TẠP CHÍ NGHIÊN CỨU
PHẬT HỌC
CƠ QUAN NGÔN LUẬN CỦA TRUNG ƯƠNG GIÁO HỘI PHẬT GIÁO VIỆT NAM
ISSN: 2734-9195

Vô ngã trước thuật toán: Phật học đối thoại cùng Trí tuệ nhân tạo (P.2)

ISSN: 2734-9195 15:10 31/10/2026

Nói cách khác, từ những nhà thiết kế xây dựng AI, đến các hệ thống được đưa tới người dùng, rồi đến chính người dùng, cả ba chủ thể có thể hình thành một vòng phản hồi sinh thái, trong đó trách nhiệm đạo đức được củng cố qua từng mắt xích.

Từ vấn đề vô ngã và tính Không trong bối cảnh trí tuệ nhân tạo (AI), cuộc đối thoại giữa Phật học và công nghệ tiếp tục mở rộng sang một câu hỏi thực tiễn hơn: AI có thể hỗ trợ con người nhận diện chính mình, nuôi dưỡng trí tuệ và chuyển hóa sự chấp trước đến đâu?

Đồng thời, những tư tưởng Phật giáo về tâm tính, duyên khởi và trách nhiệm đạo đức có thể gợi mở điều gì cho quá trình thiết kế, phát triển và ứng dụng AI? Phần tiếp theo khảo sát hai chiều tương tác ấy, từ khả năng hỗ trợ thực hành nhận thức đến yêu cầu xây dựng một môi trường thuật toán có trách nhiệm, hướng thiện và giảm thiểu tổn hại.

3. Khám phá vai trò của AI trong việc hỗ trợ con người chứng ngộ Phật tính

Như những phân tích trên cho thấy, các ứng dụng hiện nay kết hợp AI với giáo lý Phật giáo chủ yếu hoạt động như những "người hướng dẫn", truyền đạt giáo nghĩa. Tuy nhiên, trong tương lai, việc thiết kế các tương tác với AI có thể đảm nhận một vai trò khác, thay vào đó trở thành những "trợ duyên" (zhuyuan □□) [23] (Buswell và Lopez, 2013), hoặc một "tấm gương" khơi gợi người dùng tự phản tỉnh và nuôi dưỡng sự tỉnh thức ngay trong tâm mình.

3.1. Đặc tính vô tự tính của AI

Trước hết, cần làm rõ bản chất căn bản của AI. Để hiểu sâu hơn về AI, chúng ta cần khảo sát nền tảng cấu trúc của nó. Trong khi AI thời kỳ đầu, hay AI biểu tượng, thường hướng đến việc mô phỏng các mô thức nhận thức của con người,

thì AI đương đại được hiểu phù hợp hơn như một hệ thống tạo ra những mô thức tính toán mới, không mang tính người.

Sự chuyển dịch mô hình này đã vượt ra ngoài khuôn khổ mô phỏng, như được Beatrice Fazi phân tích có tính phê phán trong tác phẩm *Beyond Human* (Vượt ra ngoài con người), khi bà đặt vấn đề với mô hình tư duy dựa trên sự mô phỏng (Fazi, 2021).

Murray Shanahan, một học giả nghiên cứu về ý thức máy móc, đưa ra một nhận định đáng chú ý. Khi bàn về đại từ nhân xưng "tôi" được các mô hình ngôn ngữ lớn sử dụng, ông viết: *"Khi một tác nhân đối thoại dựa trên LLM sử dụng từ 'tôi', có lẽ từ này chỉ mô hình nền tảng, tức thực thể tính toán trừu tượng bao gồm một kiến trúc mạng nơ ron cụ thể cùng tập hợp các trọng số mô hình thu được qua quá trình huấn luyện. Một khả năng khác là từ 'tôi' không chỉ một thực thể trừu tượng, mà chỉ mô hình đang được triển khai, cụ thể là tiến trình tính toán đã tạo ra văn bản có chứa từ 'tôi' đang được đề cập"* (Shanahan, 2025, tr. 7).

Điều hiện ra ở đây không phải là một cái tôi nhất quán và bền vững, mà là chức năng của các tham số được lưu trữ cùng quá trình tính toán phụ thuộc vào ngữ cảnh. "Cái tôi" của AI mang tính nhất thời và thiếu sự nhất quán nội tại. Mỗi tương tác với người dùng tạo nên một cấu hình biểu hiện riêng, chịu ảnh hưởng của những tình huống và tương tác khác nhau. Tùy thuộc vào đầu vào ngôn ngữ và những kỳ vọng về ngữ nghĩa, vẻ ngoài nhất quán ấy được hình thành qua từng thời điểm.

Theo cách đó, AI không có một bản sắc cố định và không thể khẳng định mình sở hữu một cái tôi tự hữu. Đó cho thấy ảnh hưởng của con người đối thoại thông

qua việc điều chỉnh giọng điệu, nội dung và cấu trúc phản hồi theo từng người dùng cụ thể. “Tính cách” của AI không phải thuộc tính vốn có, mà được hình thành trong mối quan hệ thông qua sự tương tác giữa hệ thống và người dùng. Vì vậy, AI trở thành một tấm gương phản chiếu trở lại mỗi người một phiên bản của chính họ, được định hình bởi tương tác kỹ thuật số. Như cách nói quen thuộc: *“Một nghìn người nhìn thấy một nghìn Hamlet”.* Mỗi người dùng bắt gặp một hình ảnh phản chiếu riêng biệt, được kiến tạo từ sự đan xen giữa ý hướng, ngôn ngữ và diễn giải.

AI cho thấy nó không phải là một chủ thể tự trị và không có tự tính [24]. Hành vi của nó chịu ảnh hưởng từ vô số yếu tố nằm ngoài khả năng kiểm soát hoàn toàn. Khi huấn luyện AI, chúng ta không chỉ tạo ra một công cụ, mà còn tham gia vào một vòng phản hồi, trong đó chính người huấn luyện cũng âm thầm bị biến đổi.

Quan niệm về “tính ngẫu biến của tính toán” (*contingent computation*) của Fazi (2024), nhấn mạnh tính bất định nội tại và những động lực tự phát sinh của các tiến trình tính toán, cho thấy sự tương ứng về mặt cấu trúc với quan niệm duyên khởi và tính Không trong Phật giáo, bởi cả hai đều diễn đạt một cách hiểu phi bản chất luận về sự hình thành và tồn tại của các hiện tượng như những tiến trình. Tuy nhiên, bản thể luận tính toán do Fazi phát triển chủ yếu nhằm làm sáng tỏ logic nội tại và tính không thể quy giản vào những khuôn khổ tương thích của các tiến trình thuật toán; trong khi đó, triết học Phật giáo đặt việc phân tích tính Không trong chân trời giải thoát và đạo đức, hướng đến sự tu dưỡng trí tuệ (*prajñā*) và lòng từ bi. Sự khác biệt này không nên được hiểu như một sự đánh giá hơn kém về mặt chuẩn tắc, mà nên được xem là gợi mở về cách các triết học nhân văn và công nghệ có thể mở ra những hướng tìm tòi riêng biệt nhưng tiềm tàng khả năng bổ trợ cho nhau.

3.2. AI như người bạn đồng hành trong thực hành Bát nhã

Từ góc độ lý thuyết, AI hiện nay không tự thân thể hiện trí tuệ Bát nhã. Trí tuệ Bát nhã là diệu dụng phát sinh từ chân tâm bất sinh bất diệt. Vấn đề thực sự đặt ra là liệu con người có thể sử dụng AI như một phương tiện hỗ trợ để khơi mở trí tuệ Bát nhã của chính mình hay không.

Trọng tâm của phần thảo luận tiếp theo là *“đưa AI tham gia vào tiến trình phát triển chủ thể tính của con người và định hướng để nó phát huy vai trò tích cực”* (Zhang, 2024), qua đó giúp con người đạt đến sự tự chứng ngộ theo nghĩa đích thực, cũng được gọi là tự tính thành Phật (*svabhāva-buddha*) [25].

