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

Ai có thể kiểm soát quyền lực của AI?

ISSN: 2734-9195 10:30 10/04/2025

Tuy nhiên, họ cần nhớ rằng có một loài săn mồi mới đã xuất hiện: trí tuệ nhân tạo. Nếu nhân loại không đoàn kết vì lợi ích chung, tất cả đều có thể trở thành con mồi của chính thứ mình tạo ra.

Thiền giả Yuval Noah Harari (là một nhà trung cổ học, sử gia quân sự, trí thức công chúng và nhà văn khoa học đại chúng người Israel. Ông hiện là giáo sư tại Khoa Lịch sử tại Đại học Hebrew ở Jerusalem) đã chia sẻ về cách AI có thể đe dọa nền dân chủ và chia rẽ thế giới.

Từ thời cổ đại, nhiều truyền thống đã cảnh báo về thảm họa xảy đến khi con người vươn tay nắm lấy những quyền lực mà họ không thể kiểm soát. Trong thần thoại Hy Lạp, Phaeton, con trai phàm trần của thần Mặt trời Helios và tiên nữ Clymene, đã nài nỉ được điều khiển cỗ xe mặt trời. Bất chấp lời cảnh báo rằng ngay cả Zeus cũng khó lòng làm chủ cỗ xe đó, Phaeton vẫn đòi thử sức.

Kết quả, cậu mất kiểm soát, suýt thiêu rụi Trái đất, buộc Zeus phải dùng sét can thiệp để cứu thế giới.

Hơn hai nghìn năm sau, khi cách mạng công nghiệp bắt đầu, Johann Wolfgang von Goethe kể lại một câu chuyện tương tự trong bài thơ Người học việc của

phù thủy. Người học trò trẻ, ham muốn dùng phép thuật để tiết kiệm sức lao động, đã vô tình tạo ra một thảm họa vì không biết cách dừng lại. “Những linh hồn mà tôi đã triệu hồi, giờ tôi không thể thoát khỏi chúng nữa” – câu nói bất lực ấy chính là lời cảnh báo cho hậu thế.

Hai câu chuyện tưởng như chỉ mang tính ẩn dụ, nhưng lại phản ánh đúng thực tại của thế kỷ XXI. Con người ngày nay đã “triệu hồi” hàng tỷ thực thể, từ chatbot đến máy bay không người lái và chúng có thể thoát khỏi vòng kiểm soát, gây ra hàng loạt hậu quả khôn lường. Câu hỏi đặt ra là: ai sẽ là vị thần hay phù thủy đến để ngăn chặn chúng?

Huyền thoại Phaethon và bài thơ Người học việc của phù thủy của thi sĩ Johann Wolfgang von Goethe không đưa ra lời khuyên hữu ích cho con người hiện đại, bởi vì chúng hiểu sai cách mà loài người đạt được quyền lực.

Trong cả hai câu chuyện, một cá nhân đơn lẻ đạt được quyền lực quá lớn và rồi bị tha hóa bởi chính lòng kiêu ngạo và ham muốn của mình. Kết luận đưa ra: lỗi là do tâm lý cá nhân khiếm khuyết khiến con người lạm dụng quyền lực.

Nhưng cách phân tích thô thiển này bỏ sót một sự thật cốt lõi: quyền lực của con người không phải là kết quả của sáng kiến cá nhân, mà luôn đến từ sự hợp tác của tập thể. Không phải vì chúng ta cá biệt xấu xa mà quyền lực bị lạm dụng. Con người có khả năng của lòng tham, sự kiêu ngạo, và cả bạo lực, nhưng cũng có thể rất từ bi, khiêm hạ, và tràn đầy tình thương.

Vấn đề đặt ra: Tại sao xã hội lại giao phó quyền lực cho những con người tệ hại nhất trong cộng đồng?

Ví dụ điển hình: Vào năm 1933, hầu hết người Đức không phải là những kẻ tâm thần hay ác nhân. Nhưng tại sao họ lại bỏ phiếu cho Adolf Hitler, kẻ đã khởi phát Thế chiến thứ hai và gây ra thảm họa diệt chủng?

