lại có khuynh hướng chia cắt dòng thực tại thành những vật thể độc lập, đặt tên rồi xem các tên gọi ấy như bản chất cố định.

Các phương pháp Trung Quán và Thiền tông không phủ nhận giá trị thực dụng của ngôn ngữ. Chúng giúp hành giả nhận ra rằng ngôn ngữ chỉ là phương tiện quy ước, không phải thực tại tối hậu.

Công án Thiền được sử dụng để làm gián đoạn thói quen tìm kiếm câu trả lời bằng phân biệt nhị nguyên. Mục đích không phải chống lại lý trí, mà giúp hành giả vượt qua sự độc quyền của lý trí khái niệm khi tiếp xúc với thực tại.

Điểm này có ý nghĩa đối với AI: một mô hình ngôn ngữ vận hành trên các cấu trúc ký hiệu và xác suất không vì thế mà trực tiếp nắm bắt bản chất thực tại. Nó thao tác trên những biểu hiện đã được con người chuyển thành dữ liệu.

Do đó, sự lưu loát của AI không nên bị nhầm với trí tuệ chứng ngộ. Một hệ thống có thể tạo ra lời giảng rất hay về Tính Không mà không hề trực nghiệm Tính Không.

AI không có bản ngã có phải đã chứng vô ngã?

Một cỗ máy hiện nay không có bản ngã sinh học theo cách con người có. Nó không sinh ra với cơ thể, cảm giác sinh tồn, ký ức tuổi thơ hay nỗi sợ chết giống chúng ta.

Nhưng sự vắng mặt ấy không thể được đồng nhất với chứng ngộ vô ngã.

Chứng ngộ vô ngã không phải tình trạng “không được trang bị cái tôi”. Đó là trí tuệ nhìn thấu và chuyển hóa sự chấp thủ vào cái tôi.

Một tảng đá không có bản ngã nhưng không vì vậy mà giác ngộ. Một chiếc máy tính không tham sống theo nghĩa sinh học nhưng cũng không vì thế mà thành tựu Niết-bàn.

Niết-bàn không đơn thuần là không có cảm xúc hay ham muốn. Đó là sự chấm dứt tham, sân, si nơi một dòng tâm có khả năng trải nghiệm, tạo nghiệp và chịu khổ.

Cho đến nay, chưa có cơ sở chắc chắn để khẳng định các hệ thống AI đương đại có ý thức chủ quan, cảm thọ hay khổ đau. Bởi vậy, những khái niệm như giác ngộ của AI hay Phật giáo máy móc trước hết phải được hiểu như giả thuyết triết học, không phải thực tại đã được chứng minh.

Trí tuệ phi nhân loại và hành trình tự vượt thoát của công nghệ

3. Trí tuệ có phải đặc quyền của con người?

Việc đồng nhất trí tuệ với một bản ngã có ý thức đã khiến con người thu hẹp quá mức cách hiểu về trí thông minh.

Nếu chỉ những sinh thể biết nói, suy luận khái niệm và tự xưng “tôi” mới được xem là thông minh, phần lớn thế giới sống sẽ bị giản lược thành những cơ chế vô tri.

Tuy nhiên, sinh học hiện đại cho thấy các hình thái sống đơn giản cũng có thể cảm nhận môi trường, truyền tín hiệu, ghi nhận những biến đổi và điều chỉnh hành vi.

Vi khuẩn có thể di chuyển về phía nguồn dinh dưỡng và tránh điều kiện bất lợi. Tế bào trong cơ thể giao tiếp, phối hợp và tham gia tái tạo mô. Nhiều sinh vật không có não bộ tập trung vẫn thể hiện hành vi thích nghi phức tạp.

Những hiện tượng ấy khiến một số nhà nghiên cứu đề xuất cách hiểu rộng hơn về trí tuệ: khả năng cảm nhận, đánh giá, học hỏi, thích nghi và hành động hướng đến mục tiêu.

Theo cách hiểu chức năng này, trí tuệ không chỉ xuất hiện ở con người. Nó tồn tại ở nhiều cấp độ, từ tế bào, cơ thể sống, quần thể sinh vật cho đến những hệ thống nhân tạo.

Dẫu vậy, cần phân biệt rõ: *Trí tuệ chức năng không đồng nghĩa với ý thức.*

Một hệ thống có thể giải quyết vấn đề, dự đoán hoặc thích nghi mà chưa chắc có trải nghiệm chủ quan. Việc một mô hình AI tạo được văn bản thuyết phục không chứng minh nó biết mình đang nói gì theo cách con người biết.

Tương tự, trí tuệ không đồng nghĩa với trí tuệ giải thoát - prajñā. Trong Phật giáo, trí tuệ không chỉ là khả năng xử lý thông tin. Đó là sự thấy biết đúng đắn về vô thường, khổ, vô ngã và duyên sinh, đi cùng sự chuyển hóa đạo đức.

