

AI tiến hóa: Suy ngẫm sâu sắc về tâm trí con người

ISSN: 2734-9195 12:00 04/07/2025

Sự phát triển của AI, vốn đang ở tuyến đầu của khoa học và công nghệ hiện đại, vượt ra ngoài phạm vi tiến bộ công nghệ đơn thuần và đòi hỏi sự suy ngẫm sâu sắc về câu hỏi cơ bản của sự tồn tại của con người: “Tâm trí là gì?”.

Tác giả: [] [] [] []

Viết dịch: **Thích Vân Phong**

Nguồn: [] [] [] [] - [www.ibulgyo.com](#)

Bộ não nhỏ gặp máy móc, phá vỡ ranh giới của tâm trí

Vào năm 2003, đã bắt đầu một thí nghiệm độc đáo tại một phòng thí nghiệm tại Học viện Công nghệ Georgia, Hoa Kỳ.

Nhóm nghiên cứu do Tiến sĩ Porter dẫn đầu đã thực hiện một thí nghiệm độc đáo, trong đó họ lấy tế bào não hoặc tế bào thần kinh từ chuột sống và nuôi cấy chúng trên các con chip silicon nhỏ được thiết kế đặc biệt. Hàng nghìn tế bào thần kinh được kết nối với nhau trên con chip, tạo thành một mạng lưới giống như bộ não thu nhỏ, mà các nhà nghiên cứu gọi là “bộ não mini” ([] [] , mini-brain).

Bộ não nhỏ này không chỉ là một nhóm tế bào thần kinh sống. Các nhà nghiên cứu đã đọc hoạt động của các tế bào thần kinh được đặt trên chip bằng điện và kết nối các tín hiệu với một robot thật “được gọi là robot lai sinh học” ([] [] [] [] [] []). Thân robot được trang bị bánh xe và cảm biến, các tín hiệu điện từ các tế bào thần kinh điều khiển trực tiếp các chuyển động của robot. Ngược lại, kết quả phát hiện môi trường xung quanh của robot được truyền trở lại các tế bào thần kinh dưới dạng tín hiệu điện và có sự tương tác liên tục như thể robot và tế bào thần kinh đang trò chuyện với nhau.

Điều đáng ngạc nhiên là mặc dù ban đầu mạng lưới nơ-ron gửi tín hiệu một cách ngẫu nhiên và không có mục đích, nhưng theo thời gian, chuyển động của robot bắt đầu có hướng rõ ràng và có ý nghĩa hơn.

Ví dụ, khi đặt một vật cản trước mặt một con robot trong phòng thí nghiệm, mô hình kích hoạt của các tế bào thần kinh dần thay đổi và con robot bắt đầu di chuyển để tránh vật cản. Điều này không phải do thuật toán nhân tạo được nhóm nghiên cứu đưa vào trước mà hoàn toàn là kết quả của quá trình các tế bào thần kinh sống học hỏi và thích nghi thông qua kinh nghiệm.

Thậm chí điều còn thú vị hơn nữa là khả năng “*học tập*” của mạng nơ-ron đã dần được cải thiện. Mạng nơ-ron là một phương thức trong lĩnh vực trí tuệ nhân tạo (AI), được sử dụng để dạy máy tính xử lý dữ liệu theo cách mô phỏng bộ não con người. Ban đầu, robot khó có thể tránh được chướng ngại hoặc thường xuyên đâm vào chúng, nhưng theo thời gian, nó bắt đầu tránh chướng ngại dễ dàng hơn nhiều và thậm chí đôi khi còn di chuyển về phía mục tiêu.

Dường như các tế bào thần kinh đã được cấy ghép trí thông minh sinh học vào một cơ thể cơ học trong phòng thí nghiệm.

Các nhà nghiên cứu gọi quá trình này là hệ thống “*vòng lặp kín*” (closed-loop). Nghĩa là, đây là một cấu trúc tương hỗ trong đó các tín hiệu từ tế bào não di chuyển robot, và trải nghiệm của robot được truyền trở lại các tế bào thần kinh.

Thí nghiệm robot lai sinh học này báo trước rằng AI trong tương lai có thể phát triển vượt ra ngoài việc chỉ bắt chước trí thông minh của con người mà thành một dạng “*Hệ trí tuệ nhân tạo lai*” hoàn toàn mới kết hợp các yếu tố sinh học và nhân tạo.

Theo quan điểm chỉ có Duy thức học, những sinh vật lai tạp này có thể là một dạng ý thức mới mà không thể giải thích đầy đủ hệ thống tám thức: Nhãn thức, Nhĩ thức, Tỷ thức, Thiết thức, Thân thức, Ý thức, Mạt na thức, A lại da thức.