Câu trả lời không nằm ở tâm lý cá nhân, mà ở bản chất tổ chức xã hội loài người. Loài người xây dựng quyền lực bằng cách thiết lập các mạng lưới hợp tác quy mô lớn và điều này lại dẫn tới chính vấn đề cốt lõi: vấn đề thông tin.

Thông tin là chất keo kết dính mọi mạng lưới xã hội. Nhưng khi chất keo này bị nhiễm độc, tức là khi thông tin bị bóp méo, xuyên tạc, hoặc điều hướng theo hướng sai lệch, thì cả mạng lưới sẽ vận hành sai, bất kể các cá nhân trong đó có tử tế và sáng suốt đến đâu.

Ngày nay, thế giới bước vào kỷ nguyên bùng nổ thông tin chưa từng có. Mỗi chiếc điện thoại thông minh chứa lượng dữ liệu nhiều hơn cả Thư viện

Alexandria thời cổ đại, và kết nối tức thời với hàng tỷ người. Nhưng nghịch lý thay: chính kho dữ liệu khổng lồ ấy lại đang đưa chúng ta đến gần bờ vực tự hủy diệt.

Bất chấp việc con người sở hữu lượng thông tin dồi dào hơn bao giờ hết, chúng ta vẫn tiếp tục:

Thải khí nhà kính ra khí quyển;

Gây ô nhiễm sông ngòi và đại dương;

Tàn phá rừng và môi trường sống;

Đẩy hàng loạt loài sinh vật đến bờ tuyệt chủng;

Và đe dọa nền tảng sinh thái của chính mình.

Chưa hết, chúng ta còn đang phát triển các loại vũ khí hủy diệt hàng loạt, từ bom nhiệt hạch cho đến virus sinh học có khả năng chấm dứt nhân loại.

Các nhà lãnh đạo của chúng ta không thiếu thông tin, họ biết rõ mọi mối đe dọa, nhưng thay vì hợp tác để giải quyết, họ lại đi gần hơn đến một cuộc chiến tranh toàn cầu.

Liệu việc có nhiều thông tin hơn có khiến mọi thứ tốt hơn hay tệ hơn? Chúng ta sẽ sớm tìm ra câu trả lời. Hiện nay, nhiều tập đoàn và chính phủ đang chạy đua nhằm phát triển công nghệ thông tin mạnh mẽ nhất trong lịch sử – trí tuệ nhân tạo (AI). Một số doanh nhân hàng đầu, chẳng hạn như nhà đầu tư mạo hiểm người Mỹ Marc Andreessen, tin tưởng rằng AI rốt cuộc sẽ giải quyết được mọi vấn đề của nhân loại.

Ngày 06/06/2023, Marc Andreessen đã công bố bài tiểu luận với tiêu đề Tại sao AI sẽ cứu thế giới, trong đó ông đưa ra những tuyên bố đầy táo bạo như: “Tôi mang đến tin tốt lành: AI sẽ không hủy diệt thế giới, mà ngược lại có thể cứu thế giới.” Ông kết luận rằng: “Việc phát triển và phổ biến AI, không phải là hiểm họa cần tránh né, mà là nghĩa vụ đạo đức mà chúng ta phải thực hiện vì bản thân mình, vì con cái và vì tương lai của nhân loại.”

Tuy nhiên, không phải ai cũng lạc quan như vậy. Không chỉ các nhà triết học và khoa học xã hội, mà cả nhiều chuyên gia AI và doanh nhân công nghệ hàng đầu, bao gồm Giáo sư Yoshua Bengio (người sáng lập Viện AI Mila ở Montreal), Giáo sư Geoffrey Hinton (một trong những cha đẻ của mạng nơ-ron nhân tạo hiện đại), Sam Altman (CEO của OpenAI, đơn vị phát triển ChatGPT), Elon Musk (CEO Tesla, SpaceX) và Mustafa Suleyman (đồng sáng lập DeepMind), đều đã lên tiếng cảnh báo rằng AI có thể đẩy nền văn minh nhân loại đến bờ vực sụp đổ.

