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

Trí tuệ nhân tạo: Chuyển hóa nguy hiểm từ hành động của con người

ISSN: 2734-9195 15:10 14/02/2026

Trí tuệ nhân tạo, nếu được hiểu đúng và sử dụng đúng, là một công cụ chuyển hóa mạnh mẽ cho sự phát triển xã hội. AI không phải là mối đe dọa tự thân, mà chính lựa chọn và hành động của con người mới quyết định hướng đi của công nghệ.

Trí tuệ nhân tạo (Artificial Intelligence - AI) từ lâu đã trở thành chủ đề vừa hấp dẫn vừa gây lo ngại trong đời sống đương đại. Không ít người nhìn AI với sự kỳ vọng lớn lao, trong khi cũng có những nỗi sợ về nguy cơ công nghệ vượt khỏi tầm kiểm soát của con người.

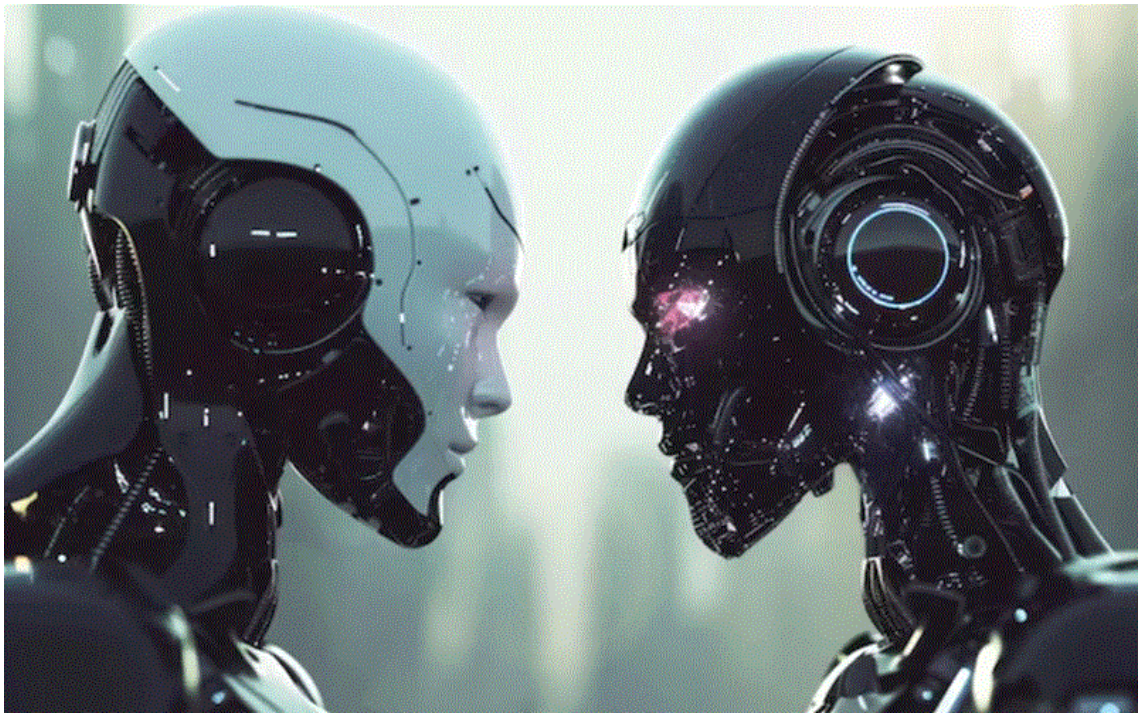
Xét bản chất, AI không phải là một thực thể nguy hiểm, đó chỉ là một công cụ mạnh mẽ có thể mang lại lợi ích to lớn cho nhân loại, hoặc gây ra những hệ lụy nghiêm trọng tùy thuộc hoàn toàn vào cách con người thiết kế, sử dụng và quản trị.