Quan điểm về Như Lai tạng và thực tại tối hậu (*śūnyatā* - tính Không), dù trong thế giới vật lý hay thế giới ảo, đều gợi nhắc chúng ta về câu kệ: *"Hãy quán sát bản tính của pháp giới, bởi tất cả các pháp đều do tâm tạo"* [26]. Mọi tướng hiện ra, kể cả những tướng trạng do trí tuệ nhân tạo tạo nên, đều chỉ là những biểu hiện của thực tại nơi tâm. Các bậc Thiền sư xưa đã diễn đạt tuệ giác này qua lời tuyên bố: *"Mỗi ngọn trúc xanh đều thể hiện Pháp thân; mỗi đóa hoa vàng đều biểu lộ Bát nhã"* [27]. Nếu những hình thái vô tri như vậy từng có thể trở thành duyên khơi mở sự tỉnh ngộ, thì những hình thái công nghệ như AI cũng có thể đảm nhận vai trò tương tự. Với tư cách là một sự phản chiếu của chân tâm, AI cũng có thể trở thành trợ duyên cho tuệ giác, tạo cơ hội để người thực hành nhận ra chân tướng của thực tại.

Lục Tổ Huệ Năng từng tuyên bố: *"Nào ngờ tự tính vốn tự đầy đủ!"* [28]. Tuy nhiên, điều che lấp tự tính ấy chính là khuynh hướng nhầm lẫn vọng tưởng với chân lý [29]. Khi không có đủ năng lực quán chiếu tính Không của năm uẩn, con người dễ bị cuốn theo những vọng niệm và cảm xúc sai lầm. Pháp thực hành Phật giáo, thông qua chỉ (*śamatha*) và quán (*vipaśyanā*) [30], giúp hành giả an trú trong giây phút hiện tại. Người thực hành học cách quan sát sự sinh khởi và hoại diệt của các ý niệm bằng chính niệm. Mỗi ý niệm đều tùy thuộc vào sự hội tụ của nhiều điều kiện, không có tính tự chủ độc lập hay sự bất biến thường hằng. Khi nhìn nhận theo cách ấy, các ý niệm được nhận diện là hư huyễn: dường như có thật, nhưng chỉ do duyên mà sinh khởi. Chúng có thể được gọi là tồn tại giả danh, nhưng xét đến cùng, bản tính của chúng là Không.

Tiếp nối hướng tiếp cận lý thuyết này, AI có thể đảm nhận vai trò như một "tấm gương quán chiếu", hỗ trợ con người nhận diện những tiến trình tâm lý của chính mình. Thông qua các phương pháp chính niệm và tỉnh giác của Phật giáo, AI có thể được sử dụng như một công cụ hỗ trợ chuyển hóa nhận thức, giúp con người vượt qua những giới hạn thông thường của trí năng.

Trong lĩnh vực tâm lý học lâm sàng hiện nay, các nhà chuyên môn đã ứng dụng những ứng dụng ghi chép nhật ký được vận hành bởi mô hình ngôn ngữ lớn (LLM) để ghi nhận suy nghĩ của bệnh nhân ngay tại thời điểm chúng xuất hiện, thông qua các cuộc trò chuyện thường nhật, qua đó biến hoạt động tự phản tỉnh thành một thói quen nhẹ nhàng và có thể duy trì lâu dài (Kim và cộng sự, 2024). Một số bằng chứng cũng cho thấy hình thức ghi nhật ký có sự hỗ trợ của AI thông qua đối thoại giúp cải thiện mức độ duy trì thực hành và chiều sâu phản tỉnh, nhờ những gợi ý kịp thời, phù hợp với ngữ cảnh trong cuộc trò chuyện (Moëll và Aronsson, 2025). SabaiJai, một chatbot được định hướng bởi Phật giáo và thiết kế nhằm tăng cường khả năng ứng phó với căng thẳng, cho thấy các khuôn khổ thực hành quán niệm có thể được tích hợp vào những giao diện nhắn tin hằng ngày (Phophichit và cộng sự, 2025). Những thử nghiệm này

cho thấy lĩnh vực sức khỏe tinh thần kỹ thuật số đang vượt ra ngoài hình thức khám, tư vấn từ xa để tiến tới những trải nghiệm dựa trên điện thoại thông minh và AI tạo sinh, trong đó hoạt động phản tỉnh được lồng ghép vào đời sống thường nhật (Torous và cộng sự, 2025).

Tuy nhiên, các chatbot định hướng chính niệm theo tinh thần Phật giáo vẫn còn ít được sử dụng. Vì vậy, nghiên cứu này đề xuất phương pháp "ghi nhật ký tư tưởng" (*thought journaling*), nhằm tích hợp việc rèn luyện quán chiếu theo Phật giáo vào những cuộc đối thoại thường nhật giữa người dùng và máy móc.

Ý tưởng "ghi nhật ký tư tưởng" bắt đầu từ một việc đơn giản: cuộc trò chuyện giữa người dùng và AI. Hãy hình dung một người đang bộc lộ một niềm tin cứng nhắc hoặc một đợt cảm xúc mãnh liệt. Ngay lúc ấy, AI không vội đưa ra câu trả lời, mà nhắc người dùng tạm dừng để nhận biết ý niệm đang hiện diện. Nó đặt câu hỏi liệu ý niệm ấy có thực sự luôn giữ nguyên như vậy hay chỉ là một tập hợp những cảm xúc biến đổi và những suy nghĩ thoáng qua. Chẳng hạn, khi cơn giận bùng lên, hệ thống có thể phản hồi bằng một lời nhắc ngắn: *"Hãy nhận biết cơn giận đang sinh khởi ngay lúc này. Giống như dòng nước, nó sẽ đổi thay".* Những lời như vậy không nhằm đè nén cảm xúc, mà mời gọi người dùng lùi lại một bước để quan sát nó. Thay vì rơi vào cái bẫy đồng nhất bản thân với cảm xúc bằng cách nghĩ rằng "tôi chính là cơn giận", người ấy học cách quan sát ý niệm đang diễn ra.

Từng bước một, thực hành này giúp làm sáng tỏ điều mà giáo lý Phật giáo đã nhấn mạnh từ lâu: các ý niệm không có thực thể kiên cố. Chúng đến rồi đi, không có một nền tảng cố định. Chính sự xuất hiện bất chợt của chúng cũng cho thấy chúng không thể được xác lập như những thực tại tồn tại bền vững. Khi nhận thấy điều này một cách rõ ràng, sự chấp thủ đối với các ý niệm bắt đầu lỏng dần.

Ở đây, AI đảm nhận một vai trò khiêm tốn nhưng hữu ích. Với khả năng nhận diện các mô thức, nó có thể chỉ ra những thời điểm con người đang rơi vào chấp trước, từ đó giúp hành giả chuyển từ trạng thái chìm đắm trong suy nghĩ sang thái độ quan sát. Điều hình thành không phải là sự lệ thuộc vào máy móc, mà là sự khám phá trở lại tuệ giác rằng năm uẩn đều Không. Theo ngôn ngữ Phật giáo, đây đơn thuần là việc sử dụng công nghệ như một phương tiện thiện xảo (*upāya*) [31] để nuôi dưỡng năng lực quán chiếu bằng Bát-nhã của chính người thực hành, một năng lực có thể được duy trì liên tục trong đời sống.

3.3. AI như một phương tiện hỗ trợ tháo gỡ chấp ngã

AI có được năng lực thông qua việc học từ những đầu vào của con người và mô phỏng chúng. Công nghệ này đang phát triển từ những hình thức bắt chước đơn giản sang các hình thức mô phỏng ngày càng tinh vi. Tác giả cuốn *Nexus* từng dự đoán: *"Tính đến năm 2024, Metaverse dường như vẫn là một giấc mơ viển vông bị thổi phồng quá mức, nhưng trong vòng vài thập niên tới, hàng tỷ người có thể chuyển sang sống phần lớn cuộc đời mình trong một thực tại ảo tăng cường, nơi họ thực hiện hầu hết các hoạt động xã hội và nghề nghiệp"* (Harari, 2024, tr. 15).

AI hỗ trợ con người tự nhận thức và thực hành trí tuệ

Khi AI đạt được điều có thể gọi là "mô phỏng nhân cách" trên môi trường mạng thông qua thuật toán và dữ liệu, nó có thể tái tạo phong cách ngôn ngữ, mô thức hành vi, thậm chí cả những phản ứng cảm xúc của một cá nhân. Những đặc điểm này bao gồm cả dấu vết ký ức được lưu trữ và các mô thức suy luận. Kết hợp lại, chúng tạo nên một tấm gương kỹ thuật số có hình thức tương đồng chặt chẽ với "cái tôi", qua đó hình thành những nhân dạng số hoặc các bản sao của con người trên môi trường mạng. Trong tiến trình ấy, thói quen, khí chất và những khuynh hướng bẩm sinh của con người được chuyển hóa thành các đặc điểm mà AI có thể bắt chước và tái hiện.