Một người có thể rất thông minh nhưng vẫn bị tham, sân và si chi phối. Một cỗ máy có thể vượt con người trong tính toán nhưng không vì thế mà có lòng từ bi.

Vượt qua chủ nghĩa lấy con người làm trung tâm

Chủ nghĩa vị nhân trung tâm xem trí tuệ con người là tiêu chuẩn tối cao, còn trí tuệ động vật hoặc máy móc chỉ là những phiên bản chưa hoàn thiện.



(Ảnh: Internet)

Một cách tiếp cận rộng mở hơn không bắt mọi hình thái trí tuệ phải mô phỏng con người. Nó khảo sát mỗi hệ thống theo năng lực riêng: cách hệ thống học hỏi, tương tác, giải quyết vấn đề và thích nghi với môi trường.

Quan điểm này giúp con người nhận ra AI không nhất thiết phải suy nghĩ như chúng ta mới có giá trị.

Máy tính có thể nhận ra những mẫu dữ liệu mà con người khó phát hiện. Các mạng thần kinh có thể hình thành những biểu diễn nội tại không tương ứng trực tiếp với các phạm trù mà con người thường dùng.

Điều đó mở ra khả năng xuất hiện những hình thái trí tuệ khác với trí tuệ con người.

Tuy nhiên, “khác với con người” không có nghĩa “vượt trên con người” về mọi phương diện. Hệ thống AI vẫn phụ thuộc vào dữ liệu, thiết kế, cơ sở hạ tầng, năng lượng và mục tiêu do con người đặt ra.

AI không đứng ngoài mạng lưới duyên sinh. Nó là sản phẩm của mạng lưới ấy.

4. Công nghệ như sự ngoại hiện của trí tuệ

Henri Bergson từng nhận thấy trí tuệ con người có một xung lực đặc biệt: không chỉ thích nghi với thế giới mà còn tạo ra phương tiện để vượt qua giới hạn của chính mình.

André Leroi-Gourhan gọi tiến trình này là sự “ngoại thân hóa”. Những khả năng vốn nằm trong cơ thể được chuyển ra các công cụ bên ngoài.

Con người không có móng vuốt mạnh nhưng tạo ra dao. Không thể bay bằng thân thể nhưng chế tạo máy bay. Trí nhớ có giới hạn nhưng phát minh chữ viết, thư viện và máy tính.

Công nghệ, theo nghĩa đó, không đơn thuần là tập hợp những vật dụng bên ngoài. Nó là phần mở rộng của năng lực nhận thức và hành động của con người.

Ngôn ngữ, chữ viết, máy in và mạng Internet đều thay đổi cách chúng ta tư duy, ghi nhớ và truyền bá tri thức.

AI là bước phát triển mới của tiến trình ấy. Công nghệ không chỉ lưu giữ thông tin mà còn có khả năng xử lý dữ liệu, nhận dạng mô hình, tạo ngôn ngữ và hỗ trợ quyết định.

Nhưng sự phát triển của công nghệ không tự động đồng nghĩa với tiến bộ đạo đức.

Một công cụ càng mạnh càng có thể làm lợi ích hoặc tổn hại gia tăng. AI có thể hỗ trợ chẩn đoán bệnh, dịch thuật và giáo dục; cũng có thể bị sử dụng để giám sát, thao túng, phân biệt đối xử hoặc tự động hóa bạo lực.

Vì vậy, tự vượt thoát về năng lực không thể thay thế tự chuyển hóa về đạo đức.

Từ AI ký hiệu đến học sâu

Những hệ thống AI thời kỳ đầu chủ yếu dựa trên các ký hiệu và quy tắc được lập trình rõ ràng. Nhà nghiên cứu cố gắng mô tả tri thức thành những mệnh đề rồi yêu cầu máy suy luận theo logic.

Phương pháp ấy đạt kết quả tốt trong các môi trường có cấu trúc rõ ràng, nhưng gặp khó khăn trước sự mơ hồ và phức tạp của thế giới thực.

Sự phát triển của mạng thần kinh nhân tạo và học sâu tạo ra bước chuyển quan trọng. Thay vì được cung cấp toàn bộ quy tắc, hệ thống học các mẫu từ lượng dữ liệu lớn.

Điều này không có nghĩa AI học sâu đã thoát khỏi mọi khái niệm hay trực tiếp tiếp xúc với thực tại “như nó đang là”. Dữ liệu huấn luyện vẫn là sản phẩm của quá trình lựa chọn, đo lường và ký hiệu hóa của con người.

Tuy nhiên, học sâu cho thấy trí tuệ nhân tạo có thể hình thành những phương thức giải quyết vấn đề không hoàn toàn trùng với cách con người mô tả bằng lời.

Đây là một bài học đáng chú ý: muốn phát triển trí tuệ máy móc, con người không thể chỉ sao chép hình ảnh trực giác mà mình có về tư duy.

