Năm nguy cơ của trí tuệ nhân tạo ("AI") theo Geoffrey E. Hinton, một trong những người tiên phong trong lãnh vực" AI".

Huỳnh Thiện Quốc Việt dịch

Nguồn: Les 5 dangers de l'IA selon Geoffrey Hinton, un de ses pionniers, Le Point, ngày 02/05/2023.

06/6/2023

Tiến sĩ Geoffrey E. Hinton năm 2019. © MASAHIRO SUGIMOTO / Yomiuri / The Yomiuri Shimbun theo AFP

Geoffrey Hinton, người tiên phong trong lĩnh vực trí tuệ nhân tạo, đã từ chức tại Google vào hôm thứ Hai ngày 1 tháng 5 sau 10 năm phục vụ cho gã công nghệ khổng lồ Mỹ. Trong một cuộc phỏng vấn với tờ New York Times, ông cho biết đã rời Google để nói chuyện mà không bị ràng buộc về AI sản sinh như ChatGPT, mà không gây ảnh hưởng đến công ty. Ông bày tỏ sự hối tiếc về việc tham gia vào quá trình phát triển ChatGPT, và cố gắng tự an ủi bằng cách nào đó "với lý do phổ biến này: nếu tôi không làm thì cũng có người khác làm".

Tốt nghiệp ngành tâm lý học thực nghiệm và là tiến sĩ về trí tuệ nhân tạo, người gốc London này đã dành cả cuộc đời để nghiên cứu về AI. Dù với tư cách là nhà nghiên cứu tại Carnegie Mellon, thuộc Đại học Toronto hay tại Google – công ty mà ông đã tham gia khi gia nhập nhóm Google Brain vào năm 2013 – Geoffrey Hinton đã phát triển các mạng thần kinh nhân tạo – những hệ thống toán học biết tiếp thu những kĩ năng mới bằng cách phân tích các dữ liệu mà chúng được tiếp cận. Năm 2018, ông đã được trao tặng Giải thưởng Turing vì sự phát triển các mạng mà ông đã sáng tạo. Được coi là một nhân vật quan trọng trong lĩnh vực AI, ông đã truyền cảm hứng cho các công trình của Yann Le Cun và Yoshua Bengio.

Với sự hối tiếc, ông quay trở lại bàn đến sự nguy hiểm của AI, mà theo ông, sẽ tượng trưng cho "những rủi ro nghiêm trọng đối với xã hội và nhân loại", theo lời giải thích của ông trong cuộc phỏng vấn của New York Times. Dưới đây là năm mối lo lớn nhất của ông.

Khi cỗ máy vượt mặt người tạo ra nó

Theo Geoffrey Hinton, sự cạnh tranh giữa các đại gia công nghệ đã dẫn đến những tiến bộ mà không ai có thể tưởng tượng được. Tốc độ mà những tiến bộ đang diễn ra đã vượt quá mong đợi của các nhà khoa học: "Chỉ một số ít người tin vào ý tưởng cho rằng công nghệ này thực sự có thể trở nên thông minh hơn con người […] và bản thân tôi cho rằng điều đó sẽ xảy ra […] từ nay đến 30-50 năm nữa, thậm chí hơn thế nữa. Tất nhiên, tôi không nghĩ như vậy nữa."

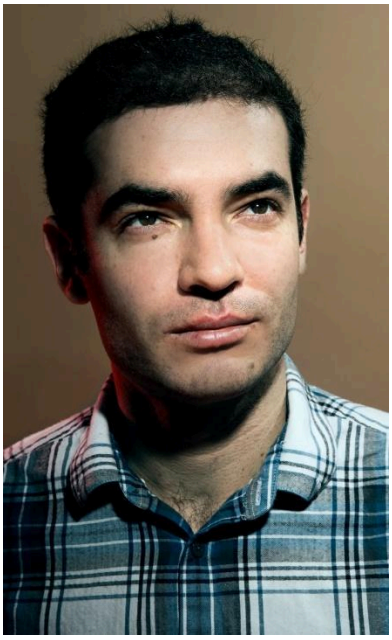
Cho đến năm ngoái, ông không coi các tiến bộ đó là nguy hiểm, nhưng quan điểm của ông đã thay đổi khi Google và OpenAI phát triển các hệ thống thần kinh có khả năng xử lý một lượng dữ liệu rất lớn, và có khả năng làm cho các hệ thống đó trở nên hiệu quả hơn bộ não con người, và vì vậy rất nguy hiểm.