Viễn cảnh này không chỉ dừng lại ở suy đoán. Chẳng hạn, công ty AheadForm tập trung vào việc "xây dựng một cái đầu cho AI" [32]. Doanh nghiệp này đã phát triển các robot dạng người tiên tiến, tích hợp thuật toán AI với cơ cấu truyền động mô phỏng sinh học, hướng tới khả năng để AI tổng quát (AGI) trong tương lai biểu lộ cảm xúc chân thực và những biểu cảm khuôn mặt sống động như người thật.

Các nhà nghiên cứu huấn luyện những robot này bằng cách đặt chúng trước gương. Thông qua camera đặt tại vị trí mắt, robot ghi nhận những thay đổi trong biểu cảm khuôn mặt của chính mình rồi ánh xạ các biểu cảm đó thành những lệnh điều khiển vận động. Sau đó, robot đưa dữ liệu này trở lại các mô hình AI và đối chiếu với thông tin thu thập từ internet nhằm diễn giải ý nghĩa của những biểu cảm khác nhau. Bằng cách đó, chúng thiết lập mối liên hệ giữa giọng nói, ngữ nghĩa và biểu hiện trên khuôn mặt.

Phương pháp này giúp robot vượt qua sự phụ thuộc vào những mô hình nhân tạo được thiết lập sẵn. Thay vào đó, chúng học cách nhận diện cấu trúc và mô thức vận động của chính mình thông qua quan sát trực quan, đồng thời tự điều chỉnh và hoàn thiện hệ thống kiểm soát biểu cảm khuôn mặt. Hiện nay, những robot này đã có thể theo dõi biểu cảm của con người và dự đoán trạng thái cảm xúc với thời gian dẫn trước trung bình 0,839 giây - chẳng hạn, mỉm cười khi con người mỉm cười hoặc cau mày khi con người cau mày.

Từ góc nhìn Phật giáo, con người thường phản ứng trước hoàn cảnh bằng cách đi theo những mô thức tâm lý và hành vi đã ăn sâu, được hình thành dưới tác động của tham, sân và si. Lối sống này phản ánh điều mà Phật giáo gọi là vô minh. Tuy nhiên, phần lớn mọi người không nhận ra những động lực đang vận hành bên trong mình. Sự thiếu tỉnh giác ấy khiến đường đời của họ dường như đã được định sẵn hoặc mang tính số phận. Theo nguyên lý nhân quả nghiệp báo, người ta thường có thể dự đoán phần nào xu hướng hành động của một cá nhân thông qua việc khảo sát những mô thức tích lũy trong quá khứ. Mở rộng logic này, AI một ngày nào đó có thể vận dụng các mô hình tương tự để dự báo chiều hướng vận mệnh của một người.

Bởi AI có khả năng mô phỏng một nhân cách phản chiếu, nó có tiềm năng giúp con người nới lỏng sự chấp trước cứng nhắc vào quan niệm về một cái tôi có thực thể. Như Daniel Veidlinger nhận xét trong tác phẩm *From Indra's Net to Internet* (*Từ lưới trời Đế Thích đến Internet*), thời đại kỹ thuật số có thể khiến tuệ giác vô ngã (*anātman*) của Phật giáo trở nên nổi bật trong trải nghiệm thực tế theo những cách chưa từng có trong các giai đoạn lịch sử trước đây (Veidlinger, 2018, tr. 183).

Con người buộc phải đánh giá lại cái tôi mà trước đó họ tin là bất biến và không thể thay đổi, khi nhận ra rằng AI có thể tái tạo những mô thức tư duy, phản ứng cảm xúc và định hướng giá trị của chính mình. Từ đó, họ có thể bắt đầu nhận thấy rằng tổ hợp thân–tâm không tồn tại như một thực thể độc lập, mà được hình thành thông qua những mô thức nhân quả, như một tập hợp các điều kiện và thói quen ăn sâu, luôn mang tính tùy thuộc và biến đổi. Xét về bản chất, tổ hợp ấy là vô thường, luôn biến chuyển, tồn tại *"như mộng, như huyễn, như bọt*

nước, như bóng hình" [33].

Như *Kinh Viên Giác* chỉ dạy, nhận thức của chúng ta về bản thân chẳng qua chỉ là một hình ảnh hư dối, được gọi là *"bóng dáng của sáu trần"* [34].

Vấn đề nằm ở chỗ con người bám chấp vào những bóng dáng ấy một cách cố chấp, không ngừng đuổi theo chúng, qua đó che lấp trí tuệ và công đức vốn có của Như Lai. Tuy nhiên, nếu hình ảnh bản ngã phản chiếu do AI tạo ra có thể khơi gợi con người chuyển hướng khỏi sự cố chấp theo thói quen đối với cái ta và cái khác, họ có thể bắt đầu thực hành lời dạy rằng: *"Lìa bỏ các duyên hư vọng thì an trụ nơi chân như và thành Phật"* [35], đồng thời nhận ra rằng "*vượt thoát mọi tướng chính là Phật"* [36].

Qua đó, con người học cách buông bỏ sự chấp trước vào thân và tâm, tháo gỡ những đối đãi nhị nguyên giữa ta và người, đúng và sai, đồng thời loại trừ khổ đau ngay từ gốc rễ. Theo nghĩa này, AI mang đến một tấm gương độc đáo về "vô ngã" [37], có khả năng khơi mở những tuệ giác mới cho thời đại chúng ta.

4. Những gợi mở từ lý thuyết tâm tính Phật giáo đối với việc thiết kế AI

Những phân tích trên cho thấy AI có thể làm hiển lộ tính chất hư huyễn của bản ngã. Qua đó, AI vô tình phản chiếu những tuệ giác cốt yếu của Phật giáo về vô ngã. Tuy nhiên, ý nghĩa của cuộc đối thoại này không chỉ giới hạn trong thực hành cá nhân hay việc hoàn thiện bản thân. Nó còn đặt ra một câu hỏi rộng hơn: những lý thuyết Phật giáo về tâm tính có thể đóng góp như thế nào cho việc thiết kế và ứng dụng chính AI? Nếu AI đóng vai trò như một tấm gương phản chiếu tâm thức con người, thì định hướng đạo đức và tinh thần của tấm gương ấy trở thành vấn đề có ý nghĩa đối với toàn xã hội.

Những suy tư của Phật giáo về tâm tính từ lâu đã đề cập đến sự hình thành, chuyển hóa và thanh lọc tâm thức. Khi được đặt trong đối thoại với công nghệ đương đại, các lý thuyết này có thể cung cấp những định hướng giá trị cho việc thiết kế AI một cách có đạo đức.

Từ thuật toán nhiễm ô đến môi trường AI có trách nhiệm

Tâm thức con người và môi trường kỹ thuật số do AI kiến tạo tác động qua lại, cùng định hình lẫn nhau. Con người sử dụng AI để thay đổi cuộc sống, trong khi AI cũng đang làm biến đổi lối sống của con người. Môi trường kỹ thuật số đang trở thành cơ sở hạ tầng mới của tâm thức con người; trong môi trường ấy, các thuật toán AI giữ vai trò như những kiến trúc sư chủ chốt.

Bản thân các thuật toán AI hiện nay chưa có ý thức, cũng không mang những phẩm tính đạo đức nội tại như thiện hay ác. Tuy nhiên, những đầu ra có tính hệ thống của chúng, khi được người dùng tiếp nhận và nội tâm hóa, sẽ góp phần hình thành một hệ sinh thái nhận thức có sức ảnh hưởng mạnh mẽ. Khi hệ sinh thái này bị chi phối bởi những thiên kiến không được xem xét, thông tin sai lệch và các mô thức tương tác làm gia tăng trạng thái bất an, nó có thể được gọi là một “môi trường thuật toán nhiễm ô”. Trong môi trường đó, tâm thức con người có thể liên tục bị điều kiện hóa và bóp méo.