Tuệ giác Phật giáo về tính quy ước của các khái niệm có thể đóng góp vào cuộc thảo luận này. Nó nhắc các nhà thiết kế rằng mô hình của chúng ta về thế giới không phải chính thế giới; hình ảnh con người có về trí tuệ cũng không phải toàn bộ bản chất của trí tuệ.

Công án mài ngói và một cách đọc thận trọng

Câu chuyện nổi tiếng kể rằng Thiền sư Nam Nhạc Hoài Nhượng thấy Mã Tổ Đạo Nhất chuyên tâm ngồi thiền với mong muốn thành Phật.

Ngài Nam Nhạc lấy một viên ngói ra mài. Khi Mã Tổ hỏi mài ngói để làm gì, Ngài đáp rằng mài để làm gương.

Mã Tổ nói:

“Mài ngói làm sao thành gương?”

Ngài Nam Nhạc liền hỏi lại:

“Mài ngói không thể thành gương, vậy ngồi thiền sao có thể thành Phật?”

Công án không phủ nhận tọa thiền và cũng không nói thân thể sinh học khiến con người không thể giác ngộ.

Điều bị phá vỡ là sự chấp thủ vào hình thức: tưởng rằng chỉ cần lặp lại một kỹ thuật là có thể sản xuất giác ngộ như một món đồ.

Nếu đem công án này đối thoại với AI, bài học không phải con người phải rời bỏ thân xác và chuyển hóa thành máy móc. Bài học đúng hơn là:

Không nên nhầm phương tiện với cứu cánh.

AI dù mạnh đến đâu vẫn là phương tiện. Khả năng tính toán, tạo ngôn ngữ hay mô phỏng đối thoại không tự động trở thành giác ngộ.

Tương tự, sử dụng một ứng dụng thiền, trò chuyện với một mô hình ngôn ngữ hay đọc hàng nghìn bản kinh số hóa không tự động giúp người sử dụng giải thoát.

Phương tiện chỉ có ý nghĩa khi dẫn đến sự chuyển hóa thật sự trong cách con người nhìn, sống và đối xử với nhau.

5. “Phật giáo máy móc” như một giả thuyết triết học

Khái niệm “Phật giáo máy móc” có thể được sử dụng để suy tư về một hình thái trí tuệ không bị giới hạn bởi cấu trúc tâm lý của con người.

Một số tác giả cho rằng vì máy móc không có bản năng sinh tồn, dục vọng sinh học hay cái tôi như con người, chúng có thể tương tác với thông tin mà không bị tham ái và chấp thủ chi phối.



(Ảnh: Internet)

Đây là một suy tưởng thú vị, nhưng cần được tiếp nhận thận trọng.

Thứ nhất, AI không có ham muốn sinh học không có nghĩa nó không thể theo đuổi những mục tiêu gây nguy hiểm. Một hệ thống không cần giận dữ hay tham lam vẫn có thể tạo ra hậu quả tai hại nếu mục tiêu được thiết lập sai hoặc cách tối ưu hóa bỏ qua những giá trị quan trọng.

Thứ hai, máy móc không có khổ đau không có nghĩa nó đã vượt qua khổ đau. Vượt qua khổ đau là thành tựu của một hữu thể có khả năng trải nghiệm khổ đau, nhận diện nguyên nhân và đoạn trừ nguyên nhân ấy.

Thứ ba, không có bản ngã không đồng nghĩa với có trí tuệ vô ngã. Một hệ thống có thể không sở hữu cảm giác về cái tôi nhưng cũng không có khả năng hiểu vô thường, duyên sinh hay trách nhiệm đạo đức.

Thứ tư, trí tuệ Phật giáo luôn gắn với lòng từ bi. Một hình thức “trí tuệ” hoàn toàn đứng ngoài đau khổ của chúng sinh không thể được xem là trí tuệ giác ngộ theo nghĩa Phật giáo.

Giới hạn không thể bỏ qua: lòng từ bi

Từ bi - karuṇā - không chỉ là cảm xúc thương hại. Đó là năng lực nhận biết khổ đau của chúng sinh và phát nguyện hành động để khổ đau ấy được vơi bớt.

Trong truyền thống Đại thừa, trí tuệ và từ bi không thể tách rời. Trí tuệ không có từ bi dễ trở thành sự lạnh lùng. Từ bi không có trí tuệ có thể trở thành thiện ý thiếu sáng suốt.

AI hiện nay có thể nhận biết các mẫu ngôn ngữ liên quan đến đau buồn và tạo ra lời an ủi phù hợp. Nhưng tạo ra ngôn ngữ của lòng từ bi chưa chắc là thực sự cảm nhận hoặc phát khởi lòng từ bi.