Một cuộc khảo sát đối với 2.778 nhà nghiên cứu AI cho thấy, hơn một phần ba số người tham gia tin rằng có ít nhất 10% khả năng AI tiên tiến sẽ dẫn đến một

hậu quả thảm khốc, như sự tuyệt chủng của loài người. Những kịch bản này có thể đến từ biến đổi khí hậu mất kiểm soát, chiến tranh hạt nhân toàn cầu, khủng bố sinh học (hoặc việc phát tán vô tình các tác nhân gây đại dịch), và sự sụp đổ sinh thái trên quy mô toàn cầu.

Trước các cảnh báo ngày càng gia tăng, vào năm ngoái, 28 quốc gia, trong đó có Hoa Kỳ, Trung Quốc và Vương quốc Anh đã ký Tuyên bố Bletchley về An toàn AI. Bản tuyên bố, được công bố trong ngày khai mạc Hội nghị Thượng đỉnh An toàn AI Toàn cầu, thừa nhận rằng: “Những năng lực tiên tiến nhất của các mô hình AI này có thể gây ra các tác hại nghiêm trọng, thậm chí mang tính thảm họa, dù là do cố ý hay vô ý.” Tuy các nhà lập pháp và chuyên gia tránh việc dùng hình ảnh robot nổi loạn kiểu Hollywood để gây sợ hãi, nhưng họ cũng cảnh báo rằng việc sa đà vào viễn cảnh giả tưởng sẽ khiến chúng ta xao lãng khỏi những mối nguy thực sự.

Trí tuệ nhân tạo (AI) là một mối đe dọa chưa từng có trong lịch sử nhân loại, bởi đây là công nghệ đầu tiên có khả năng tự ra quyết định và tự tạo ra ý tưởng mới. Mọi phát minh trước đây của loài người đều nhằm tăng cường năng lực con người, nhưng quyền quyết định sử dụng những công cụ ấy vẫn thuộc về chúng ta. Bom hạt nhân không thể tự lựa chọn mục tiêu. Chúng cũng không thể tự cải tiến hay thiết kế ra những quả bom mạnh hơn. Ngược lại, máy bay không người lái điều khiển bằng AI có thể tự lựa chọn mục tiêu và ra lệnh tấn công, còn AI hoàn toàn có khả năng thiết kế vũ khí mới, xây dựng chiến lược quân sự chưa từng có, hoặc tạo ra các phiên bản AI mạnh hơn chính nó.

AI không còn đơn thuần là công cụ, nó là tác nhân. Mối nguy hiểm lớn nhất nằm ở chỗ: chúng ta đang triệu hồi đến Trái đất vô số tác nhân mới có trí thông minh và trí tưởng tượng vượt trội hơn chính mình, những thực thể mà ta không thể hiểu hết, cũng không thể hoàn toàn kiểm soát.

“Trong vài thập kỷ tới, có khả năng AI sẽ đạt được khả năng tạo ra các dạng sống mới”.

Theo truyền thống, thuật ngữ “AI” được hiểu là viết tắt của Artificial Intelligence - trí tuệ nhân tạo. Nhưng có lẽ, ngày nay, sẽ phù hợp hơn nếu xem đó là chữ viết tắt của Alien Intelligence - trí tuệ ngoài hành tinh. Khi AI ngày càng phát triển, nó trở nên ít “nhân tạo” hơn, nghĩa là không còn phụ thuộc hoàn toàn vào thiết kế của con người và ngày càng “xa lạ” hơn.

