



Đạo luật Trí tuệ nhân tạo của châu Âu cấm một số ứng dụng của trí tuệ nhân tạo thu thập hình ảnh khuôn mặt không có chủ đích từ internet hoặc đoạn phim camera giám sát an ninh.

NGHỊ VIỆN CHÂU ÂU THÔNG QUA ĐẠO LUẬT VỀ TRÍ TUỆ NHÂN TẠO

“

Ngày 13/03/2024, các nghị sỹ của Nghị viện châu Âu (EP) đã bỏ 523 phiếu ủng hộ, 46 phiếu chống và 49 phiếu trắng, nhất trí thông qua Đạo luật Trí tuệ nhân tạo (Artificial Intelligence Act - AIA) nhằm đảm bảo sự an toàn và tuân thủ các quyền cơ bản trong tiến trình thúc đẩy đổi mới sáng tạo. Trước đó, toàn văn nội dung dự luật trên cũng đã được thống nhất sau các cuộc đàm phán giữa EP với tất cả các quốc gia thành viên vào ngày 09/12/2023. AIA gồm 459 trang, 13 chương, 113 điều và 13 phụ lục.

”

Một số nội dung quan trọng của Đạo luật Trí tuệ nhân tạo

Thông cáo báo chí gửi giới truyền thông của EP cho biết, AIA được xây dựng với mục đích bảo vệ các quyền cơ bản, dân chủ và pháp quyền cũng như sự bền vững môi trường khỏi AI có độ rủi ro cao, đồng thời thúc đẩy đổi mới sáng tạo và đưa châu Âu trở thành khu vực dẫn

đầu trong lĩnh vực AI. Quy định của AIA thiết lập các nghĩa vụ đối với AI dựa trên những rủi ro tiềm ẩn và mức độ ảnh hưởng. Bên cạnh đó, AIA cũng đề cập đến một số nội dung chính quan trọng khác như: các ứng dụng của AI bị cấm; miễn trừ việc thực thi pháp luật; những nghĩa vụ đối với các hệ thống AI có rủi ro cao; các yêu cầu về tính minh bạch.



Các ứng dụng của trí tuệ nhân tạo bị cấm

AIA cấm một số ứng dụng của AI đe dọa đến quyền công dân, bao gồm hệ thống phân loại sinh trắc học dựa trên các đặc điểm nhạy cảm và việc thu thập hình ảnh khuôn mặt không có chủ đích từ internet hoặc đoạn phim camera giám sát an ninh để tạo cơ sở dữ liệu nhận dạng khuôn mặt. Sự nhận dạng cảm xúc ở nơi làm việc và trường học, chấm điểm xã hội dựa trên hành vi xã hội hoặc đặc điểm cá nhân, chính sách dự đoán (khi AI chỉ dựa trên hồ sơ cá nhân hoặc đánh giá đặc điểm của họ) và AI thao túng hành vi con người hoặc khai thác lỗ hổng của con người sẽ đều bị cấm.

Miễn trừ việc thực thi pháp luật

Về nguyên tắc, việc sử dụng hệ thống nhận dạng sinh trắc học (biometric identification systems - RBI) của cơ quan thực thi pháp luật đều bị cấm, ngoại trừ trong các tình huống được liệt kê đầy đủ và được xác định ở phạm vi hẹp. RBI “thời gian thực” chỉ có thể được triển khai nếu đáp ứng các biện pháp bảo vệ nghiêm ngặt, ví dụ như việc sử dụng AI bị giới hạn về mặt thời gian và phạm vi địa lý cũng như phải tuân theo sự cho phép về mặt hành chính hoặc tư pháp cụ thể, chẳng hạn như tìm kiếm có chủ đích một người mất tích hoặc ngăn chặn một cuộc tấn công khủng bố. Việc sử dụng các hệ thống hậu thực tế (“RBI hậu từ xa”) được coi là trường hợp sử dụng có độ rủi ro cao, do vậy, phải có sự cho phép của cơ quan tư pháp vì có liên quan đến hành vi phạm tội hình sự.

Những nghĩa vụ đối với các hệ thống trí tuệ nhân tạo có rủi ro cao

AIA cũng quy định các nghĩa vụ rõ ràng đối với các hệ thống AI có rủi ro cao (do chúng có khả năng gây hại đáng kể đối với sức khỏe, sự an toàn, các quyền cơ bản, môi trường, dân chủ, pháp quyền). Việc sử dụng AI có rủi ro cao có thể bao gồm: cơ sở hạ tầng quan trọng, giáo dục - đào tạo nghề, việc làm, các dịch vụ công - tư thiết yếu (ví dụ như chăm sóc sức khỏe y tế, ngân hàng), một số hệ thống nhất định trong việc thực thi pháp luật, di cư và quản lý biên giới, các quy trình tư pháp và dân chủ (ví dụ như gây ảnh hưởng đến các cuộc bầu cử). Các hệ thống như vậy phải được đánh giá, giảm thiểu rủi ro và duy trì nhật ký sử dụng; cần có sự minh bạch, chính xác cũng như đảm bảo sự giám sát của con người. Mọi công dân sẽ có quyền gửi khiếu nại về các hệ thống AI và nhận được sự giải thích về các quyết định dựa trên các hệ thống AI có mức độ rủi ro cao, gây ảnh hưởng đến quyền lợi của họ.