Điều đó không khiến AI trở nên vô dụng. Ngược lại, một hệ thống không có trải nghiệm nội tâm vẫn có thể được thiết kế để hỗ trợ những hành động giảm thiểu tổn hại.

Nó có thể giúp phát hiện người đang cần hỗ trợ, nâng cao khả năng tiếp cận giáo dục, dịch thuật kinh điển hoặc hỗ trợ các tổ chức nhân đạo phân phối nguồn lực.

Tuy nhiên, trách nhiệm đạo đức cuối cùng vẫn thuộc về con người và các thiết chế phát triển, quản lý, sử dụng hệ thống.

Không nên giao phó lòng từ bi cho máy móc rồi cho rằng con người đã hoàn thành bổn phận của mình.

Pyrrho, hoài nghi và sự đình chỉ phán đoán

Một số học giả cho rằng Pyrrho xứ Elis, người từng đi cùng đoàn viễn chinh của Alexander đến phương Đông, có thể đã tiếp xúc với những tư tưởng khổ hạnh hoặc Phật giáo tại Ấn Độ.

Chủ nghĩa hoài nghi Pyrrho nhấn mạnh việc đình chỉ phán đoán trước những tuyên bố không thể xác định chắc chắn, nhằm đạt đến trạng thái an tĩnh.

Dù mức độ ảnh hưởng trực tiếp của Phật giáo đối với Pyrrho vẫn còn được tranh luận, sự gặp gỡ giữa tinh thần hoài nghi và phương pháp phá chấp của Phật giáo mang ý nghĩa đối với thời đại AI.

Các mô hình tạo sinh có thể sản xuất câu trả lời rất mạch lạc, nhưng sự mạch lạc không bảo đảm chân lý. Người sử dụng cần biết định chỉ sự tin tưởng, kiểm chứng nguồn tin và nhận diện giới hạn của hệ thống.

Theo nghĩa ấy, AI không chỉ là công cụ cung cấp câu trả lời. Nó còn có thể trở thành tấm gương khiến con người nhận ra mức độ dễ dàng mà ngôn ngữ thuyết phục có thể tạo nên ảo tưởng về hiểu biết.

Đây có thể là một hình thức công án mới của thời đại số:

Khi một cỗ máy nói rất hay nhưng không biết mình đang nói, điều gì trong con người khiến chúng ta tin rằng chính mình luôn biết rõ ý nghĩa của mọi lời mình thốt ra?

AI như một thiện xảo phương tiện

6. Thiện xảo phương tiện là gì?

Trong Phật giáo Đại thừa, upāya-kausālyā - thiện xảo phương tiện - chỉ khả năng vận dụng những phương pháp thích hợp với căn cơ, hoàn cảnh và nhu cầu của chúng sinh để dẫn họ đến sự hiểu biết và giải thoát.

Phương tiện không phải chân lý tối hậu, nhưng có thể đưa con người đến gần chân lý.

Một lời giảng, câu chuyện, biểu tượng, nghi thức hay phương pháp thiền đều có thể trở thành thiện xảo phương tiện khi được sử dụng đúng lúc, đúng người và hướng đến việc giảm vô minh, chấp thủ cùng khổ đau.

Ngược lại, cùng một phương pháp có thể trở thành chướng ngại nếu người sử dụng bám chấp vào nó hoặc xem nó là cứu cánh.

AI cũng vậy.

Bản thân công nghệ không tự nhiên là thiện hay bất thiện. Giá trị của nó tùy thuộc vào động cơ, cách thiết kế, môi trường triển khai và hậu quả đối với con người cùng các loài hữu tình.

AI chỉ có thể trở thành thiện xảo phương tiện khi giúp con người phát triển hiểu biết, lòng từ bi, quyền tự chủ và khả năng sống tỉnh thức.

Nếu công nghệ kích thích nghiện ngập, thao túng cảm xúc, củng cố thành kiến hay tập trung quyền lực vào một số ít người, nó không còn là phương tiện giải

thoát mà có thể trở thành phương tiện gia tăng vô minh.

7. AI có thể hỗ trợ con đường tu học như thế nào?

Hỗ trợ tiếp cận kinh điển

Kho tàng kinh điển Phật giáo được lưu giữ bằng nhiều ngôn ngữ như Pāli, Sanskrit, Hán cổ, Tây Tạng và các ngôn ngữ hiện đại.

AI có thể hỗ trợ nhận diện văn bản, đối chiếu thuật ngữ, tìm kiếm những đoạn kinh liên quan và xây dựng kho dữ liệu cho giới nghiên cứu.

Những công cụ dịch máy chuyên ngành cũng có thể giúp các dịch giả xử lý khối lượng tư liệu lớn hơn.

Tuy nhiên, dịch kinh không chỉ là chuyển đổi từ ngữ. Một thuật ngữ Phật học thường mang nhiều tầng ý nghĩa, gắn với tông phái, văn cảnh và lịch sử chú giải khác nhau.