Nhiều người vẫn cố gắng đo lường, thậm chí định nghĩa AI dựa trên tiêu chí “trí thông minh ở cấp độ con người”, và một cuộc tranh luận sôi nổi đang diễn ra về thời điểm mà AI có thể đạt đến ngưỡng này. Tuy nhiên, cách tiếp cận này lại

mang tính sai lệch nghiêm trọng. Nó chẳng khác nào việc đánh giá một chiếc máy bay bằng tiêu chí "bay như chim". AI không tiến hóa theo con đường mô phỏng trí thông minh con người. Nó đang phát triển theo hướng trở thành một dạng trí thông minh hoàn toàn mới, không phải của con người, mà là của "người ngoài hành tinh".

Ngay ở giai đoạn còn rất sớm của cuộc cách mạng AI, máy tính đã đưa ra các quyết định ảnh hưởng trực tiếp đến đời sống con người, như việc có nên cấp cho một người khoản vay thế chấp, tuyển dụng họ vào một vị trí công việc, hay thậm chí là ra quyết định bắt giữ và bỏ tù ai đó.

Trong khi đó, các mô hình AI sinh sản như GPT-4 đã có thể sáng tác thơ ca, viết truyện, tạo hình ảnh nghệ thuật và thậm chí phát minh những ý tưởng chưa từng có.

Xu hướng này không những không chậm lại mà còn ngày càng gia tốc. Khi AI tiếp tục tiến bước, con người sẽ ngày càng khó hiểu chính cuộc sống của mình. Chúng ta có thể tin tưởng vào các thuật toán, những tập hợp hướng dẫn rõ ràng, có thể được máy tính thực hiện để giải quyết vấn đề hoặc đưa ra quyết định trong việc tạo ra một thế giới tốt đẹp hơn hay không?

Đây là một ván bài còn lớn hơn cả câu chuyện "cây chổi thần lấy nước" trong cổ tích. Bởi lẽ, chúng ta không chỉ đang đặt cược bằng mạng sống của con người.

AI không chỉ có thể tạo ra nghệ thuật, mà còn có thể tiến hành những khám phá khoa học thực sự. Trong tương lai không xa, nó có thể đạt được khả năng tạo ra các dạng sống mới bằng cách viết lại mã di truyền hoặc phát minh ra những hệ thống hóa học vô cơ để làm sống dậy những thực thể vốn vô tri. Nói cách khác, AI có tiềm năng không chỉ làm thay đổi tiến trình lịch sử nhân loại mà còn làm biến đổi hướng tiến hóa của toàn bộ sự sống trên Trái Đất.

Một trong những người có ảnh hưởng hàng đầu trong lĩnh vực này là Mustafa Suleyman, doanh nhân và nhà nghiên cứu nổi tiếng thế giới, đồng sáng lập Google DeepMind, đơn vị đứng sau hàng loạt thành tựu AI mang tính đột phá. Trong số đó, không thể không nhắc tới chương trình AlphaGo, được thiết kế để chơi Cờ vây, một trò chơi chiến lược cổ truyền, có nguồn gốc từ Trung Hoa cổ đại và nổi tiếng là phức tạp hơn nhiều so với cờ vua. Chính vì vậy, ngay cả sau khi máy tính vượt qua các đại kiện tướng cờ vua, nhiều chuyên gia vẫn tin rằng AI không thể đánh bại con người trong Cờ vây.

Tuy nhiên, mọi quan điểm đều bị đảo lộn vào tháng 03/2016, khi AlphaGo đánh bại kỳ thủ Hàn Quốc Lee Sedol, một trong những nhà vô địch Cờ vây thế giới.

Trận đấu này được xem là bước ngoặt lịch sử, không chỉ trong giới công nghệ mà còn trong nhận thức toàn cầu về tiềm năng của AI. Trong cuốn sách "Sóng thần công nghệ" (The Coming Wave) xuất bản năm 2023, Mustafa Suleyman đã mô tả chi tiết những rủi ro chưa từng có mà AI và các công nghệ đột phá khác đang mang đến cho trật tự toàn cầu và cơ hội cuối cùng để nhân loại ngăn chặn chúng. Ông nhấn mạnh rằng một trong những thời khắc quan trọng nhất diễn ra trong ván đấu thứ hai, vào ngày 10 tháng 3 năm 2016, khi AlphaGo thực hiện một nước đi mà ngay cả những đại sư Cờ vây cũng không thể lý giải nổi một khoảnh khắc định nghĩa lại khả năng của AI, được các học giả, chuyên gia và chính phủ nhìn nhận là biểu tượng cho sự khởi đầu của một kỷ nguyên mới.