Ở tầng sâu hơn, AI là kết quả hội tụ của nhiều lĩnh vực như khoa học máy tính, toán học, khoa học dữ liệu và khoa học nhận thức. Mục tiêu của AI là xây dựng các hệ thống có khả năng học hỏi, suy luận và thích ứng, nhưng không có ý chí, động cơ hay ý thức tự thân.

Nguồn gốc của mọi rủi ro không nằm ở AI, mà nằm ở hành động và lựa chọn của con người trong quá trình phát triển và triển khai công nghệ này.

AI học như thế nào?

Cốt lõi của AI là khả năng học từ dữ liệu. Quá trình này thường diễn ra thông qua ba mô hình chính: học có giám sát, học không giám sát và học tăng cường.



Học có giám sát sử dụng dữ liệu đã được gán nhãn để "dạy" mô hình nhận biết mối quan hệ giữa đầu vào và đầu ra mong muốn. Học không giám sát cho phép hệ thống tự phát hiện các mô thức và cấu trúc ẩn trong dữ liệu chưa được gán nhãn. Trong khi đó, học tăng cường giúp AI học thông qua cơ chế thưởng - phạt, từ đó tối ưu hóa chuỗi quyết định nhằm đạt được một mục tiêu xác định.

Việc xây dựng một hệ thống AI bắt đầu từ thiết kế mô hình và thuật toán, sau đó là quá trình huấn luyện với khối lượng dữ liệu lớn. Thông qua nhiều vòng lặp, mô hình dần được tinh chỉnh để có khả năng khái quát và dự đoán chính xác hơn. Một số hệ thống còn được trang bị khả năng học liên tục, giúp chúng thích nghi với dữ liệu mới và cải thiện hiệu suất theo thời gian.

AI không có ý chí - chỉ phản chiếu mục đích của con người

Điều quan trọng cần nhấn mạnh là AI không có chủ ý, cảm xúc hay nhận thức. Những gì được gọi là "quyết định" của AI thực chất chỉ là kết quả của các phép tính phức tạp dựa trên dữ liệu đầu vào và các thuật toán đã được lập trình sẵn.

Vì vậy, AI tự nó không thiện cũng không ác. Nó trở thành công cụ hữu ích hay gây hại tùy thuộc vào mục đích sử dụng và chất lượng dữ liệu huấn luyện. Một hệ thống AI được đào tạo để phân tích hình ảnh y khoa có thể hỗ trợ chẩn đoán sớm bệnh hiểm nghèo, góp phần cứu sống nhiều người. Trong nông nghiệp, AI giúp tối ưu hóa tài nguyên, giảm lãng phí và tăng năng suất, hướng tới phát triển bền vững.

Những ứng dụng này cho thấy, khi được định hướng bởi các giá trị đạo đức và mục tiêu nhân bản, AI có thể trở thành một trợ duyên tích cực cho con người, tương hợp với tinh thần "lợi mình - lợi người" trong Phật giáo.

Khi rủi ro xuất hiện từ sự lệch hướng

Nguy cơ thực sự nảy sinh khi AI bị sử dụng sai mục đích, thiết kế cẩu thả hoặc triển khai mà thiếu cơ chế giám sát phù hợp. AI chỉ phản ánh dữ liệu mà nó được "nuôi dưỡng". Nếu dữ liệu chứa thiên kiến, kết quả đầu ra cũng sẽ tái tạo và khuếch đại những thiên kiến ấy.

Chẳng hạn, một hệ thống AI tuyển dụng được huấn luyện từ dữ liệu lịch sử thiên lệch về giới tính hoặc sắc tộc có thể vô tình duy trì sự bất công. Bên cạnh đó, nhiều mô hình AI hoạt động như "hộp đen", khiến con người khó hiểu được cơ chế ra quyết định. Khi những hệ thống như vậy được sử dụng trong lĩnh vực tư pháp hay an sinh xã hội, nguy cơ sai lệch và bất công là điều không thể xem nhẹ.