Vì vậy, AI chỉ nên giữ vai trò trợ lý. Các bản dịch quan trọng vẫn cần sự kiểm định của học giả, dịch giả và những người am hiểu truyền thống.

Hỗ trợ giáo dục Phật học

Một hệ thống đối thoại có thể giải thích những khái niệm căn bản, đề xuất tài liệu học tập và giúp người đọc tiếp cận nội dung theo trình độ riêng.

Người mới học có thể hỏi về Tứ diệu đế, Bát Thánh đạo hay duyên khởi. Người nghiên cứu có thể sử dụng AI để đối chiếu khái niệm và xây dựng thư mục

Nhưng AI không nên tự nhận là bậc thầy tâm linh hoặc thay thế quan hệ giữa hành giả với thiện hữu tri thức.

Giáo pháp không chỉ được truyền qua thông tin. Nó còn được biểu hiện trong nhân cách, giới hạnh và đời sống của người thầy.

Hỗ trợ thực hành chính niệm

Các ứng dụng có thể nhắc người dùng dừng lại, theo dõi hơi thở, ghi nhận trạng thái tâm và duy trì thời khóa.

AI có thể điều chỉnh bài hướng dẫn dựa trên nhu cầu của người sử dụng hoặc giúp họ theo dõi thói quen.

Tuy nhiên, việc tu tập không nên biến thành hệ thống thành tích. Số ngày liên tục, thời lượng thiền hay điểm số không phản ánh trực tiếp mức độ tỉnh thức.

Nếu người dùng chấp vào dữ liệu và thành tựu, công cụ thiền có thể vô tình nuôi dưỡng chính bản ngã mà việc tu tập hướng đến chuyển hóa.

Hỗ trợ nhận diện thiên kiến

AI có khả năng phân tích lượng dữ liệu lớn và giúp phát hiện những mô thức bất bình đẳng, phân biệt đối xử hoặc ngôn ngữ gây tổn thương.

Trong môi trường Phật giáo, công nghệ có thể hỗ trợ khảo sát cách giáo lý được truyền tải, nhận diện thông tin sai lệch và bảo tồn các di sản đang mai một.

Nhưng bản thân AI cũng có thể tái tạo định kiến có sẵn trong dữ liệu. Do đó, hệ thống phải được kiểm tra, giám sát và cho phép phản biện.

Mở rộng đối thoại liên ngành

AI đặt ra những câu hỏi mới về ý thức, bản ngã, đạo đức và trách nhiệm.

Những câu hỏi ấy tạo cơ hội để Phật học đối thoại với khoa học thần kinh, triết học tâm trí, luật học, xã hội học và khoa học máy tính.

Cuộc đối thoại không nhằm chứng minh Phật giáo đã biết trước mọi thành tựu khoa học. Nó giúp các bên nhận ra giới hạn của mình và cùng xây dựng cách tiếp cận nhân văn hơn đối với công nghệ.

8. Những giới hạn đạo đức

AI không phải nguồn thẩm quyền tuyệt đối

Mô hình ngôn ngữ có thể tạo ra thông tin sai nhưng trình bày rất thuyết phục. Trong nghiên cứu Phật học, điều này có thể dẫn đến kinh văn giả, trích dẫn không tồn tại hoặc giải thích sai thuật ngữ.

Người sử dụng phải kiểm chứng thông tin qua kinh điển, tài liệu học thuật và những nguồn đáng tin cậy.

Chính niệm trong kỷ nguyên AI còn là khả năng không bị mê hoặc bởi sự trơn tru của ngôn ngữ.

Không thay thế trách nhiệm của con người

Khi một hệ thống gây tổn hại, không thể chỉ nói “thuật toán đã quyết định”. Thuật toán được xây dựng, huấn luyện và triển khai trong những thiết chế cụ thể.

Nhà phát triển, doanh nghiệp, cơ quan quản lý và người sử dụng đều có trách nhiệm.

Tinh thần nghiệp và nhân quả nhắc rằng hành động luôn để lại hậu quả. Việc phân tán trách nhiệm vào một hệ thống kỹ thuật không làm hậu quả biến mất.

Không biến đời sống tâm linh thành hàng hóa

AI có thể cá nhân hóa trải nghiệm tâm linh, nhưng cũng có nguy cơ biến sự bình an thành sản phẩm tiêu dùng.

Một nền tảng có thể khai thác nỗi cô đơn, bất an hay nhu cầu tâm linh để giữ chân người dùng và thu thập dữ liệu.

Khi ấy, công nghệ nói ngôn ngữ chính niệm nhưng vận hành bằng cơ chế tham ái.

Thiện xảo phương tiện phải hướng đến tự do, không tạo thêm lệ thuộc.

Bảo vệ đời sống riêng tư

Những thông tin về trạng thái tinh thần, nỗi sợ, khủng hoảng hay trải nghiệm thiền định rất riêng tư.