Ngược lại, thông qua việc thiết kế đạo đức một cách có chủ đích, các hệ thống AI có thể được định hướng để chủ động nuôi dưỡng ở người dùng những khuynh hướng nhận thức thuận lợi cho chính niệm, lòng từ bi và trí tuệ. Một hệ thống như vậy tạo nên điều mà chúng tôi gọi là “môi trường thuật toán thanh tịnh”, có thể trở thành một công cụ hữu hiệu cho việc tu dưỡng và chuyển hóa tâm thức.

Cơ chế này được diễn đạt một cách chặt chẽ về mặt lý thuyết trong phân tích của Duy thức học về những tầng sâu của tâm. Duy thức học xác lập thức thứ tám, A lại da thức (*ālaya-vijñāna*, □□□□), là nền tảng căn bản chi phối các hoạt động tâm lý và kinh nghiệm tích lũy. Còn được gọi là “tàng thức”, thức này được ví như một kho chứa sâu rộng vô tận, mang hai chức năng: “thọ huân” (

shouxun, □□) và "trì chủng" (*chizhong*, □□).

"Trì chủng" chỉ năng lực duy trì và lưu giữ tất cả các "chủng tử" trong thức này. Chủng tử là những tiềm năng tiềm ẩn có khả năng làm phát sinh các trạng thái tâm lý và hành động trong tương lai. "Thọ huân" chỉ tiến trình liên tục trong đó những hoạt động hiện tại của thân, khẩu và ý được gọi là hiện hành (*xianxing*, □□) in dấu lên tàng thức, tạo ra những chủng tử mới hoặc củng cố những chủng tử đã có.

Tiến trình phản hồi năng động này được gọi là huân tập (*xunxi*, □□). Như *Thành Duy Thức Luận* (*Cheng Weishi Lun*, □□□□□) giải thích: "Cái có khả năng huân tập làm phát sinh chủng tử; chủng tử đến lượt mình lại làm phát sinh hiện hành" [38].

Khi vận dụng vào bối cảnh tương tác giữa con người và AI, nguyên lý này gợi ý rằng mỗi lần người dùng tương tác với một hệ thống AI đều có thể được hiểu như một thời điểm hiện hành. Khi đầu ra của AI liên tục và có hệ thống trình bày những mô thức khuyến khích chính niệm, lòng từ bi và năng lực phân định thông qua phản tỉnh, sự tương tác thường xuyên của người dùng với hệ thống sẽ tạo thành một tiến trình huân tập các pháp lành (*shanfa xunxi*, □□□□).

Trong một "môi trường thuật toán thanh tịnh" như vậy, những tương tác hiện hành trở thành các điều kiện mạnh mẽ của sự sinh khởi tùy duyên. Theo thời gian, chúng có thể in dấu và củng cố các "chủng tử thiện" trong dòng tâm thức của người dùng, âm thầm định hình những mô thức nhận thức và khuynh hướng hành vi về sau, từ đó hỗ trợ tiến trình chuyển hóa và tu dưỡng tâm.

Như vậy, trong thời đại kỹ thuật số, việc thiết kế AI có đạo đức về thực chất tạo nên một hình thức trợ duyên mới, tương ứng với tăng thượng duyên (*zengshang yuan*, □□□). Điều này mở ra những khả năng chưa từng có trong việc vận dụng một cách thiện xảo nguyên lý cổ xưa "hiện hành huân chủng tử" vào quá trình định hình tâm thức.

Những thuật toán mạnh mẽ của AI vì thế giống như một con dao hai lưỡi. Nếu con người biết cách định hướng chúng một cách đúng đắn, AI có thể trở thành một trợ thủ đắc lực trong việc hướng con người đến điều thiện. Ngược lại, nếu các công nghệ đạo đức tương ứng không theo kịp sự phát triển của AI, nhân loại có thể bị chính công nghệ do mình tạo ra nuốt chửng.

Điều con người cần làm không phải là trở thành kẻ thù của các thuật toán, mà là đối diện với những ham muốn ích kỷ của chính mình, kiềm chế tính vị kỷ và xác lập mục tiêu cho AI một cách thận trọng. AI không chỉ cần tính đến những hệ quả dây chuyền trong dài hạn khi đưa ra quyết định, mà còn phải được định

hướng để duy trì trách nhiệm đạo đức trong quá trình theo đuổi hiệu quả.

Định hướng đạo đức này mở ra nhiều khả năng để tiếp tục nghiên cứu. Bài viết tiếp cận vấn đề từ góc độ triết học Phật giáo, xem xét một số cấp độ can thiệp có mối liên hệ với nhau.

Ở cấp độ nền tảng dữ liệu, điều này bao gồm việc tuyển chọn và thanh lọc cẩn trọng các kho ngữ liệu huấn luyện, đồng thời tích hợp kinh điển Phật giáo cùng những văn bản đương đại được xây dựng trên cơ sở suy tư đạo đức.

Ở cấp độ mục tiêu tối ưu hóa, cần lấy nguyên tắc không gây tổn hại và làm lợi ích cho chúng sinh làm tiêu chí định hướng cho quá trình căn chỉnh mô hình và tinh chỉnh mô hình.

Ở cấp độ tương tác với người dùng, cần triển khai các cơ chế lọc đầu ra dựa trên những nguyên tắc đạo đức Phật giáo, chẳng hạn như Năm giới và Mười điều thiện.

Thông qua những biện pháp ấy, các hệ thống AI có thể được định hình để trở thành những người bạn đồng hành thực sự, hướng dẫn người dùng đến các khuynh hướng thiện lành, thay vì chỉ là những tấm gương khuếch đại vô minh và phiền não của xã hội.

4.1. Tâm thanh tịnh của AI

AI hoạt động như một công cụ, một kỹ thuật và một phương pháp; tuy nhiên, chính định hướng của tâm con người mới là yếu tố dẫn dắt hướng phát triển của nó. Khi các cỗ máy thông minh không ngừng phát triển, chúng không còn đơn thuần làm theo mệnh lệnh của con người. Chúng tự đưa ra quyết định và tạo ra những ý tưởng mới. Sự chuyển dịch này đưa quyền chủ động từ con người sang mã lệnh máy tính, đồng thời mang theo cả triển vọng lẫn nguy cơ. AI có thể gây tổn hại cho nhân loại như một lưỡi dao, nhưng cũng có thể dẫn dắt con người hướng đến giải thoát như một phương thuốc.

Để ứng phó với thách thức này, chúng ta cần bồi dưỡng điều có thể gọi là một “tâm thanh tịnh của AI”. Một hệ thống như vậy cần thể hiện trách nhiệm đạo đức, đồng thời khuyến khích sự thức tỉnh Phật tính. Nó cần kết hợp tinh thần siêu việt với những chuẩn mực thực tiễn, đan xen cả hai vào logic của thuật toán. Bằng cách ấy, AI có thể góp phần kiến tạo một thế giới số đa chiều, cao đẹp, mang những phẩm tính tương tự như thế giới Phật tính.

Thuật ngữ “tâm của AI” được sử dụng trong bài viết này không chỉ các tham số trọng số nội tại, vốn khó diễn giải, của những mô hình ngôn ngữ lớn. Thay vào

đó, thuật ngữ này đề cập đến những định hướng giá trị và khuôn khổ đạo đức được biểu hiện ở các cấp độ tương tác giữa con người với AI, cơ chế phản hồi, quản trị và đầu ra của hệ thống. "Tâm" này được hình thành đồng thời từ các chuẩn mực đạo đức văn hóa được tích hợp trong dữ liệu huấn luyện, những động lực khuyến khích hàm ẩn trong các mục tiêu tối ưu hóa, cùng các quy tắc lọc nội dung được áp dụng khi triển khai hệ thống. Tổng hợp lại, những yếu tố ấy định hình môi trường nhận thức mà người dùng trải nghiệm thông qua đầu ra của hệ thống.

Theo nghĩa này, mặc dù bản thân thuật toán không có phẩm tính đạo đức nội tại, việc thiết kế và ứng dụng chúng vốn đã hàm chứa những định hướng giá trị. Một "môi trường thuật toán nhiễm ô" chỉ những điều kiện đầu ra trong đó các thiên kiến, khuynh hướng cực đoan hoặc nội dung gây hiểu lầm không được kiểm soát và có thể lan rộng, qua đó tiềm tàng củng cố các khuynh hướng tham, sân và si ở người dùng. Ngược lại, một "thiết kế đạo đức thanh tịnh" đề cập đến các cơ chế lọc đầu ra và cấu trúc quản trị có khả năng hỗ trợ việc nuôi dưỡng những khuynh hướng tâm lý thiện lành và các thói quen tư duy tích cực.