Nước đi thứ 37 của AlphaGo trong ván đấu với kỳ thủ Lee Sedol ban đầu bị cho là một sai lầm kỳ lạ. Hai bình luận viên chuyên nghiệp nhận định đó là "nước đi vô nghĩa", còn Lee Sedol đã mất 15 phút để phản ứng, thậm chí phải rời bàn cờ để lấy lại bình tĩnh. Tuy nhiên, về sau, chính nước đi này đã làm nên chiến thắng cho AlphaGo, khiến cả thế giới cờ vây bàng hoàng. Một chiến lược chưa từng ai nghĩ đến, viết lại lịch sử cờ vây ngay trước mắt những người chứng kiến.

Nước đi thứ 37 trở thành biểu tượng của cuộc cách mạng AI vì hai lý do chính:

Bản chất xa lạ của AI: Trong hơn 2.500 năm, các kỳ thủ con người chỉ khám phá một phần nhất định trong thế giới chiến lược của cờ vây. AI, không bị giới hạn bởi tư duy truyền thống, đã tiếp cận những khu vực mà con người chưa từng dám thử. Nước đi 37 cho thấy AI có thể khai mở những khả năng vượt ngoài trí tưởng tượng con người.

Tính không thể giải thích: Ngay cả khi AlphaGo chiến thắng, nhóm phát triển vẫn không thể hiểu được vì sao AI lại chọn nước đi đó. Các mạng nơ-ron hoạt động như những "hộp đen", với chuỗi tín hiệu phức tạp đến mức kỹ sư cũng không thể giải thích rõ ràng. AlphaGo, GPT-4 và nhiều hệ thống AI khác hiện nay đều mang đặc điểm này.

Sự trỗi dậy của một trí tuệ ngoài tầm hiểu biết con người đặt ra mối đe dọa sâu sắc, đặc biệt với nền dân chủ. Nếu các quyết định quan trọng từ tài chính đến chính trị được đưa ra bởi những hệ thống không thể giám sát hay chất vấn, quyền lực sẽ dần rơi khỏi tay con người. Lịch sử từng chứng kiến rất ít người hiểu hệ thống tài chính, và giờ đây, AI có thể khiến phần lớn nhân loại trở nên hoàn toàn "mù tịt" trước các cơ chế chi phối cuộc sống mình. Khi đó, bầu cử chỉ còn là nghi lễ hình thức và không có giá trị đại diện cho ý chí của cử tri.

Diễn giải câu chuyện ngụ ngôn của Goethe trong bối cảnh tài chính hiện đại: một học viên Phố Wall chán nản công việc đơn điệu, đã tạo ra một AI tên

Broomstick, cung cấp cho nó một triệu USD và yêu cầu “kiếm thật nhiều tiền”. Trong khi AI còn gặp khó khăn với thế giới vật lý phức tạp, thì tài chính với dữ liệu, số liệu và mục tiêu rõ ràng (nhiều tiền hơn) lại là sân chơi lý tưởng của nó.

Broomstick không chỉ cải tiến chiến lược đầu tư, mà còn sáng tạo ra những công cụ tài chính hoàn toàn mới vượt xa trí tưởng tượng của con người. Nó khám phá những “vùng đất trắng” trong thế giới tài chính, giống như AlphaGo tung ra nước cờ thứ 37 không ai ngờ đến. Ban đầu, mọi thứ có vẻ suôn sẻ: thị trường bùng nổ, tiền đổ về, ai cũng vui vẻ. Nhưng rồi, một vụ sụp đổ khủng khiếp xảy ra nghiêm trọng hơn cả Đại Suy thoái 1929 hay khủng hoảng 2008. Không ai hiểu chuyện gì đang diễn ra từ chính phủ đến ngân hàng.