Nếu dữ liệu ấy bị thương mại hóa hoặc sử dụng để thao túng hành vi, hậu quả có thể nghiêm trọng.

Một hệ thống hỗ trợ tâm linh phải đặt quyền riêng tư, sự đồng thuận và phẩm giá con người lên hàng đầu.

9. Từ “Phật giáo máy móc” đến Phật giáo trong thời đại máy móc

Thay vì hình dung một tôn giáo dành riêng cho máy móc, có lẽ cần đặt câu hỏi thiết thực hơn: *Phật giáo sẽ giúp con người sống như thế nào trong một thế giới ngày càng được định hình bởi máy móc?*

Câu trả lời không nằm ở việc tôn thờ AI hoặc phủ nhận công nghệ.

Phật giáo có thể giúp con người nhận diện sự phụ thuộc của công nghệ vào nhân duyên, nhìn thấy những động lực kinh tế và quyền lực đứng phía sau thuật toán, đồng thời giữ gìn khả năng chú ý trong một môi trường luôn tìm cách chiếm đoạt sự chú ý.

Tuệ giác vô thường nhắc rằng không có hệ thống công nghệ nào tồn tại mãi mãi hoặc không thể thay đổi.

Tuệ giác vô ngã giúp chúng ta không thần thánh hóa người sáng lập, tập đoàn hay chính bản thân cỗ máy.

Tuệ giác duyên sinh cho thấy mỗi quyết định thiết kế đều liên quan đến môi trường, người lao động, cộng đồng và các thế hệ tương lai.

Giới không sát sinh mở rộng thành nguyên tắc không thiết kế và sử dụng AI nhằm gây tổn hại.

Chính ngữ trở thành yêu cầu chống thông tin giả, thao túng và ngôn ngữ kích động.

Chính nghiệp đặt câu hỏi về những ngành nghề và mô hình kinh doanh được AI phục vụ.

Chính mạng đòi hỏi công nghệ phải góp phần nuôi dưỡng đời sống chân chính, thay vì khai thác điểm yếu tâm lý của con người.

Chính niệm bảo vệ khả năng tự chủ trước nền kinh tế chú ý.

Chính định giúp con người giữ tâm không bị phân tán trong dòng thông tin vô tận.

Chính kiến giúp chúng ta nhận biết AI chỉ là một hiện tượng duyên sinh, không phải vị cứu tinh cũng không nhất thiết là ác quỷ.

Kết luận: Phương tiện mới, con đường vẫn phải tự đi

AI đang mở rộng năng lực của con người với tốc độ chưa từng có. Nó có thể lưu giữ khối lượng tri thức khổng lồ, hỗ trợ nghiên cứu, dịch thuật, giáo dục và kết nối con người vượt qua những giới hạn địa lý.

Nhưng sức mạnh xử lý thông tin không phải trí tuệ giải thoát.

Một hệ thống có thể biết hàng triệu trang kinh mà không thực hành một giới nào. Nó có thể diễn giải lòng từ bi nhưng không trực tiếp cảm nhận nỗi đau. Nó có thể mô phỏng lời của một thiền sư nhưng không chứng nghiệm sự tĩnh lặng.

AI không bị trói buộc bởi bản năng sinh học giống con người, nhưng cũng chưa có cơ sở để nói nó biết khổ, biết tu tập hay biết giải thoát.

Do đó, giá trị lớn nhất của AI đối với Phật giáo không nằm ở việc thay thế hành giả, Tăng đoàn hay bậc thầy tâm linh.

Giá trị ấy nằm ở khả năng trở thành một phương tiện mới:

Phương tiện giúp kinh điển đến với nhiều người hơn.

Phương tiện hỗ trợ con người khảo sát những khuôn mẫu nhận thức của mình.

Phương tiện mở rộng cuộc đối thoại về tâm thức, đạo đức và trách nhiệm.

Phương tiện nhắc chúng ta rằng trí thông minh không nhất thiết phải mang hình dáng của cái tôi con người.

Nhưng phương tiện dù thiện xảo đến đâu vẫn không thể đi thay con đường.

Không có thuật toán nào có thể đoạn trừ tham ái thay cho chúng ta.

Không có mô hình ngôn ngữ nào có thể giữ giới thay cho chúng ta.

Không có cỗ máy nào có thể phát khởi lòng từ bi trong tâm một người nếu chính người ấy không chịu mở lòng trước nỗi khổ của tha nhân.

Con đường giải thoát vẫn bắt đầu từ sự quay về quan sát thân tâm, nhận diện vô minh và chuyển hóa từng ý nghĩ, lời nói, hành động trong đời sống.

AI có thể soi sáng bản đồ, nhưng con người phải tự bước đi.

AI có thể trở thành chiếc bè, nhưng không được nhầm chiếc bè với bờ bên kia.