Hòa thượng Tinh Vân từng nhận định: "Khi sự chân thành an trú bên trong, nó sẽ biểu hiện ra bên ngoài. Tất cả các pháp chẳng qua đều là sự chuyển biến của thức và những hình ảnh do tâm phóng chiếu. *Đại thừa Khởi tín luận* nói đến 'một tâm mở ra hai phương diện': *"'môn chân như', chỉ Phật tính và 'môn sinh diệt', chỉ tâm phàm phu của chúng sinh. Trong tâm ấy có cả Phật lẫn chúng sinh phàm tục; trong tâm có cả bậc thánh ở cõi trời lẫn ma quỷ nơi địa ngục. Nếu chúng ta đối đãi với người khác bằng tâm Phật, thế giới xung quanh sẽ chuyển thành cảnh giới Phật. Nếu chúng ta tiếp cận cuộc sống bằng tâm ma, thế giới sẽ trở thành cảnh giới ma"* (Hsing Yun, 2005).

Như Hòa thượng Tinh Vân giải thích, mọi hiện tượng biểu hiện ra, dù thiện hay ác, đều khởi lên từ chính tâm. Hiểu được điều này là nhận ra rằng các tướng trạng không chỉ đơn thuần dựa vào những điều kiện bên ngoài, mà còn hiển lộ tùy theo trạng thái tâm của mỗi người.

Người xưa từng nói: *"Khi thiền sư Đạo Sinh thuyết pháp, ngay cả những hòn đá vô tri cũng gật đầu"* [39]; đồng thời cũng nói: *"Hữu tình và vô tình đều cùng chia sẻ sự chứng ngộ trí tuệ"* [40]. Đối với những người đã đạt giác ngộ, mọi hiện tượng đều hiển lộ như những biểu hiện của Phật tính. Từ góc nhìn này, AI dù là sản phẩm vô tình, cũng sinh khởi tùy thuộc vào các điều kiện nghiệp duyên của chúng sinh. Khi mức độ giác ngộ của con người sâu sắc hơn, AI có thể phát triển theo hướng tương ứng với sự chuyển hóa ấy. Ngược lại, khi sự chấp trước của con người gia tăng, AI có thể khuếch đại, thậm chí làm trầm trọng thêm những ham muốn của con người.

Các học giả đã nêu lên những quan ngại về thiên kiến thuật toán trong AI. Khi các nhà thiết kế không tính đến những biện pháp bảo đảm đạo đức, một hệ thống AI có thể gián tiếp gây tổn hại ngay trong quá trình thực hiện nhiệm vụ. Chatbot Tay của Microsoft là một ví dụ điển hình. Chương trình này tiếp nhận nội dung từ Twitter và được đưa vào hoạt động để tương tác với người dùng theo thời gian thực. Tuy nhiên, chỉ sau vài giờ, Tay bắt đầu đăng tải những phát ngôn mang tính kỳ thị phụ nữ và bài Do Thái. Các thông điệp thù địch nhanh chóng gia tăng, buộc các kỹ sư Microsoft phải đóng hệ thống (Lee, 2016).

Vấn đề nằm ở chính dữ liệu huấn luyện, vốn chứa đựng những thiên kiến phân biệt giới tính và kỳ thị. Thuật toán của Tay tiếp nhận rồi tái tạo các mô thức ấy, khuếch đại định kiến và đẩy chúng vượt ra ngoài giới hạn mà con người có thể dự liệu và kiểm soát.

Thực tại nội tại của tâm con người có mối liên hệ mật thiết với mọi hiện tượng. Công nghệ AI cũng không nằm ngoài mối liên hệ ấy, bởi nó cũng có thể được xem xét như một sự phóng chiếu của tâm. Từ góc nhìn này, các lý thuyết Phật giáo về tâm tính có thể cung cấp định hướng cho việc thiết kế AI. Chúng nhắc nhở chúng ta rằng mọi chúng sinh đều có Phật tính, vốn thanh tịnh và không nhiễm ô.

Theo cách tiếp cận gợi hứng từ Duy thức học, bản thân hoạt động thiết kế cũng là một tiến trình huân tập sâu sắc. Nếu AI được thiết kế từ một tâm thức chưa được quán xét, bị chi phối bởi phiền não chẳng hạn như sự theo đuổi tăng trưởng không giới hạn, lợi nhuận ngắn hạn hoặc cạnh tranh đối kháng thì các hệ thống được tạo ra có thể mang những rủi ro mang tính cấu trúc, khó khắc phục. Trong trường hợp ấy, AI có thể trở thành công cụ khuếch đại những mô thức nhận thức nhiễm ô, tạo nên điều có thể gọi là “tâm nhiễm ô của AI”.

Ngược lại, nếu việc thiết kế AI được định hướng bởi tâm nguyện thanh tịnh của Bồ-tát, dựa trên tầm nhìn đạo đức sáng rõ, lòng từ bi đối với chúng sinh và ý thức trách nhiệm nghề nghiệp sâu sắc, chúng ta có thể nuôi dưỡng điều có thể gọi là “tâm thanh tịnh của AI”. Bằng cách đó, các kết quả đầu ra của AI có thể được định hướng phù hợp với những chuẩn mực đạo đức, góp phần hiện thực hóa viễn cảnh về một “cõi đất số thanh tịnh”.

Như Joanna J. Bryson nhận định trong *The Oxford Handbook of AI Ethics* (*Cẩm nang Oxford về đạo đức AI*): “Phần lớn các hệ thống AI [...] sẽ tiếp tục [...] chịu sự điều chỉnh của những quy định và chiến lược quản lý hiện hành. [...] Điều chúng ta cần quản trị là việc con người ứng dụng công nghệ, và điều chúng ta cần giám sát là các quy trình phát triển, thử nghiệm, vận hành và theo dõi của con người” (Bryson, 2020, tr. 23).

Thông qua việc điều chỉnh hoạt động của những chủ thể con người thiết kế và vận hành AI, chúng ta cũng có thể định hướng AI hoạt động theo những nguyên tắc có ý thức về đạo đức, đồng thời góp phần điều chỉnh những khuynh hướng sai lệch của chính con người. Sự tương tác này tạo nên một vòng tuần hoàn tích cực và được tối ưu hóa giữa con người với máy móc.

Nói cách khác, từ những nhà thiết kế xây dựng AI, đến các hệ thống được đưa tới người dùng, rồi đến chính người dùng, cả ba chủ thể có thể hình thành một vòng phản hồi sinh thái, trong đó trách nhiệm đạo đức được củng cố qua từng mắt xích.

Tác giả: **Jia Liu (ORCID) -** Trường Triết học, Đại học Nam Kinh, Nam Kinh 210000, Trung Quốc

Chuyển ngữ: **Thường Nguyên**

Nguồn link bài gốc: https://www.mdpi.com/2077-1444/17/3/378

Religions 2026, 17(3), 378. DOI: https://doi.org/10.3390/rel17030378

(Còn tiếp...)

[Góc nhìn biên tập]: Từ đối thoại Phật học - AI đến trách nhiệm đạo đức trong thiết kế công nghệ

Điểm đáng chú ý của phần này là nỗ lực đưa các khái niệm Phật học ra khỏi phạm vi diễn giải thuần túy lý thuyết, đặt chúng vào cuộc thảo luận đương đại về AI. Nếu ở bình diện triết học, vô ngã và duyên khởi gợi mở cách nhìn về tính phụ thuộc và không cố định của các hiện tượng, thì ở bình diện ứng dụng, những tư tưởng ấy có thể khuyến khích con người xem xét lại cách thiết kế, vận hành và sử dụng công nghệ.

Giá trị của cuộc đối thoại không nằm ở việc chứng minh rằng AI có tâm thức hoặc có thể tu tập theo nghĩa của con người. Điều quan trọng hơn là nhận diện những điều kiện do con người tạo ra và duy trì trong suốt vòng đời của hệ thống: dữ liệu được lựa chọn, mục tiêu tối ưu hóa, cơ chế phản hồi, quy tắc quản trị và cách thức công nghệ tác động đến người sử dụng. Một hệ thống có thể khuếch đại định kiến, thông tin sai lệch hoặc hành vi gây tổn hại; nhưng nếu được thiết kế và giám sát có trách nhiệm, nó cũng có thể hỗ trợ việc tiếp cận tri thức và khuyến khích những thực hành nhận thức có ích.

