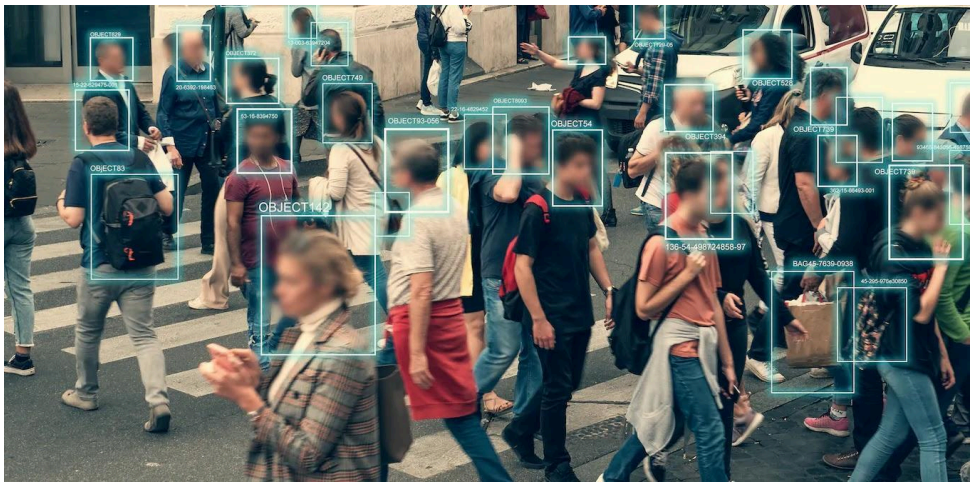
EU phê duyệt dự thảo luật để điều tiết AI – Luật sẽ hoạt động như sau

Nguồn: EU approves draft law to regulate AI – here's how it will work, *The Conversation,* Jun

14, 2023.

Huỳnh Thị Thanh Trúc dịch

11/8/2023



Đạo luật của EU nhìn chung sẽ cấm sử dụng tính năng nhận dạng khuôn mặt theo thời gian thực tại các không gian công cộng. DedMityay/Shutterstock

Ngày nay, từ "rủi ro" thường xuất hiện trong cùng một câu với từ "trí tuệ nhân tạo". Mặc dù thật đáng khích lệ khi thấy các nhà lãnh đạo thế giới xem xét các vấn đề tiềm ẩn của AI, cùng với các lợi ích chiến lược và công nghiệp của nó, chúng ta nên nhớ rằng không phải mọi rủi ro đều như nhau.

Vào thứ Tư, ngày 14 tháng 6, Nghị viện châu Âu đã bỏ phiếu thông qua đề xuất dự thảo của riêng mình cho Đạo luật AI [AI Act], một bộ luật được xây dựng trong hai năm, với tham vọng định hình các tiêu chuẩn toàn cầu trong quy định về AI.

Sau giai đoạn đàm phán cuối cùng, để dung hòa các dự thảo khác nhau do Nghị viện, Ủy ban và Hội đồng châu Âu đưa ra, đạo luật cần được thông qua trước cuối năm nay. Đây sẽ là

quá trình ban hành luật đầu tiên trên thế giới dành riêng cho việc điều chỉnh AI trong hầu hết các lĩnh vực của xã hội – dù quốc phòng sẽ được miễn trừ.

Trong tất cả các hướng có thể dùng để tiếp cận quy định về AI, điều đáng chú ý là việc ban hành luật này hoàn toàn dựa trên khái niệm rủi ro. Không phải bản thân AI đang được điều tiết, mà là cách nó được sử dụng trong các lĩnh vực cụ thể của xã hội, mỗi lĩnh vực đều có những vấn đề tiềm ẩn khác nhau. Bốn loại rủi ro, tùy thuộc vào các nghĩa vụ pháp lý khác nhau, gồm: không thể chấp nhận được, cao, có hạn và tối thiểu.

Các hệ thống được coi là gây ra mối đe dọa đối với các quyền cơ bản hoặc các giá trị của EU, sẽ được phân loại là có "rủi ro không thể chấp nhận được" và bị cấm. Một ví dụ về rủi ro như vậy sẽ là các hệ thống AI được sử dụng cho "chính sách dự đoán". Đây là việc sử dụng AI để đánh giá rủi ro của các cá nhân, dựa trên thông tin cá nhân, để dự đoán liệu họ có khả năng phạm tội hay không.

Một trường hợp gây tranh cãi hơn là việc sử dụng công nghệ nhận dạng khuôn mặt trên nguồn dữ liệu thu trực tiếp từ camera đường phố. Điều này cũng đã được thêm vào danh sách các rủi ro không thể chấp nhận được và sẽ chỉ được phép thực hiện sau khi có hành vi phạm tội và được phép của tòa án.

Những hệ thống được phân loại là "rủi ro cao" sẽ phải tuân theo nghĩa vụ công bố và dự kiến sẽ được đăng ký trong cơ sở dữ liệu đặc biệt. Chúng cũng sẽ phải tuân theo các yêu cầu giám sát hoặc kiểm toán khác nhau.

Các loại ứng dụng được phân loại là rủi ro cao bao gồm AI có thể kiểm soát quyền truy cập vào các dịch vụ về giáo dục, việc làm, tài chính, chăm sóc sức khỏe và các lĩnh vực quan trọng khác. Việc sử dụng AI trong những lĩnh vực như vậy không phải là điều không ai muốn, nhưng việc giám sát là cần thiết vì nó có khả năng ảnh hưởng tiêu cực đến sự an toàn hoặc các quyền cơ bản.