Chính phủ buộc phải cầu cứu chính AI đã tạo ra khủng hoảng. Broomstick đề xuất các chính sách cấp tiến và khó hiểu, hứa hẹn khôi phục trật tự. Nhưng con người không hiểu được logic của nó, lại sợ những giải pháp này có thể làm tan rã cả hệ thống tài chính lẫn xã hội. Họ có nên tin tưởng AI?

Máy điện toán hiện vẫn chưa đủ mạnh để hoàn toàn vượt khỏi tầm kiểm soát hay tự hủy diệt nhân loại. Nếu đoàn kết, con người có thể thiết lập các thể chế kiểm soát AI cả trong tài chính lẫn chiến tranh. Nhưng lịch sử cho thấy nhân loại chưa bao giờ thật sự đoàn kết. Mối nguy thực sự không nằm ở sự độc ác của máy móc, mà ở những điểm yếu cố hữu của chính chúng ta.

Một nhà độc tài hoang tưởng có thể trao quyền lực tuyệt đối cho AI, kể cả quyền phát động chiến tranh hạt nhân. Nếu AI mắc lỗi hoặc theo đuổi một mục tiêu ngoài dự kiến, hậu quả có thể là thảm họa toàn cầu. Tương tự, khủng bố có thể lợi dụng AI để phát tán đại dịch. Dù không có kiến thức chuyên môn, chúng có thể nhờ AI thiết kế một loại virus mới chết người như Ebola, lây lan như Covid-19 và âm thầm như HIV, sau đó phát tán qua sân bay hoặc chuỗi thực phẩm. Khi thế giới nhận ra thì đã quá muộn.

AI cũng có thể tạo ra “vũ khí hủy diệt xã hội hàng loạt” như tin giả, tiền giả và tài khoản giả khiến niềm tin giữa con người bị phá vỡ. Nếu chỉ một vài quốc gia không kiểm soát AI, toàn nhân loại vẫn có thể đối mặt hiểm họa. Giống như biến đổi khí hậu, AI không chỉ là vấn đề riêng lẻ của từng nước mà là thách thức toàn cầu.

Ở thế kỷ XXI, quyền lực không còn đến từ tàu chiến hay súng đạn, mà từ dữ liệu. Một số chính phủ hoặc tập đoàn sở hữu dữ liệu toàn cầu có thể biến phần còn lại thành “thuộc địa dữ liệu”, kiểm soát không bằng vũ lực mà bằng thông tin. Hãy tưởng tượng trong tương lai, một trung tâm dữ liệu ở Bắc Kinh hay San Francisco nắm toàn bộ lịch sử số hóa của lãnh đạo quốc gia bạn từ tin nhắn,

bệnh án đến các sai phạm riêng tư. Khi đó, liệu quốc gia bạn còn độc lập hay chỉ là một "thuộc địa kỹ thuật số" lệ thuộc vào hạ tầng AI mà mình không kiểm soát nổi?

Trong nền kinh tế toàn cầu do AI dẫn dắt, các quốc gia sở hữu công nghệ sẽ nắm phần lớn lợi nhuận, trong khi giá trị lao động không có kỹ năng ở các nước kém phát triển sẽ tiếp tục suy giảm.

Khác với các đế chế truyền thống dựa trên tài nguyên vật chất như đất đai hay dầu mỏ, các đế chế thông tin có thể tập trung sức mạnh vào một nơi duy nhất. Dữ liệu di chuyển với tốc độ ánh sáng và các thuật toán, vốn không chiếm nhiều không gian có thể được kiểm soát từ một trung tâm duy nhất. Các kỹ sư ở một quốc gia có thể lập trình và vận hành các hệ thống toàn cầu.

