Các nhà khoa học lo lắng AI tổng hợp sẽ hủy diệt loài người

(Dân trí) - Trí tuệ nhân tạo (AI) có thể dẫn đến sự tuyệt chủng của chúng ta? Một số nhà khoa học, vốn rất coi trọng những rủi ro liên quan đến sự phát triển nhanh chóng của công nghệ này đã tin như vậy.



AI đang ngày càng tham gia vào cuộc sống của chúng ta (Ảnh minh họa: Trust my science).

Theo họ, khi AI nói chung xuất hiện, một số sự phát triển của nó có thể dẫn đến kết cục diệt vong của nhân loại, mặc dù khả năng xảy ra khá thấp, ước tính là 5%.

AI đang phát triển nhanh chóng và theo các chuyên gia, xu hướng này sẽ tiếp tục trong những năm tới. Một ngày nào đó, trí tuệ nhân tạo vượt trội hơn con người ở một số kỹ năng nhất định và đây mới chỉ là khởi đầu.

Một số nhà khoa học tin rằng, AI cuối cùng có thể là nguồn gốc dẫn đến sự tuyệt chủng của loài người. Điều này đã được tiết lộ bởi một cuộc khảo sát gần đây từ 2.778 nhà khoa học đã công bố các nghiên cứu về AI.

Những người tham gia đã chia sẻ ý kiến của họ về các mốc thời gian để đạt được các cột mốc quan trọng nhất định trong quá trình phát triển AI.

Họ cũng ý kiến về những tác động tiềm tàng mà những tiến bộ công nghệ này có thể gây ra cho xã hội.

Những tiến bộ lớn trong những thập kỷ tới

Các nhà nghiên cứu tham gia khảo sát tin rằng, ít nhất 50% khả năng hệ thống AI sẽ hoàn thành phần lớn các nhiệm vụ trí tuệ định hình con người ngày nay, cột mốc để đạt được điều này là vào năm 2028.

Điều tương tự cũng xảy ra với các kỹ năng liên quan đến nghệ thuật, điển hình như việc sáng tác độc lập các bài hát hoàn chỉnh khó có thể phân biệt được với các tác phẩm của con người hoặc việc tạo ra toàn bộ trang web dành riêng cho việc xử lý các giao dịch tài chính.

Tuy nhiên, không nhiều chuyên gia tin rằng đến năm 2028, AI sẽ có khả năng thực hiện các nhiệm vụ phức tạp liên quan đến lĩnh vực vật lý, chẳng hạn như lắp đặt hệ thống dây điện trong một ngôi nhà mới.

Một số nhà nghiên cứu quan điểm, các hệ thống AI cũng sẽ không thể giải quyết các vấn đề đòi hỏi sự hiểu biết sâu sắc trong những năm tới đây.

Theo ước tính từ các chuyên gia, có 50% khả năng AI vượt trội hơn con người trong mọi việc, từ giải quyết các vấn đề phức tạp đến đưa ra các quyết định quan trọng, thông qua sáng tạo và đổi mới nghệ thuật vào năm 2047.

Đến thế kỷ 22, tất cả công việc sẽ được AI tự động hóa hoàn toàn, khả năng xảy ra điều này là 50%.

Tuy nhiên, những xác suất này dựa trên giả định rằng, sự phát triển của AI tuân theo một tốc độ nhất định và có nhiều đổi mới hơn trong ngành.

Sự tuyệt chủng của loài người do AI gây ra?

Gần 58% nhà nghiên cứu được khảo sát tin rằng, có 5% khả năng AI có thể dẫn đến sự tuyệt chủng của loài người, đặc biệt là sau khi trí tuệ nhân tạo tổng hợp (AGI-Artificial General Intelligence) ra đời.

Đây là một dạng AI rất tiên tiến, vượt xa khả năng trí tuệ và ra quyết định của con người. Vì vậy, các nhà nghiên cứu này tin rằng, nếu công nghệ như vậy thoát khỏi sự kiểm soát của chúng ta, nó có thể gây ra hậu quả nghiêm trọng.

Những kết quả này có thể được xem là tích cực hoặc tiêu cực, tùy theo quan điểm của mỗi người.

Điều quan trọng là chúng ta phải đưa ra các hướng dẫn đạo đức chặt chẽ hơn, cũng như quy định của chính phủ để kiểm soát sự phát triển của AI theo cách tối đa hóa lợi ích, trong khi rủi ro được giảm thiểu.

Hơn 70% nhà nghiên cứu được khảo sát lo ngại về các rủi ro khác nhau liên quan đến AI (thậm chí không tính đến những rủi ro của siêu AI).

Họ lo sợ nhiều kịch bản kịch tính khác nhau, chẳng hạn như việc sử dụng hàng loạt công nghệ để thao túng dư luận, phát triển vũ khí hóa học hoặc tự động.