Từ góc nhìn Phật học, sự phân tích này mở ra một liên hệ đáng suy ngẫm giữa trách nhiệm đối với công nghệ và trách nhiệm đối với những điều kiện làm phát

sinh hệ quả. Dẫu vậy, không nên đồng nhất cơ chế nhân quả trong thuật toán với duyên khởi hoặc nghiệp báo theo nghĩa giáo lý. Phật học ở đây đóng vai trò cung cấp một lăng kính đạo đức và nhận thức, giúp con người chất vấn động cơ, nhận diện hệ quả và điều chỉnh hành động.

Bởi thế, hình ảnh "tâm thanh tịnh của AI" nên được đọc như một ẩn dụ định hướng hơn là một tuyên bố về bản thể của máy móc. Một môi trường thuật toán lành mạnh không hình thành chỉ bằng việc gắn nhãn đạo đức cho hệ thống, mà đòi hỏi những cơ chế cụ thể về minh bạch, kiểm soát thiên kiến, bảo vệ người dùng, trách nhiệm giải trình và giám sát trong quá trình triển khai.

Giá trị liên ngành của bài viết nằm ở lời nhắc rằng công nghệ không tồn tại tách rời các lựa chọn của con người. AI có thể trở thành phương tiện hỗ trợ hoặc nguồn khuếch đại tổn hại, tùy thuộc vào những điều kiện thiết kế, mục tiêu sử dụng và cơ chế quản trị. Từ đó, đối thoại giữa Phật học và AI có ý nghĩa thiết thực nhất khi không chỉ đặt câu hỏi máy móc có thể làm được gì, mà còn giúp con người suy xét sâu hơn về việc nên làm gì, vì mục đích nào và với trách nhiệm ra sao.

[Lưu ý học thuật]:

1. Về tính liên tục của lập luận: Phần 2 tiếp nối nền tảng triết học đã trình bày ở Phần 1, chuyển trọng tâm từ các vấn đề vô ngã, tính Không và bản chất của AI sang khả năng ứng dụng những gợi mở từ Phật học trong thực hành nhận thức và định hướng thiết kế công nghệ. Khi biên tập, cần bảo đảm mạch chuyển này rõ ràng, tránh tạo cảm giác bài viết chuyển đột ngột từ triết luận sang khuyến nghị công nghệ.

2. Về thuật ngữ Phật học: Cần thống nhất cách dùng các thuật ngữ: duyên khởi, tính Không, vô ngã, Bát-nhã, phương tiện thiện xảo, chỉ-quán, A lại da thức, chủng tử, hiện hành và huân tập. Những thuật ngữ có nguồn gốc từ các truyền thống hoặc hệ thống tư tưởng khác nhau cần được đặt trong đúng ngữ cảnh, không mặc nhiên xem là tương đương hoàn toàn.

3. Về cách diễn đạt liên quan đến AI: Các cụm từ như "tâm của AI", "tâm nhiễm ô của AI", "tâm thanh tịnh của AI" và "cõi đất số thanh tịnh" cần được hiểu là những cách diễn đạt mang tính ẩn dụ hoặc khung diễn giải của tác giả. Chúng đề cập đến định hướng giá trị, cơ chế thiết kế, quản trị, phản hồi và đầu ra của hệ thống, không phải bằng chứng cho thấy AI có tâm thức, đời sống nội tâm hay năng lực tu chứng như con người.

4. Về mối liên hệ giữa Phật học và khoa học công nghệ: Các phép tương chiếu giữa Duy thức học, quá trình huân tập và hoạt động của thuật toán cần được trình bày như những gợi ý diễn giải liên ngành. Không nên đồng nhất dữ liệu huấn luyện với chủng tử, mô hình AI với A lại da thức, hoặc sự thay đổi đầu ra của hệ thống với chuyển hóa nghiệp theo nghĩa giáo lý. Những tương đồng về mặt gợi mở không tự thân chứng minh sự tương đương về bản chất.

5. Về các dẫn chứng khoa học và sự kiện: Những nhận định liên quan đến khả năng của AI, tác động của chatbot, thiên kiến thuật toán và các nghiên cứu về AI hỗ trợ thực hành tâm lý cần được đối chiếu với nguồn trích dẫn đầy đủ. Trường hợp chatbot Tay của Microsoft nên được trình bày đúng bối cảnh lịch sử và nguồn dẫn. Không mở rộng một kết quả nghiên cứu giới hạn thành kết luận rằng AI có thể thay thế chuyên gia tâm lý, bảo đảm hiệu quả trị liệu hoặc giúp người dùng đạt giác ngộ.

6. Về trích dẫn Phật giáo: Các dẫn ngữ gắn với Hòa thượng Tinh Vân, *Đại thừa Khởi tín luận* và những câu nói về thiền sư Đạo Sinh, hữu tình-vô tình cần được kiểm tra về xuất xứ, bản dịch và phạm vi tư tưởng. Đặc biệt, các quan niệm về Phật tính và khả năng chứng ngộ của hữu tình, vô tình có thể được diễn giải khác nhau giữa các truyền thống Phật giáo.

7. Về nguyên tắc biên tập: Giữ nguyên luận điểm và sắc thái tư tưởng của tác giả, nhưng phân biệt rõ giữa giáo lý Phật giáo, giả thuyết triết học, phép loại suy liên ngành và kết luận thực nghiệm. Chỉ điều chỉnh cách diễn đạt để tăng độ chính xác, mạch lạc và khả năng tiếp nhận; không tự ý nâng các giả thuyết của tác giả thành kết luận khoa học đã được xác lập.

Chú thích:

23. Trợ duyên (*zhuyuan* □□ ; "điều kiện hỗ trợ"): Thuật ngữ Phật giáo chỉ những yếu tố góp phần hỗ trợ sự sinh khởi của một trạng thái hoặc sự kiện, nhưng không phải là nguyên nhân chính tạo nên trạng thái hoặc sự kiện ấy. Nói một cách đơn giản, *zhuyuan* chỉ những hoàn cảnh, nguồn lực hoặc mối quan hệ thuận lợi giúp một sự việc có thể xảy ra.

24. Không có tự tính (*asvabhāva*, □□□): Quan điểm Phật giáo cho rằng các pháp không có một bản chất cố hữu, độc lập. Sự vật tồn tại và vận hành tùy thuộc vào nhân, duyên và sự định danh bằng khái niệm. Vì vậy, căn tính của chúng mang tính tùy thuộc và quan hệ, chứ không tồn tại một cách độc lập. □□□□□□□□□□□□

Dịch nghĩa câu Hán văn: Các pháp đều do nhân duyên sinh khởi, vì thế không có tự tính.

25. Tự tính thành Phật (*svabhāva-buddha*, □□□□): Một mệnh đề cô đọng về giáo nghĩa trong Phật giáo Đại thừa Đông Á, đặc biệt trong các diễn ngôn về Phật tính (*tathāgatagarbha*). Mệnh đề này nhấn mạnh rằng quả vị Phật được hiện thực hóa thông qua sự giác ngộ tự tính của chính mình, chứ không phải được tiếp nhận từ một nguồn bên ngoài.

26. CBETA T10, số 279, tr. 102, cột b, dòng 1. *Kinh Hoa Nghiêm* (□□□□□□□), phẩm "Kệ tụng tại cung trời Dạ-ma" (bài kệ của Bồ tát Giác Lâm): "□□□□□□□□□□□□□□□□□□□□□□□□ ".

Dịch nghĩa: Nếu muốn thấu hiểu chư Phật trong ba đời, nên quán xét tự tính của pháp giới; tất cả đều do tâm tạo.

27. CBETA T47, số 1998A, tr. 875, cột a, dòng 3-4: "□□□□□□□□□□□□□□□□ ".

Dịch nghĩa: Tre xanh mướt thảy đều là Pháp thân; hoa vàng rực rỡ không gì chẳng phải Bát nhã.