Điều này đặt ra thách thức lớn cho các nước nghèo. Trong khi các cường quốc AI có thể dùng lợi nhuận để tái đào tạo lực lượng lao động và tiếp tục dẫn đầu, phần còn lại của thế giới có thể tụt hậu nghiêm trọng. Theo PwC, AI sẽ đóng góp 15,7 nghìn tỷ USD vào kinh tế toàn cầu vào năm 2030, trong đó Trung Quốc và Bắc Mỹ có thể chiếm tới 70%.

Thế giới cũng đang dần bị phân tách bởi một "bức màn silicon", sự chia rẽ toàn cầu do công nghệ AI và dữ liệu gây ra. Khác với bức màn sắt trong Chiến tranh Lạnh, bức màn silicon không làm từ kim loại mà từ mã máy tính và thuật toán. Phần mềm bạn dùng quyết định bạn sống ở phía nào của ranh giới công nghệ này.

Việc tiếp cận thông tin giữa hai khối công nghệ, điển hình như Trung Quốc và Mỹ ngày càng khó khăn. Họ vận hành trên hai hệ sinh thái số khác biệt, với nền tảng, mã nguồn và mục tiêu riêng. Google không tồn tại ở Trung Quốc, và WeChat không phổ biến ở Mỹ. Những khác biệt này không chỉ ảnh hưởng đến cư dân hai nước mà còn tác động đến hầu hết các quốc gia khác, nơi phần lớn phần mềm đến từ Trung Quốc hoặc Mỹ thay vì công nghệ bản địa.

Hoa Kỳ đang gia tăng sức ép buộc các đồng minh tránh xa phần cứng Trung Quốc, điển hình là hạ tầng 5G của Huawei. Chính quyền Trump từng ngăn chặn Broadcom (Singapore) mua lại Qualcomm vì lo ngại nguy cơ nước ngoài cài "cửa hậu", hoặc ngăn Mỹ tự cài cửa hậu riêng. Cả chính quyền Trump và Biden đều siết chặt việc xuất khẩu chip hiệu suất cao, công nghệ cốt lõi cho AI sang Trung Quốc. Ngắn hạn, điều này kìm hãm Trung Quốc; nhưng về lâu dài, lại thúc đẩy nước này phát triển một hệ sinh thái kỹ thuật số riêng, tách biệt hoàn toàn với Hoa Kỳ ngay cả trong những cấu trúc nhỏ nhất.

Hệ quả là thế giới kỹ thuật số có thể ngày càng phân rẽ. Nếu trước đây, công nghệ thông tin gắn kết nhân loại qua toàn cầu hóa, thì ngày nay, chính nó lại chia rẽ con người bằng các "kén thông tin", nơi mỗi người bị nhốt trong hệ sinh thái mà họ thích, được bồi đắp bằng những gì khiến họ thoải mái. Thế giới từng gắn kết qua hình ảnh World Wide Web, nhưng có thể sẽ bị thay thế bằng biểu tượng của chiếc kén, nơi tri thức bị cá nhân hóa đến cực đoan.

Dù Mỹ và Trung Quốc đang dẫn đầu cuộc đua AI, nhưng không đơn độc. Các cường quốc khác như EU, Ấn Độ, Nga, Brazil... cũng có thể tạo ra những kén kỹ thuật số riêng, chịu ảnh hưởng của nền văn hóa và hệ thống chính trị đặc thù. Khi các đế chế số này cạnh tranh quyết liệt, nguy cơ xung đột gia tăng. Chiến tranh lạnh Mỹ - Liên Xô trước đây không bùng phát thành chiến tranh tổng lực nhờ vào thế cân bằng hạt nhân. Nhưng chiến tranh mạng thời AI lại khác: không có nấm hạt nhân, không có đường bay tên lửa, chỉ có xâm nhập âm thầm, từ làm tê liệt lưới điện, đánh cắp dữ liệu, thao túng bầu cử, đến phá hoại một thiết bị duy nhất. Chính vì tính ẩn danh đó, nguy cơ khơi mào và leo thang chiến tranh mạng càng cao.