Các yêu cầu về tính minh bạch

AIA cũng quy định, các hệ thống AI có mục đích chung (GPAI) - là một hệ thống AI có thể thực hiện các chức năng thường được áp dụng (như nhận dạng hình ảnh/giọng nói, tạo âm thanh/video, phát hiện mẫu, trả lời câu hỏi, dịch thuật...) thì cần đáp ứng các yêu cầu nhất định về tính minh bạch, bao gồm việc tuân thủ đạo luật về bản quyền của Liên minh châu Âu (EU) và công bố các bản tóm tắt chi tiết về nội dung được sử dụng cho hoạt động đào tạo. Các mẫu GPAI mạnh hơn có thể gây ra rủi ro mang tính hệ thống sẽ phải đối mặt với các yêu cầu bổ sung, bao gồm việc thực hiện nhận xét từng mẫu, đánh giá và giảm thiểu rủi ro của hệ thống cũng như báo cáo về sự cố. Ngoài ra, nội dung hình ảnh, âm thanh hoặc video giả tạo hoặc bị thao túng (“giả mạo”- deepfakes) cần phải được dán nhãn một cách rõ ràng.

Đạo luật đầu tiên trên thế giới về trí tuệ nhân tạo

Tại phiên tranh luận toàn thể tại nghị trường ngày 13/03/2024, Brando Benifei - Đảng S&D (Italy) và là thành viên Ủy ban thị trường nội bộ và bảo vệ người tiêu dùng phát biểu: “Cuối cùng thì chúng ta cũng sở hữu một đạo luật ràng buộc đầu tiên trên thế giới về AI nhằm giảm thiểu rủi ro, tạo cơ hội và chống phân biệt đối xử cũng như đem lại sự minh bạch; các hoạt động AI không được chấp nhận sẽ bị cấm ở châu Âu và quyền của mọi người lao động cũng như các công dân sẽ được bảo vệ. Văn phòng về AI sẽ được thành lập để hỗ trợ các công ty bắt đầu tuân thủ các quy tắc trước khi AIA có hiệu lực. Chúng tôi đảm bảo rằng, con người và các giá trị châu Âu là trung tâm của sự phát triển AI”.

Về những bước đi tiếp theo, thông cáo báo chí của EP cho biết, AIA vẫn phải trải qua cuộc kiểm tra lần cuối cùng về kỹ thuật, ngôn ngữ và dự kiến sẽ được thông qua trước khi cơ quan lập pháp kết thúc kỳ họp. Đạo luật này cũng cần phải được Hội đồng châu Âu (EC) chính thức thông qua. AIA sẽ có hiệu lực 20 ngày sau khi được công bố chính thức và được áp dụng đầy đủ sau 24 tháng kể từ khi chính thức có hiệu lực, ngoại trừ: lệnh cấm các hành vi thực tiễn bị cấm sẽ áp dụng 6 tháng sau ngày có hiệu lực; quy tắc thực hành (9 tháng sau khi có hiệu lực); các quy tắc AI có mục đích chung bao gồm quản trị (12 tháng sau khi có hiệu lực) và nghĩa vụ đối với các hệ thống có rủi ro cao (36 tháng sau khi có hiệu lực).

Trước đó, ngày 09/12/2023, các nhà đàm phán của EP và EC đã đạt được thỏa thuận tạm thời về dự luật. AIA được thiết kế nhằm đảm bảo các quyền cơ bản, dân

chủ, pháp quyền và tính bền vững của môi trường được bảo vệ khỏi AI có rủi ro cao, đồng thời thúc đẩy đổi mới sáng tạo và đưa châu Âu trở thành khối dẫn đầu thế giới trong lĩnh vực này. Các quy tắc thiết lập nghĩa vụ đối với AI dựa trên những rủi ro tiềm ẩn và mức độ tác động của nó.

AIA ra đời nhằm phản hồi trực tiếp đối với các đề xuất của công dân tại Hội nghị về tương lai châu Âu năm 2022 đối với việc cải cách EU (COFE), cụ thể nhất là đề xuất 12(10) về nâng cao khả năng cạnh tranh của EU trong các lĩnh vực chiến lược; đề xuất 33(5) về một xã hội an toàn và đáng tin cậy, bao gồm phòng chống các thông tin sai lệch và đảm bảo con người nắm quyền kiểm soát cuối cùng; đề xuất 35 về thúc đẩy đổi mới kỹ thuật số (điểm 3), đồng thời đảm bảo sự giám sát của con người (điểm 8) và việc sử dụng AI một cách đáng tin cậy, có trách nhiệm, thiết lập các biện pháp bảo vệ, đảm bảo tính minh bạch; đề xuất 37(3) về việc sử dụng các công cụ

AI và kỹ thuật số để cải thiện khả năng tiếp cận thông tin của người dân, bao gồm cả người khuyết tật.

Theo trang Business Insider ngày 14/03/2024, AIA là nỗ lực đầu tiên của cơ quan lập pháp của EU nhằm bảo vệ công dân của mình khỏi những rủi ro tiềm ẩn của công nghệ AI. Tuy nhiên, AIA cũng bị một số nhà bình luận đặt câu hỏi. Chuyên gia về AI Henry Ajder cho rằng, đây là bước đi “đầy tham vọng”, song ông cũng cảnh báo, AIA có nguy cơ khiến châu Âu kém cạnh tranh hơn trên toàn cầu. Trả lời phỏng vấn của Business Insider, ông Ajder cho biết thêm: Mỗi quan tâm của tôi là chúng ta thấy các công ty rõ ràng sẽ tránh phát triển ở một số nơi mà có những quy định rất chặt chẽ và toàn diện. Do đó, sẽ có những quốc gia gần như đóng vai trò là thiên đường cho phát triển AI - nơi mà họ tránh áp dụng luật pháp chặt chẽ về AI ✍

Hùng Tuấn (Hà Nội)



Ngày 13/03/2024, các nghị sỹ của Nghị viện châu Âu đã bỏ phiếu nhất trí thông qua Đạo luật Trí tuệ nhân tạo (Ảnh: EU/EP/UPI).