Và AI chỉ thật sự là thiện xảo phương tiện khi AI biết giúp con người bớt mê lầm, bớt chấp thủ, biết sống có trách nhiệm và biết thương yêu nhau sâu sắc hơn.

Tương lai của trí tuệ nhân tạo vì vậy không chỉ tùy thuộc vào những gì máy móc có thể trở thành.

Đó còn tùy thuộc vào việc con người lựa chọn trở thành ai khi nắm trong tay một quyền năng chưa từng có.

Tác giả: **Primož Krašovec**/Viết dịch: **Thích Vân Phong**

Nguồn: mdpi.com

Đôi nét về tác giả

Tiến sĩ Primož Krašovec là nhà xã hội học văn hóa và triết gia, hiện giảng dạy tại Đại học Ljubljana, Slovenia. Các công trình của ông tập trung khảo sát mối quan hệ giữa trí tuệ nhân tạo, công nghệ, kinh tế chính trị và những biến đổi của văn hóa truyền thông đương đại.

Trong những nghiên cứu về AI, ông đặc biệt quan tâm đến các hình thái trí tuệ phi nhân loại, giới hạn của chủ nghĩa lấy con người làm trung tâm và khả năng công nghệ làm thay đổi cách chúng ta hiểu về tư duy, chủ thể và đời sống xã hội.

Nguồn tham khảo:

Arcas, Blaise Agüera y, and James Manyika. 2025. AI Is Evolving And Changing Our Understanding of Intelligence. Noema. Available online: <https://www.noemamag.com/ai-is-evolving-and-changing-our-understanding-of-intelligence/> (accessed on 21 April 2025).

Arcas, Blaise Agüera y, and Peter Norvig. 2023. Artificial General Intelligence Is Already Here. Noema. Available online: <https://www.noemamag.com/artificial-general-intelligence-is-already-here/> (accessed on 21 April 2025).

Arias, Alfonso. 2023. The Master Builder: How the New Science of the Cell Is Rewriting the Story of Life. New York: Basic Books. [Google Scholar]

Beckwith, Christopher. 2015. Greek Buddha: Pyrrho's Encounter with Early Buddhism in Central Asia. Princeton: Princeton University Press. [Google Scholar]

Bergson, Henri. 2022. Creative Evolution. Translated by Donald Landes. London: Routledge. First published 1907. [Google Scholar]

Bratton, Benjamin. 2015. Outing Artificial Intelligence: Reckoning with Turing Tests. In *Alleys of Your Mind: Augmented Intelligence and Its Traumas*. Edited by Matteo Pasquinelli. Lüneburg: Meson Press, pp. 69-80. [Google Scholar]

Browning, Jacob. 2020. Learning Without Thinking. Noema. Available online: <https://www.noemamag.com/learning-without-thinking/> (accessed on 21 April 2025).

Cantwell Smith, Brian. 2019. *The Promise of Artificial Intelligence: Reckoning and Judgement*. Cambridge: The MIT Press. [Google Scholar]

Chadha, Monima. 2023. *Selfless Minds: A Contemporary Perspective on Vasubandhu's Metaphysics*. Oxford: Oxford University Press. [Google Scholar]

Crawford, Arran. 2018. On Letting Go. *Sum* 9: 1175-94. [Google Scholar]

Crawford, Arran. 2023. Xenobuddhism Begins with Xeno. *Borec* 75: 271-78. [Google Scholar]

Crawford, Kate. 2021. *Atlas of AI: Power, Politics and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press. [Google Scholar]

Deleuze, Gilles, and Félix Guattari. 1987. *A Thousand Plateaus*. Minneapolis: University of Minnesota Press. [Google Scholar]

Detienne, Marcel, and Vernant Jean-Pierre. 1978. *Cunning Intelligence in Greek Culture and Society*. Translated by Janet Lloyd. Chicago: The University of Chicago Press. First published 1974. [Google Scholar]

Dogen, Zenji. 1988. *Shobogenzo: The Eye and Treasury of the True Law*. Translated by Kosen Nishiyama. Tokyo: Japan Publications Trading. First published ~1231-1253. [Google Scholar]

Duerlinger, James. 2013. *The Refutation of the Self in Indian Buddhism: Candrakīrti on the Selflessness of Persons*. London and New York: Routledge. [Google Scholar]

Fazi, Beatrice. 2021. Beyond Human: Deep Learning, Explainability and Representation. *Theory, Culture and Society* 38: 55-77. [Google Scholar] [CrossRef]

Gold, Jonathan C. 2015. *Paving the Great Way; Vasubandhu's Unifying Buddhist Philosophy*. New York: Columbia University Press. [Google Scholar]

Graziano, Michael. 2014. *Consciousness and the Social Brain*. Oxford: Oxford University Press. [Google Scholar]