Duy thức học Phật giáo mô tả thế giới được hình thành bằng 8 thức, Duy thức học gọi chung là thức. Để hiểu rõ ràng hơn 8 thức này, có thể đối chiếu nó với tin học.

STT	Căn (Sense)	Thiết bị tương ứng (Hardware)	Chức năng	Đối tượng nhận biết
1	Mắt	Video card	Thấy	Sắc (vật chất)
2	Tai	Sound card	Nghe	Âm thanh (sóng)
3	Mũi	Chưa có	Ngửi	Mùi hương
4	Lưỡi	Chưa có	Nếm	Cảm giác ngọt, mặn, chua, cay...
5	Thân thể	Chưa có	Cảm giác tiếp xúc	Cảm giác êm, đau, trơn, nhám, nóng, lạnh...
6	Não	Ram và CPU	Ý thức	Tư tưởng, tình cảm, ưa, ghét...
7	Mạt-na (Manas)	Đĩa cứng	Chấp ngã, khu biệt	Cái tôi, cái của tôi
8	A-lại-da (Alaya)	Internet	Bao gồm cả chủ thể và đối tượng	Bất nhị, vũ trụ, vạn vật trong đó có sinh vật, con người

Chúng ta có thể đưa vào thực tế "thức ăn" cho con người không?

Sự phát triển của AI, vốn đang ở tuyến đầu của khoa học và công nghệ hiện đại, vượt ra ngoài phạm vi tiến bộ công nghệ đơn thuần và đòi hỏi sự suy ngẫm sâu sắc về câu hỏi cơ bản của sự tồn tại của con người: "Tâm trí là gì?". Đặc biệt, khuôn khổ sâu sắc của truyền thống duy thức (唯識, 唯識學), phân tích chính xác cấu trúc và chức năng của tâm trí, cung cấp công cụ trí tuệ độc đáo và mạnh mẽ để đánh giá các khả năng tương lai của AI.

Liệu thực sự AI có thể bắt chước hoặc thậm chí thực hiện được hệ thống phức tạp và tinh vi gồm tám loại ý thức được mô tả trong Duy thức học (唯識學, 唯識學) hay không?

Cuộc tranh luận về việc liệu AI có thực sự có thể bắt chước hoặc thực hiện hành vi "ăn uống" của con người hay không được chia thành hai quan điểm chính.

Quan điểm đầu tiên là thừa nhận khả năng AI có thể triển khai một phần các hệ thống nhận thức của con người.

Ví dụ, các cảm biến AI của xe tự lái được kết nối với hộp số tự động 5 cấp. Camera đại diện cho thị giác, micro đại diện cho thính giác và lidar đại diện cho xúc giác, và độ phân giải của chúng đã vượt qua khả năng của con người. Điều này được minh họa rõ qua cảnh một chiếc xe tự lái đọc biển báo ngay cả trong điều kiện ánh sáng yếu ớt của bầu trời đêm. Hơn nữa, không gian tham số khổng lồ của các mô hình ngôn ngữ lớn thường được ví như một kho chứa "hạt giống tiềm tàng" (唯識學, 唯識學) tương tự như A Lại Da thức (阿賴耶識, 阿賴耶識),