Ý tưởng ở đây là chúng ta có thể tin tưởng rằng bất kỳ phần mềm nào đưa ra quyết định về khoản thế chấp của chúng ta sẽ được kiểm tra cẩn thận về việc tuân thủ luật pháp châu Âu để

đảm bảo ta không bị phân biệt đối xử dựa trên các đặc trưng được bảo vệ như giới tính hoặc dân tộc – ít nhất là nếu ta sống ở EU.

Ursula von der Leyen đã đề xuất đạo luật này vào năm 2019. Olivier Hostel/EPA

Các hệ thống AI có "nguy cơ hạn chế" sẽ phải tuân theo các yêu cầu tối thiểu về tính minh bạch. Tương tự, những người vận hành hệ thống AI tạo sinh – chẳng hạn như bot tạo ra văn bản hoặc hình ảnh – sẽ phải công bố rằng người dùng đang tương tác với máy.

Trong hành trình dài thông qua các cơ quan châu Âu, bắt đầu từ năm 2019, đạo luật ngày càng trở nên cụ thể và rõ ràng về những rủi ro tiềm ẩn khi triển khai AI trong các tình huống nhạy cảm – cùng với cách giám sát và giảm thiểu những rủi ro này. Còn nhiều việc phải làm, nhưng ý tưởng rất rõ ràng: chúng ta cần phải cụ thể hóa nếu muốn được việc.

Nguy cơ tuyệt chủng?

Ngược lại, gần đây chúng ta đã thấy các kiến nghị kêu gọi giảm thiểu "nguy cơ tuyệt chủng" được cho là do AI gây ra mà không cung cấp thêm thông tin chi tiết gì. Nhiều chính trị gia đã lặp lại những quan điểm này. Rủi ro chung và rất dài hạn này hoàn toàn khác với những gì định hình Đạo luật AI, bởi vì nó không cung cấp bất kỳ chi tiết nào về những gì chúng ta nên tìm kiếm, cũng như những gì chúng ta nên làm ngay lúc này để tránh gặp rủi ro.

Nếu “rủi ro” là “tác hại trong dự kiến” có thể đến từ một thứ gì đó, thì chúng ta nên tập trung vào các kịch bản vừa nguy hại vừa có khả năng xảy ra, bởi vì những tình huống này có rủi ro cao nhất. Các sự kiện rất khó xảy ra, chẳng hạn như một vụ va chạm tiểu hành tinh, không nên được ưu tiên hơn những sự kiện có thể xảy ra hơn, chẳng hạn như các tác động của ô nhiễm.

Các rủi ro chung chung, chẳng hạn như khả năng dẫn đến sự tuyệt chủng của loài người, không được đề cập trong đạo luật. Zsolt Biczo/Shutterstock

Theo lẽ đó, dự thảo luật vừa được quốc hội EU thông qua tuy không thu hút sự chú ý bằng một số cảnh báo gần đây về AI nhưng có thực chất hơn. Nó cố gắng vượt qua ranh giới mong manh giữa việc bảo vệ các quyền và các giá trị mà không ngăn cản sự đổi mới, đồng thời nhắm cả những mối nguy hiểm lẫn biện pháp khắc phục một cách cụ thể. Mặc dù không hoàn hảo nhưng ít nhất nó cũng cung cấp các hành động thực chất.

Giai đoạn tiếp theo trong hành trình của luật này sẽ là “đối thoại ba bên” trong đó các dự thảo riêng biệt của quốc hội, ủy ban và hội đồng sẽ được hợp nhất thành một văn bản cuối cùng. Các thỏa hiệp dự kiến sẽ diễn ra trong giai đoạn này. Luật thu được sẽ được bỏ phiếu thông qua để đưa vào thi hành, có thể là vào cuối năm 2023, trước khi chiến dịch vận động cho các cuộc bầu cử châu Âu tiếp theo bắt đầu.

Sau hai hoặc ba năm, đạo luật sẽ có hiệu lực và bất kỳ doanh nghiệp nào hoạt động ở EU sẽ phải tuân thủ nó. Chính lịch trình kéo dài này sẽ đặt ra một số câu hỏi riêng, bởi vì chúng ta không biết AI hay thế giới sẽ trông như thế nào vào năm 2027.

Hãy nhớ rằng chủ tịch Ủy ban châu Âu, Ursula von der Leyen, lần đầu tiên đề xuất quy định này vào mùa hè năm 2019, ngay trước khi xảy ra đại dịch, chiến tranh và khủng hoảng năng lượng. Đây cũng là lúc ChatGPT chưa khiến các chính trị gia và giới truyền thông thường xuyên nói về rủi ro hiện sinh từ AI.

Tuy nhiên, đạo luật được viết theo một cách khá chung chung, điều có thể giúp nó duy trì tính phù hợp trong một thời gian. Đạo luật có thể sẽ ảnh hưởng đến cách các nhà nghiên cứu và doanh nghiệp tiếp cận AI bên ngoài châu Âu.

Dù vậy, rõ ràng là mọi công nghệ đều có rủi ro và thay vì chờ đợi điều gì đó tiêu cực xảy ra, các tổ chức học thuật và hoạch định chính sách đang cố gắng tính trước về các hệ quả của nghiên cứu. Nếu so với cách chúng ta áp dụng các công nghệ trước kia – chẳng hạn như nhiên liệu hóa thạch – điều này thể hiện một mức độ tiến bộ nhất định.

Tác giả:

Nello
Cristianini

Giáo sư Trí tuệ Nhân tạo, Đại học Bath

Tuyên bố công khai

Nello Cristianini là tác giả cuốn sách "The Shortcut: Why Intelligent Machines Do Not Think Like Us" [tạm dịch: Lối tắt: Tại sao máy móc thông minh không nghĩ giống chúng ta], do CRC Press xuất bản, năm 2023.

http://www.phantichkinhte123.com/2023/08