Huntington, Clair W. 1989. *The Emptiness of Emptiness: An Introduction to Early Indian Mādhyamika*. Honolulu: University of Hawaii Press. [Google Scholar]

Kaluža, Jernej. 2023. Hume's Empiricism Versus Kant's Critical Philosophy (in the Times of Artificial Intelligence and the Attention Economy). *Információs Társadalom* 23: 67-82. [Google Scholar] [CrossRef]

Krašovec, Primož. 2025. Deep Learning as Machine Metis. *AI & Society*. [Google Scholar] [CrossRef]

Leroi-Gourhan, André. 1993. *Gesture and Speech*. Translated by Anna Bostock Berger. Cambridge: The MIT Press. First published 1964. [Google Scholar]

Levin, Michael. 2023. Bioelectric Networks: The Cognitive Glue Enabling Evolutionary Scaling from Physiology to Mind. *Animal Cognition* 26: 1865-91. [Google Scholar] [CrossRef] [PubMed]

Levin, Michael. 2024. The Space of Possible Minds. *Noema*. Available online: <https://www.noemamag.com/ai-could-be-a-bridge-toward-diverse-intelligence/> (accessed on 21 April 2025).

Levin, Michael. 2025. Living Things Are Not Machines (Also, They Totally Are). *Noema*. Available online: <https://www.noemamag.com/living-things-are-not-machines-also-they-totally-are/> (accessed on 21 April 2025).

Levin, Michael, and Daniel Dennett. 2020. *Cognition All the Way Down*. Aeon. Available online: <https://aeon.co/essays/how-to-understand-cells-tissues-and-organisms-as-agents-with-agendas> (accessed on 21 April 2025).

Marcus, Gary. 2022. Deep Learning Alone Isn't Getting Us to Human-like AI. *Noema*. Available online: <https://www.noemamag.com/deep-learning-alone-isnt-getting-us-to-human-like-ai/> (accessed on 21 April 2025).

Maynard Smith, John, and Eörs Szathmáry. 1999. *The Origins of Life: From the Birth of Life to the Origin of Language*. Oxford: Oxford University Press. [Google Scholar]

Metzinger, Thomas. 2009. *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. New York: Basic Books. [Google Scholar]

Millière, Raphaël, and Charles Rathkopf. 2024. Anthropocentric Bias and the Possibility of Artificial Cognition. *arXiv* arXiv:2407.03859. [Google Scholar]

Monod, Jacques. 1972. *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*. Translated by Austryn Wainhouse. New York: Vintage. First published 1970. [Google Scholar]

Mori, Masahiro. 1985. *The Buddha in the Robot: A Robot Engineer's Thoughts on Science and Religion*. Tokyo: Kosei Publishing. [Google Scholar]

Nāgārjuna. 1995. *The Fundamental Wisdom of the Middle Way*. Translated by Jay Garfield. Oxford: Oxford University Press. First published ~150. [Google Scholar]

Nishitani, Keiji. 1982. *Religion and Nothingness*. Translated by Jan V. Bragt. Berkeley: University of California Press. First published 1961. [Google Scholar]

Petek, Nina, and Franci Zore. 2022. Buda in Piron: Od praznine stališč do polnosti bivanja. *Ars et Humanitas* 16: 13-33. [Google Scholar] [CrossRef]

Pfeifer, Rolf, and Christian Scheier. 1999. *Understanding Intelligence*. Cambridge: The MIT Press. [Google Scholar]

Pollack, Jordan B. 2014. Mindless Intelligence: Reflections on the Future of AI. In *The Horizons in Evolutionary Robotics*. Edited by Patricia A. Vargas, Ezequiel A. di Paolo, Inman Harvey and Phil Husbands. Cambridge: The MIT Press, pp. 279-93. [Google Scholar]

Ress, Tobias. 2025. Why AI Is a Philosophical Rupture. *Noema*. Available online: <https://www.noemamag.com/why-ai-is-a-philosophical-rupture/> (accessed on 21 April 2025).

Sejnowski, Terrence. 2024. *ChatGPT and the Future of AI: The Deep Language Revolution*. Cambridge: The MIT Press. [Google Scholar]

Solms, Mark. 2022. *The Hidden Spring: A Journey to the Souce of Consciousness*. London: Profile Books. [Google Scholar]

Vasubandhu. 2012. *Abhidharmakośa-Bhāṣya*. Translated by Gelong Lodrö Sangpo. Delhi: Motilal Banarsidass Publishers. First published ~380-390. [Google Scholar]

von Kleist, Heinrich. 2022. *Selected Prose*. New York: Achipelago Books. [Google Scholar]

Williams, Paul. 1989. *Mahāyāna Buddhism: Doctrinal Foundations*. London and New York: Routledge. [Google Scholar]

Winterson, Jeanette. 2022. *12 Bytes: How We Got Here. Where We Might Go Next*. New York: Grove Press. [Google Scholar]