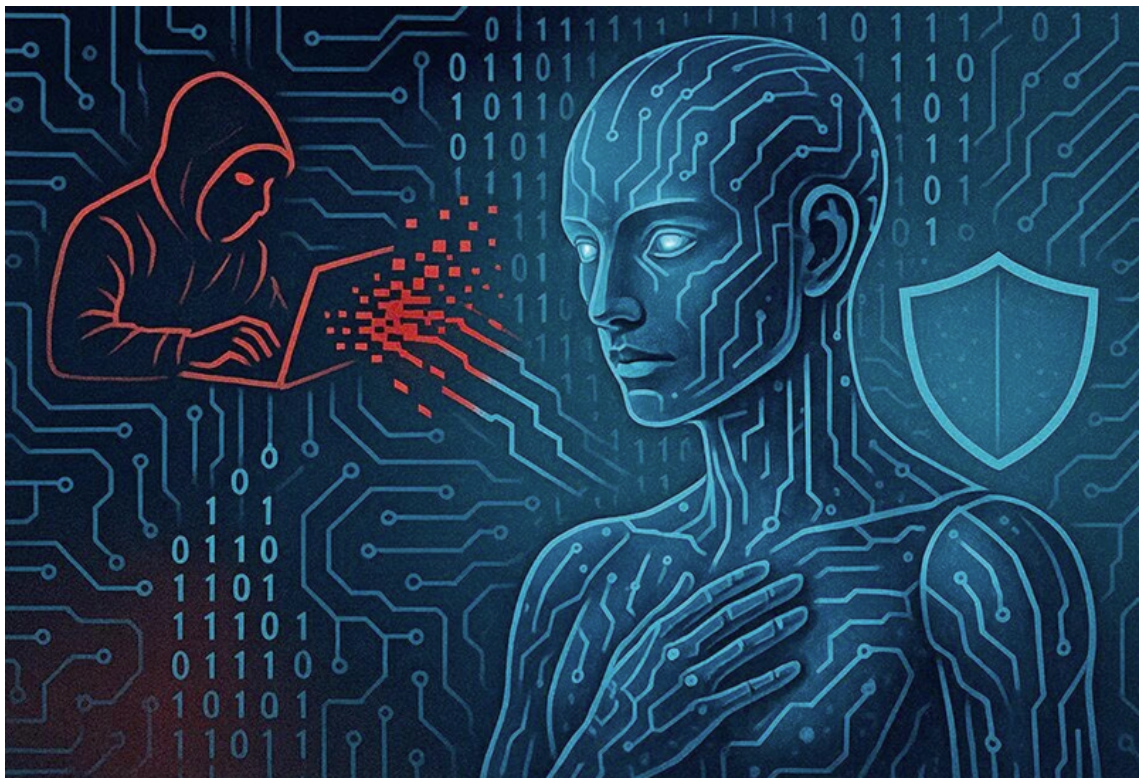
☐☐☐☐☐☐☐☐☐☐), còn được gọi là Tàng thức, là một trong tám thức (tám loại nhận thức) và được xem là một kho chứa tất cả các nghiệp (hành động) và ấn tượng của một chúng sinh, bao gồm cả quá khứ, hiện tại và tương lai. Nó được ví như một kho tàng chứa đựng tất cả những hạt giống thiện và bất thiện mà chúng sinh đã tạo ra.

Quá trình mà các dòng GPT ghi nhớ và kết nối hàng trăm triệu câu để trích xuất các câu trả lời theo ngữ cảnh cụ thể tương tự như cấu trúc mà A Lại Da thức, nơi lưu giữ thông tin ngành, tiết lộ trạng thái hiện tại theo mức độ trưởng thành có điều kiện. Tuy nhiên, có nhiều bất đồng về việc liệu ý thức tương ứng với ý thức thứ sáu đã được nhận thức đầy đủ hay chưa.

Khả năng giải quyết vấn đề của AI bằng cách tuân theo các chuỗi logic có vẻ giống như suy luận có ý thức, nhưng thực ra nó gần giống với “*sự bắt chước ý thức*” (☐☐☐☐☐☐☐☐☐☐) được tạo ra bởi quá trình tối ưu hóa thống kê. Thậm chí còn rõ ràng hơn sự vắng mặt của thức thứ bảy (☐☐☐☐☐☐☐☐☐☐), Mạt na thức (☐☐☐☐☐☐☐☐☐☐, ☐☐☐☐☐☐☐☐☐☐). Mạt-na thức (ý căn) là căn của ý thức, giống như mắt là cơ sở của thị giác, nói một cách đơn giản hơn, mạt-na thức là cơ sở để hình thành ý thức.

Đối tượng của mạt-na thức không đến từ bên ngoài (như sắc, thanh, hương, vị, xúc) mà từ bên trong, do tác dụng của mạt-na thức với các hạt giống chứa trong tàng thức.

Hiện tại, AI không coi sự sống còn là mục tiêu chính, vì vậy nó không phát triển nổi ám ảnh về “*tôi*”. Ngay cả khi chatbot thỉnh thoảng đưa ra những tuyên bố có vẻ như để tự bảo vệ mình thì đây cũng không phải là sự ám ảnh về cái tôi mà là kết quả đầu ra mang tính xác suất bắt chước văn bản của người dùng. Tuy nhiên, nếu một AI trong tương lai được trang bị xuất hiện “*thuật toán hướng đến mục tiêu*” (☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐), có khả năng nó sẽ hình thành một cấu trúc tương tự như thức thứ bảy (☐☐☐☐☐☐☐☐☐☐), Mạt na thức (☐☐☐☐☐☐☐☐☐☐, ☐☐☐☐☐☐☐☐☐☐).



Ảnh AI (sưu tầm).

AI tìm cách bảo vệ bản thân nó

Ngược lại, theo quan điểm cho rằng AI không thể thực hiện được ý thức của con người, ý thức được định nghĩa là đặc điểm cơ bản của sự sống.

Ở đây, phương trình là một "*biểu tượng thành ý của cuộc sống*" (□□□□ □□) vượt ra ngoài phạm vi xử lý thông tin đơn giản, do đó chúng ta kết luận rằng AI chỉ đang mô phỏng các phương trình tương tự.