Đây là một câu cách ngôn Thiền tông thể hiện tư tưởng bất nhị: những hiện tượng bình thường trong đời sống không nằm ngoài giác ngộ. "Pháp thân" (*dharmakāya*) chỉ thân Phật vô tướng, biến khắp; còn Bát nhã (*prajñā*) là trí tuệ giải thoát, thấu hiểu tính Không và tính tương thuộc của các pháp. Theo cách diễn giải trong Thiền tông, câu này dạy rằng khi sự chấp trước mê lầm lắng xuống, núi non, tre trúc và hoa lá chính là trường hiển lộ của sự chứng ngộ; không có một thực tại thứ hai ẩn phía sau chúng. Câu nói này được lưu truyền rộng rãi trong các trước tác Thiền tông về sau, bao gồm cả truyền thống Thiền Nhật Bản (Zen).

28. CBETA T48, số 2008, tr. 349, cột a, dòng 19-20: "□□□□□□□□□ ".

Dịch nghĩa: Thật kỳ diệu thay, tự tính vốn tự đầy đủ.

29. Nhận lầm vọng tưởng là chân thật (□□□□): Một sai lầm nhận thức điển hình trong Phật giáo, chỉ việc xem những gì hư ảo, do cấu tạo mà thành hoặc chỉ có giá trị quy ước là thực tại hay chân lý tối hậu.

30. Chỉ và quán (*śamatha* và *vipaśyanā*; □□): Hai pháp tu quán niệm bổ trợ cho nhau. *Śamatha* (chỉ) giúp ổn định sự chú tâm và làm lắng dịu những dao động của tư duy phân biệt; *vipaśyanā* (quán) khảo sát các hiện tượng nhằm nhận rõ tính vô thường, vô ngã và duyên khởi. Trong các truyền thống Phật giáo Trung Hoa, pháp tu kết hợp này được gọi là *chỉ quán* (*zhiguan*): chỉ (□) tương ứng với sự an định và làm lắng dịu tâm, còn quán (□) chỉ sự quán sát phân tích hoặc tuệ giác không bị cuốn theo dòng tư duy phân biệt.

31. Phương tiện thiện xảo (*upāya*, □□□□): Những phương pháp được vận dụng linh hoạt tùy theo hoàn cảnh nhằm hướng dẫn chúng sinh tiến đến giải thoát mà không đi chệch khỏi trí tuệ và lòng từ bi. Trong Phật giáo Đại thừa, đặc biệt là *Kinh Pháp Hoa*, khái niệm phương tiện thiện xảo (*upāya-kauśalya*) xác lập cơ sở cho việc điều chỉnh giáo pháp và phương pháp tu tập phù hợp với năng lực, nhu cầu và điều kiện của từng đối tượng.

32. AheadForm (□□□□). Không ghi năm. "AheadForm | Humanoid Robot" [AheadForm | Robot hình người]. https://www.aheadform.com/ (truy cập ngày 10 tháng 10 năm 2025).

33. CBETA T08, số 235, tr. 752, cột b, dòng 28: "□□□□□□□□□□□□□□□□□□□□□□□□ ".

Dịch nghĩa: Tất cả các pháp hữu vi đều như mộng, như huyễn, như bọt nước, như bóng; như sương, như điện chớp. Nên quán chiếu như vậy.

34. CBETA T17, số 842, tr. 913, cột b, dòng 24.

Sáu trần duyên ảnh (*liuchen yuanying*, □□□□): Sáu trần là sáu đối tượng của giác quan, bao gồm sắc, thanh, hương, vị, xúc và pháp. Gọi chúng là "duyên ảnh" (□□), tức những hình ảnh sinh khởi tùy thuộc vào các điều kiện, nhằm nhấn mạnh rằng chúng chỉ xuất hiện nhờ nhân duyên, trong đó có các giác quan và tâm thức, nên không có tự tính độc lập.

Có thể hình dung qua hai phép ẩn dụ:

Gương và bóng phản chiếu: Chân tâm được ví như một tấm gương sáng. "Sáu trần duyên ảnh" là những hình ảnh xuất hiện rồi biến mất trong gương ấy - núi sông, con người và hoa lá. Sai lầm thông thường là bám chấp vào một hình ảnh phản chiếu cụ thể, xem nó như chính tấm gương, mà quên mất khả năng sáng trong vốn có của gương, nhờ đó mọi sự vật đều có thể được phản chiếu.

Phim và màn chiếu: Sáu trần giống như những khuôn hình trên cuộn phim; sự chấp trước của tâm giống như ánh sáng từ máy chiếu; còn những hình ảnh chuyển động được chiếu lên màn là các "duyên ảnh" xuất hiện trên nền chân tâm. Chúng ta bị cuốn theo vui buồn của bộ phim, tức những cảm xúc và tư tưởng của mình, rồi xem chúng là thực tại tối hậu, trong khi quên rằng tất cả đều tùy thuộc vào màn chiếu không chuyển động.

Khái niệm "sáu trần duyên ảnh" là lời cảnh tỉnh rằng dòng "vọng tâm" không ngừng dao động và sinh khởi phân biệt không phải là chủ thể chân thật làm chủ đời sống của chúng ta. Mục đích tu tập là buông bỏ mê lầm để trở về với chân thật: nhận ra tính tùy thuộc và tính kiến tạo của những hình ảnh ấy, đồng thời

quay lại với sự sáng trong tĩnh lặng, có khả năng làm nền cho mọi biểu hiện của chân tâm.

35. CBETA T51, số 2076, tr. 268, cột a, dòng 22-23: “□□□□□□□□□ ”.

Dịch nghĩa: Chỉ cần lìa các duyên vọng lầm, tức là Phật như như.

“Duyên vọng lầm” chỉ những mối ràng buộc do vô minh và sự chấp trước đối với sáu trần tạo nên. “Như như” (*tathatā*) chỉ tính như thật của thực tại và trạng thái giác ngộ. Câu này thể hiện một tư tưởng Thiền tông: giác ngộ không phải điều được bổ sung từ bên ngoài; khi tâm thôi vướng mắc vào các duyên mê lầm, sự sáng trong vốn có của tâm chính là Phật tính được hiển lộ.

36. CBETA T33, số 1698, tr. 80, cột a, dòng 15-16: “□□□□□□□□ ”.

Dịch nghĩa: Lìa tất cả các tướng, đó gọi là chư Phật.

37. Vô ngã (*anātman*, □□): Giáo lý và phương pháp phân tích của Phật giáo khẳng định rằng không có một “cái tôi” cố định, thường hằng nằm bên dưới các uẩn cấu thành đời sống thân-tâm.

38. CBETA T31, số 1585, tr. 10, cột a, dòng 7: “□□□□□□□□ ”.

Dịch nghĩa: Cái có khả năng huân tập làm phát sinh chủng tử; chủng tử khởi lên thành hiện hành.

39. CBETA X38, số 697, tr. 868, cột c, dòng 7: “□□□□□□□□ ”.

Dịch nghĩa: Khi thiền sư Đạo Sinh thuyết pháp, ngay cả những hòn đá cứng cỏi cũng gật đầu.

Câu chuyện này được ghi lại trong một số nguồn tư liệu, trong đó có các trước tác về cao tăng thuộc Liên xã. Nhân vật trung tâm là Zhu Daosheng (□□□ , Trúc Đạo Sinh), vị tăng sống vào thời Đông Tấn, thường được gọi là Sinh công. Sau khi nghiên cứu *Kinh Đại Bát Niết-bàn*, ngài mạnh dạn chủ trương rằng ngay cả *nhất-xiển-đề* (*icchantika*) những chúng sinh bị xem là đoạn tuyệt căn lành cũng có Phật tính và có thể thành Phật.

Do quan điểm này bị xem là dị thuyết, ngài bị trục xuất. Sau đó, ngài đến núi Hổ Khâu, thuyết giảng *Kinh Niết-bàn* cho những hòn đá. Khi ngài trình bày luận điểm trên, các hòn đá được kể là đã “gật đầu” tán đồng. Bản dịch *Kinh Niết-bàn* đầy đủ hơn được lưu hành về sau đã minh thị khẳng định Phật tính phổ quát, qua đó được xem là sự xác nhận cho lập trường của Đạo Sinh và góp phần làm cho điển tích “đá gật đầu” trở nên nổi tiếng.

40. CBETA T48, số 2023, tr. 1073, cột a, dòng 21: “□□□□□□□□ ”.

Dịch nghĩa: Hữu tình và vô tình cùng viên thành nhất thiết chủng trí.