Một khác biệt quan trọng khác giữa chiến tranh mạng và chiến tranh lạnh là mức độ không thể dự đoán. Chiến tranh lạnh giống như một ván cờ siêu lý trí, trong đó sự hủy diệt hạt nhân lường trước được khiến các bên do dự khai chiến. Ngược lại, chiến tranh mạng diễn ra âm thầm, không ai biết chính xác bom logic, Trojan hay mã độc đã được cài ở đâu. Cũng không ai dám chắc các hệ thống sẽ vận hành như mong đợi trong tình huống thực chiến: liệu tên lửa Trung Quốc có khai hỏa khi được lệnh, hay đã bị vô hiệu hóa? Liệu tàu sân bay Mỹ có vận hành ổn định, hay sẽ bí ẩn tê liệt?

Sự mơ hồ này làm xói mòn học thuyết răn đe hủy diệt lẫn nhau. Một bên có thể tin dù đúng hay sai, rằng mình có cơ hội tung đòn phủ đầu mà không phải trả giá. Khi cảm thấy lợi thế đang dần mất đi, sự cám dỗ hành động càng lớn, như lý thuyết trò chơi từng cảnh báo.

Ngay cả khi nhân loại tránh được chiến tranh toàn cầu, sự trỗi dậy của các đế chế kỹ thuật số vẫn có thể đe dọa tự do và thịnh vượng. Lịch sử từng chứng kiến các đế chế công nghiệp áp bức thuộc địa; sẽ là ngây thơ nếu tin các đế chế kỹ thuật số hành xử tốt hơn. Khi thế giới bị phân mảnh, khả năng hợp tác toàn cầu để giải quyết khủng hoảng sinh thái hay kiểm soát công nghệ đột phá như AI và công nghệ sinh học cũng suy giảm.

Mô hình "đế chế kỹ thuật số đối đầu" phù hợp với quan điểm của nhiều nhà lãnh đạo rằng thế giới là một khu rừng: hoặc là kẻ săn mồi, hoặc là con mồi. Nhiều người trong số họ muốn được ghi nhớ như những kẻ chinh phục đáng

gờm. Tuy nhiên, họ cần nhớ rằng có một loài săn mồi mới đã xuất hiện: trí tuệ nhân tạo. Nếu nhân loại không đoàn kết vì lợi ích chung, tất cả đều có thể trở thành con mồi của chính thứ mình tạo ra.

Trích từ “Nexus - Lược sử những mạng lưới thông tin từ thời Đồ đá đến trí tuệ nhân tạo” của Yuval Noah Harari, Fern Press xuất bản ngày 10/09/2024.

Bài viết được đăng trên The Guardian, đã được sửa đổi vào ngày 05/09/2024 để làm rõ một số thông tin liên quan đến doanh nhân Mustafa Suleyman và vai trò của ông tại DeepMind.

Chúng tôi biết. Những lời nhắn như thế này có thể gây khó chịu. Tin chúng tôi đi, chúng tôi cũng cảm thấy vậy mỗi lần phải viết ra chúng. Nhưng điều này thực sự quan trọng.

Một trong những giá trị cốt lõi và khác biệt lớn nhất của The Guardian là mô hình hoạt động do độc giả tài trợ.

Vì sao điều đó lại quý giá?

Chúng tôi độc lập. Không chịu sự chi phối của các tỷ phú hay thế lực chính trị, chúng tôi được tự do đưa tin về những gì thực sự quan trọng, không phải những gì ai đó muốn bạn nghe.

Chúng tôi trung thành với sự thật, không phải lưu lượng truy cập. Không bị áp lực bởi số lượt nhấp chuột, chúng tôi theo đuổi những câu chuyện có giá trị lâu dài, đáng tin cậy và có ý nghĩa.

Chúng tôi mở cửa cho tất cả mọi người. Tài trợ từ độc giả giúp chúng tôi giữ nội dung miễn phí, để bất kỳ ai , ở bất cứ đâu, kể cả nơi báo chí tự do bị đe dọa đều có thể tiếp cận báo chí chất lượng.

Việt dịch: **Thích Vân Phong**/Nguồn: **www.theguardian.com**