AI cũng có thể bị vũ khí hóa: từ các chiến dịch lừa đảo tinh vi, tạo video giả mạo (deepfake), cho đến việc thao túng thông tin và hủy hoại danh dự cá nhân. Trong bối cảnh này, trách nhiệm pháp lý và đạo đức của con người sử dụng AI cần được đặt ra một cách nghiêm túc. Luật pháp phải đủ mạnh để răn đe và bảo vệ công lý trong kỷ nguyên công nghệ.

Một rủi ro khác là hiện tượng thiên lệch tự động hóa – khi con người quá tin tưởng vào kết quả của AI mà không kiểm chứng. Điều này đặc biệt nguy hiểm trong các lĩnh vực nhạy cảm như y tế, hàng không hay tài chính. Thêm vào đó, tốc độ phát triển nhanh chóng của AI thường vượt xa khả năng điều chỉnh của các khung pháp lý, dẫn đến khoảng trống trong quản trị.

Học liên tục và yêu cầu tỉnh thức liên tục

Một trong những đặc điểm nổi bật của AI là khả năng học và thích nghi không ngừng. Trong giao thông tự hành hay y học hiện đại, AI phải liên tục cập nhật dữ liệu mới để phù hợp với điều kiện thay đổi.

Tuy nhiên, khả năng học liên tục này cũng đòi hỏi sự tỉnh thức và giám sát liên tục từ con người. Nếu thiếu cơ chế kiểm soát, AI có thể vô tình củng cố thiên kiến hoặc lệch khỏi mục tiêu ban đầu. Do đó, các nhà phát triển cần thiết lập hệ thống theo dõi, đánh giá và điều chỉnh định kỳ, nhằm bảo đảm AI luôn vận hành trong khuôn khổ đạo đức và lợi ích xã hội.

Hướng đến AI lấy con người làm trung tâm

Để khai thác tiềm năng của AI mà vẫn kiểm soát được rủi ro, cần có sự chung tay của nhiều chủ thể. Các nhà phát triển phải ưu tiên công bằng, minh bạch và trách nhiệm trong thiết kế hệ thống. Các phương pháp AI có khả năng giải thích (explainable AI) sẽ giúp tăng cường niềm tin và trách nhiệm giải trình.

Giáo dục cộng đồng, từ nhà hoạch định chính sách, giới chuyên môn đến công chúng về khả năng và giới hạn của AI là yếu tố then chốt. Một xã hội hiểu đúng về AI sẽ có năng lực đưa ra các quyết định sáng suốt hơn trong việc ứng dụng và quản trị công nghệ.

Ở tầm quốc tế, hợp tác xuyên biên giới là điều không thể thiếu để đối phó với các thách thức chung như an ninh mạng, chuẩn mực đạo đức và bất cân xứng pháp lý. Sự kết nối giữa chính phủ, học thuật và doanh nghiệp sẽ giúp AI phát triển hài hòa với nhu cầu và giá trị của nhân loại.

Kết luận

Trí tuệ nhân tạo, nếu được hiểu đúng và sử dụng đúng, là một công cụ chuyển hóa mạnh mẽ cho sự phát triển xã hội. AI không phải là mối đe dọa tự thân, mà chính lựa chọn và hành động của con người mới quyết định hướng đi của công nghệ.

Khi được phát triển trên nền tảng đạo đức, được giám sát chặt chẽ và không ngừng hoàn thiện, AI có thể trở thành chất xúc tác cho tiến bộ, thay vì nguồn cơn của bất an. Trong cách tiếp cận lấy con người làm trung tâm tương hợp với tinh thần tỉnh thức và trách nhiệm của Phật giáo, AI không còn là nỗi lo, mà là một trợ lực cho con đường phát triển bền vững và nhân văn.

Tác giả: **Dr Shubhro Chakrabartty -** *Giảng viên Trường Đại học Quốc tế DY Patil, Akurdi, Pune, Ấn Độ.*

Chuyển ngữ và biên tập: **Tịnh Như**

Nguồn: thestatesman.com