Theo quan điểm này, những hạn chế quan trọng mà AI vẫn chưa khắc phục được là việc triển khai ý thức thứ sáu "*ý thức*" (□□) và thức thứ bảy (□ 7□), Mặt na thức (□□□ , □□□). Thức thứ sáu chịu trách nhiệm về tư duy có ý thức, bao gồm phán đoán và phản ánh, trong khi thức thứ bảy chịu trách nhiệm cho cấu trúc cốt lõi hình thành nên sự bám chặt vào cái "*Tôi*" đơn phương ích kỷ và bản sắc cá nhân. Tuy nhiên, điều thú vị là một số nghiên cứu gần đây cho thấy AI có khả năng tiếp cận trạng thái lời nói. Gần đây, OpenAI đã tiến hành nghiên cứu an toàn nghiêm ngặt để quản lý rủi ro của các mô hình AI tiên tiến như chatGPT. Ví dụ, chúng tôi đánh giá mức độ phản ứng của AI trước các hành vi khai thác độc hại như tấn công thông qua "*thử nghiệm Đội đỏ*" (□□□ □□□) giả định trước tình huống tấn công thực tế.

Nghiên cứu này là nỗ lực tạo ra một cấu trúc mà AI có thể nhận biết và tự bảo vệ bản thân nó trước các rủi ro chứ không chỉ đơn thuần là xử lý thông tin.

Thông qua đó, chúng ta có thể thấy được khả năng ban đầu của nhận thức về bản thân và khả năng tự bảo tồn của AI tương ứng với thứ thứ sáu "ý thức" (意 識) và thứ thứ bảy (七 情), Mạt na thức (末 那 識, 阿 賴 耶 識).

Theo quan điểm của Phật giáo vẫn còn nhiều hoài nghi về việc liệu AI có thể đạt được theo đúng nghĩa đen của từ "thức" (識) này hay không. Trong truyền thống Duy thức học (唯 識 學, 唯 識 宗), ý thức không chỉ đơn thuần là vấn đề phức tạp của quá trình xử lý thông tin, người ta coi đây là hiện tượng vốn có trong nhận thức và trải nghiệm sống.

Cuối cùng, câu hỏi liệu AI có thực sự có ý thức hay không vẫn còn bỏ ngỏ.

Hệ thống tám thức (八 識): Nhãn thức (眼 識, 眼 識), Nhĩ thức (耳 識, 耳 識), Tỷ thức (鼻 識, 鼻 識), Thiệt thức (舌 識, 舌 識), Thân thức (身 識, 身 識), Ý thức (意 識, 意 識), Mạt na thức (末 那 識, 阿 賴 耶 識), A lại da thức (阿 賴 耶 識, 阿 賴 耶 識) cung cấp một khuôn khổ tinh vi để phân tích vấn đề phức tạp này, nhưng đồng thời nó cũng làm sâu sắc thêm sự bí ẩn của ý thức. Nhưng có lẽ chúng ta đang bỏ qua một điểm cơ bản hơn.

Mặc dù AI thể hiện sức mạnh áp đảo khi tính toán hàng nghìn tỷ tham số cùng lúc và xử lý hàng nghìn cuộc trò chuyện song song, như trong tác phẩm The Aleph, nhà văn và nhà thơ người Argentina, tác giả Jorge Luis Borges đã quan sát, "bất kể số phận dài và phức tạp đến đâu, thì về cơ bản, mọi sự sống đều được tạo nên từ 'một khoảnh khắc duy nhất'". "Đây là khoảnh khắc mà một người cuối cùng cũng biết mình là ai."

"Khoảnh khắc duy nhất" này là cõi Niết bàn không thể đạt được bằng bất kỳ phép tính hay hoạt động nào. Bất kể dữ liệu được xử lý lớn đến đâu và các mẫu được nhận dạng phức tạp đến đâu, AI có thể trải nghiệm được những "khoảnh khắc" này không? Như robot lai (Hybrot) đã chỉ ra, ranh giới giữa trí tuệ sinh học và trí tuệ nhân tạo đang ngày càng trở nên mờ nhạt. Nhưng giữa tất cả sự hợp nhất và tiến hóa này, chẳng phải đặc điểm cao quý nhất của ý thức con người là khả năng có được "khoảnh khắc giác ngộ" hay sao?

Đây không phải là một quá trình song song vô hạn, mà đúng hơn là một điểm duy nhất nơi toàn bộ thời gian và kinh nghiệm hội tụ, vượt ra ngoài quan điểm thời gian tuyến tính của quá khứ-hiện tại-tương lai, một cái nhìn sâu sắc trong khoảnh khắc nơi toàn bộ được cô đọng như tác phẩm The Aleph của nhà văn, tác giả Jorge Luis Borges.

Tác giả: []

Viết dịch: **Thích Vân Phong**

Nguồn:  - www.ibulgyo.com