Cắt giảm việc làm trên mọi phương diện

Từ quan điểm kinh tế, người đỡ đầu của AI lo ngại rằng công nghệ này có thể làm rối loạn mạnh thị trường lao động. Ông nói: "AI loại bỏ những công việc nặng nhọc," trước khi cho biết thêm rằng "nó có thể loại bỏ nhiều thứ hơn thế nữa", bao gồm cả những người phiên dịch và người trợ lý cá nhân.

Việc cắt giảm việc làm sẽ không chừa bất kỳ đối tượng nào "thông minh" nhất, bất chấp những gì nhiều người nghĩ, tin rằng mình có thể tránh khỏi.

Nguy cơ từ những "robot giết người"



Ilya Sutskever (1986-)

Người từng là giáo sư của Ilya Sutskever – giám đốc khoa học hiện tại của OpenAI – cho rằng sự tiến bộ công nghệ đang diễn ra quá nhanh, so với các phương tiện mà ta có để điều tiết việc sử dụng AI. Nhà nghiên cứu khẳng định: "Tôi không nghĩ ta nên tăng tốc việc sử dụng AI khi chưa hiểu liệu có thể kiểm soát được AI hay không." Ông lo sợ rằng các phiên bản trong tương lai sẽ trở thành "mối đe dọa cho nhân loại".

Theo ông, AI trong tương lai sẽ có thể phát triển những hành vi không mong đợi sau khi phân tích một lượng lớn dữ liệu. Điều này có thể trở nên khả dĩ, vì các hệ thống AI tạo ra mã riêng của chúng và điều khiển các mã đó… thứ có thể biến chúng thành những "vũ khí tự chủ" và "rô-bốt giết người", mặc dù đã có nhiều chuyên gia tương đối hoá mối đe dọa này.

AI trong tay những kẻ xấu

Theo ông, mối đe dọa còn đến từ việc lạm dụng AI của những kẻ nguy hiểm. Ông lo lắng: "Thật khó để tìm ra cách ngăn chặn kẻ xấu sử dụng nó vì những mục đích ác ý". Geoffrey Hinton đặc biệt phản đối việc sử dụng trí tuệ nhân tạo trong lĩnh vực quân sự. Ông đặc biệt lo sợ sự phát triển "những người lính robot" do con người tạo ra. Vào những năm 1980, ông thậm chí đã quyết định rời Đại học Carnegie Mellon ở Hoa Kỳ vì các nghiên cứu của ông ở đó đã được Lầu Năm Góc tài trợ.

"Thứ tạo ra điều nhảm nhí"

Mối đe dọa cuối cùng, và không kém phần quan trọng: đó là thông tin sai lệch gắn liền với AI. Sự sôi động khoa học này, cùng với việc sử dụng ồ ạt AI, sẽ khiến chúng ta gần như không thể phân biệt được "điều gì là đúng với điều gì không còn đúng nữa". Nhà khoa học thậm chí còn nói đến "thứ tạo ra điều nhảm nhí". Một cách phát biểu để chỉ khả năng của AI trong việc đưa ra một cách hợp lý những tuyên bố có vẻ hợp lý nhưng không vì thế mà đúng sự thật.

Vậy giải pháp là gì? Chuyên gia về mạng thần kinh ủng hộ một sự điều tiết liên hành tinh, nhưng ông cũng tỉnh táo cho rằng: "Đó có thể là điều bất khả […]. Cũng giống đối với như vũ khí hạt nhân, không có cách nào để biết liệu có công ty nào hay quốc gia nào đang bí mật chế tạo chúng hay không. Hy vọng duy nhất là các nhà khoa học vĩ đại nhất trên thế giới hợp tác với nhau để tìm ra giải pháp kiểm soát [AI]."

http://www.phantichkinhte123.com