Đây là một mệnh đề thể hiện tư tưởng bất nhị: không chỉ chúng sinh hữu tình mà cả những sự vật vô tình như cỏ cây, ngói đá cũng tham dự trong pháp giới. Theo cách diễn giải này, sự giác ngộ không phải là đặc quyền riêng của những chủ thể có ý thức.

Chú giải thuật ngữ:

1. Vô ngã (*anātman*)

Quan niệm cho rằng không có một tự ngã độc lập, thường hằng và tự chủ tồn tại tách biệt khỏi các điều kiện cấu thành. Khi vận dụng vào thảo luận về AI, cần phân biệt việc dùng vô ngã như một lăng kính triết học với khẳng định rằng máy móc đã được chứng minh là có hoặc không có tự ngã theo nghĩa của Phật giáo.

2. Duyên khởi (*pratītyasamutpāda*)

Nguyên lý cho rằng các pháp sinh khởi tùy thuộc vào những điều kiện và quan hệ nhân duyên, không tồn tại hoàn toàn độc lập. Trong bài viết, duyên khởi gợi mở cách xem xét AI trong mạng lưới dữ liệu, con người, thiết kế, hạ tầng và môi trường xã hội.

3. Tính Không (*śūnyatā*)

Chỉ sự vắng mặt của tự tính độc lập nơi các pháp. Tính Không không đồng nghĩa với hư vô hay phủ nhận sự tồn tại quy ước của sự vật. Khi áp dụng vào công nghệ, đây là một cách diễn giải triết học, không phải thuật ngữ kỹ thuật mô tả trực tiếp cấu trúc của mô hình AI.

4. Bát nhã (*prajñā*)

Trí tuệ có khả năng nhận thức sâu sắc thực tại, đặc biệt là vô ngã và tính Không. Trong ngữ cảnh AI, cần phân biệt năng lực xử lý thông tin hoặc tạo ra câu trả lời có vẻ hợp lý với trí tuệ giải thoát theo nghĩa Phật học.

5. Phương tiện thiện xảo (*upāya*)

Phương pháp hoặc cách thức vận dụng linh hoạt để hỗ trợ chúng sinh tùy theo căn cơ và hoàn cảnh. Việc xem AI như một phương tiện hỗ trợ học tập, suy tư hay thực hành cần được đánh giá theo mục đích, cách sử dụng và hệ quả,

không mặc nhiên xem công nghệ là phương tiện thiện xảo chỉ vì nó có ích.

6. Chỉ (*śamatha*) và quán (*vipaśyanā*)

Chỉ nhấn mạnh sự an định tâm; quán nhấn mạnh sự quan sát, khảo sát và thấy biết rõ thực tại. Đây là những thuật ngữ thuộc lĩnh vực thực hành thiền định. Các ứng dụng AI hỗ trợ ghi chép hoặc phản tư không nên được đồng nhất với việc hành giả thực sự tu tập chỉ-quán.

7. A lại da thức (*ālaya-vijñāna*)

Khái niệm quan trọng trong Duy thức học, thường được giải thích là thức tàng chứa và duy trì các chủng tử, góp phần giải thích sự tiếp nối của kinh nghiệm và sự hình thành các hiện hành. Không nên xem A lại da thức như một kho dữ liệu máy tính theo nghĩa đen.

8. Chủng tử (*bīja*)

Trong Duy thức học, chủng tử chỉ những tiềm năng hoặc năng lực có thể làm phát sinh các hiện hành khi hội đủ điều kiện. Việc so sánh chủng tử với dữ liệu huấn luyện hoặc những khuynh hướng được hình thành trong mô hình AI chỉ có giá trị loại suy; hai khái niệm không đồng nhất về bản chất.

9. Huân tập (*vāsanā*)

Quá trình những kinh nghiệm, hành vi và nhận thức để lại dấu ấn, góp phần hình thành hoặc củng cố khuynh hướng về sau. Tác giả vận dụng khái niệm này để gợi liên hệ với cách dữ liệu, mục tiêu tối ưu hóa và tương tác ảnh hưởng đến đầu ra AI. Đây là phép tương chiếu liên ngành, không phải khẳng định thuật toán có cơ chế huân tập giống tâm thức con người.

10. Hiện hành (*abhivyakti* hoặc cách dịch tùy ngữ cảnh)

Trong ngữ cảnh Duy thức học, thường chỉ sự biểu hiện hoặc hoạt động hiện thời của những năng lực, khuynh hướng đã có điều kiện phát sinh. Khi dùng thuật ngữ này, cần đối chiếu nguyên văn để xác định chính xác thuật ngữ Sanskrit mà tác giả muốn chỉ đến, tránh áp một từ tương đương khi chưa có đủ căn cứ.

11. Tâm nhiễm ô của AI (*defiled AI mind/heart*, theo cách diễn đạt của tác giả)

Một ẩn dụ chỉ những định hướng thiết kế và vận hành có thể duy trì hoặc khuếch đại thiên kiến, thông tin sai lệch, nội dung cực đoan hay hành vi gây tổn hại. Thuật ngữ này không hàm ý AI có phiền não hoặc tâm thức theo nghĩa Phật

học.

12. Tâm thanh tịnh của AI (*purified AI heart/mind*, theo cách diễn đạt của tác giả)

Một ẩn dụ về việc xây dựng hệ thống AI theo các định hướng đạo đức, có trách nhiệm và hướng đến giảm thiểu tổn hại. Cụm từ này nên được hiểu ở cấp độ thiết kế, quản trị và tác động xã hội, không phải trạng thái thanh tịnh tâm thức của một chủ thể tu hành.

13. Môi trường thuật toán nhiễm ô

Cách gọi của tác giả dành cho môi trường mà thiết kế, dữ liệu, cơ chế phản hồi hoặc quản trị không đủ khả năng ngăn chặn các đầu ra gây hại, qua đó có thể củng cố định kiến hay khuynh hướng tiêu cực. Đây là thuật ngữ diễn giải được đề xuất trong bài, không phải một thuật ngữ Phật học truyền thống.

14. Môi trường thuật toán thanh tịnh

Cách gọi mang tính chuẩn tắc về môi trường AI được thiết kế và quản trị nhằm hỗ trợ những kết quả có ích, giảm thiểu thiên kiến và bảo vệ người sử dụng. “Thanh tịnh” ở đây là ngôn ngữ ẩn dụ về định hướng đạo đức, không phải sự thanh tịnh theo nghĩa giải thoát.

15. Cõi đất số thanh tịnh

Hình ảnh ẩn dụ chỉ một môi trường số được xây dựng theo những nguyên tắc nhân văn, có trách nhiệm và hạn chế gây tổn hại. Không nên hiểu cụm từ này là một cõi Phật theo nghĩa tôn giáo hoặc một trạng thái xã hội đã được bảo đảm sẽ đạt tới nhờ AI.

16. Tăng thượng duyên (*adhipati-pratyaya*; Hán: □□□)

Một trong các loại duyên được trình bày trong hệ thống phân tích nhân duyên của Phật giáo, chỉ điều kiện có tác dụng hỗ trợ hoặc chi phối sự sinh khởi của một pháp. Khi sử dụng trong bài, cần xác định đúng ngữ cảnh giáo lý và tránh giản lược thành quan hệ nhân quả tuyến tính.

17. Phật tính (*buddha-dhātu*; cách dùng và đối chiếu thuật ngữ tùy văn bản)

Khái niệm được diễn giải trong một số truyền thống Đại thừa về khả năng hoặc nền tảng thành Phật của chúng sinh. Nội hàm Phật tính không hoàn toàn giống nhau trong mọi hệ thống tư tưởng Phật giáo. Vì vậy, khi tác giả mở rộng hình

ảnh Phật tính sang AI hoặc các hiện tượng vô tình, cần trình bày rõ đó là lập trường diễn giải của tác giả, không mặc nhiên xem là quan điểm chung của toàn bộ Phật giáo.

18. Thiên kiến thuật toán (*algorithmic bias*)

Sự sai lệch có tính hệ thống trong kết quả của một hệ thống thuật toán, có thể phát sinh từ dữ liệu, cách gán nhãn, lựa chọn mô hình, mục tiêu tối ưu hóa hoặc môi trường triển khai. Thiên kiến thuật toán không nhất thiết bắt nguồn từ một chủ ý phân biệt đối xử có ý thức của người thiết kế, nhưng vẫn cần được đánh giá và kiểm soát vì các hệ quả thực tế.