Đạo luật AI: Quy định về trí tuệ nhân tạo đầu tiên ở EU

**Ngày 14/5, Nghị viện châu Âu đã thông qua quan điểm đàm phán về Luật Trí tuệ nhân tạo (AI) với 499 phiếu thuận, 28 phiếu chống và 93 phiếu trắng. Các quy định nhằm khuyến khích việc áp dụng AI đáng tin cậy và lấy con người làm trung tâm, đồng thời bảo vệ “sức khỏe, sự an toàn, các quyền cơ bản và nền dân chủ khỏi những tác động có hại của AI”.**

**Xếp hạng AI theo “Mức độ rủi ro”**

Các quy tắc tuân theo cách tiếp cận dựa trên rủi ro và thiết lập nghĩa vụ cho các nhà cung cấp và những người triển khai hệ thống AI tùy thuộc vào mức độ rủi ro mà AI có thể tạo ra. Dựa trên mức độ này, những người vận hành AI tương ứng sau đó được trao quyền, đồng thời cũng có các nghĩa vụ nhất định. Do đó, các hệ thống AI có mức độ rủi ro không thể chấp nhận được đối với sự an toàn của mọi người sẽ bị cấm, chẳng hạn như những hệ thống được sử dụng để chấm điểm xã hội (phân loại mọi người dựa trên hành vi xã hội hoặc đặc điểm cá nhân của họ).

Các [ứng dụng AI](https://antoanthongtin.vn/hoat-dong---su-kien-attt/nhung-rui-ro-tiem-an-tu-tri-tue-nhan-tao-109005) xâm phạm quyền riêng tư và phân biệt đối xử sẽ bị cấm. Điều này áp dụng cho:

- Hệ thống nhận dạng sinh trắc học từ xa “thời gian thực” trong các không gian có thể truy cập công khai;

- Đăng các hệ thống nhận dạng sinh trắc học từ xa, ngoại trừ duy nhất việc thực thi pháp luật để truy tố các tội phạm nghiêm trọng và chỉ sau khi được tư pháp cho phép;

- Hệ thống phân loại sinh trắc học sử dụng các đặc điểm nhạy cảm (ví dụ: giới tính, chủng tộc, dân tộc, tình trạng công dân, tôn giáo, khuynh hướng chính trị);

- Hệ thống kiểm soát dự đoán (dựa trên hồ sơ, vị trí hoặc hành vi tội phạm trong quá khứ);

- Hệ thống nhận dạng cảm xúc trong thực thi pháp luật, quản lý biên giới, nơi làm việc và các cơ sở giáo dục;

- Chỉnh sửa không có mục tiêu các hình ảnh khuôn mặt từ internet hoặc đoạn phim từ camera quan sát (Closed Circuit Television - CCTV) để tạo cơ sở dữ liệu nhận dạng khuôn mặt (vi phạm nhân quyền và quyền riêng tư).

**Bốn nhóm rủi ro**

Đạo luật AI phân biệt các quy tắc cho các mức độ rủi ro khác nhau. Các quy tắc mới thiết lập nghĩa vụ cho nhà cung cấp và người dùng tùy thuộc vào mức độ rủi ro từ trí tuệ nhân tạo. Mặc dù nhiều hệ thống AI gây ra rủi ro tối thiểu, nhưng chúng cần phải được đánh giá.

Nhóm I. Rủi ro không thể chấp nhận được

Thuộc nhóm I, Hệ thống AI được coi là mối đe dọa đối với con người và sẽ bị cấm khi chúng có những biểu hiện sau:

- Thao túng hành vi nhận thức của con người hoặc các nhóm dễ bị tổn thương cụ thể: ví dụ như đồ chơi kích hoạt bằng giọng nói khuyến khích hành vi nguy hiểm ở trẻ em;

- Tính điểm xã hội: phân loại mọi người dựa trên hành vi, tình trạng kinh tế xã hội hoặc đặc điểm cá nhân;

- Hệ thống nhận dạng sinh trắc học từ xa và thời gian thực, chẳng hạn như nhận dạng khuôn mặt.

Nhóm II. Hệ thống AI rủi ro cao

Các hệ thống AI gây ảnh hưởng tiêu cực đến an toàn hoặc các quyền cơ bản sẽ được coi là có rủi ro cao và sẽ được chia thành hai loại:

- Các hệ thống AI được sử dụng trong các sản phẩm thuộc luật an toàn sản phẩm của EU. Điều này bao gồm đồ chơi, hàng không, ô tô, thiết bị y tế và thang máy.

- Các hệ thống AI thuộc tám lĩnh vực cụ thể sẽ phải đăng ký trong cơ sở dữ liệu của EU, bao gồm: Nhận dạng [sinh trắc học](https://antoanthongtin.vn/giai-phap-khac/mo-hinh-xac-thuc-sinh-trac-hoc-da-nhan-to-107788) và phân loại thể nhân; Quản lý và vận hành cơ sở hạ tầng quan trọng; Giáo dục và dạy nghề; Việc làm, quản lý người lao động và khả năng tự tạo việc làm; Tiếp cận và thụ hưởng các dịch vụ tư nhân thiết yếu và các dịch vụ và lợi ích công cộng; Thực thi pháp luật; Quản lý di cư, tị nạn và kiểm soát biên giới; Hỗ trợ giải thích pháp luật và áp dụng pháp luật.

Nhóm III. Trí tuệ nhân tạo

Nhóm AI, điển hình như ChatGPT, sẽ phải tuân thủ các yêu cầu về tính minh bạch:

- Tiết lộ rằng nội dung được tạo bởi AI;

- Thiết kế mô hình để ngăn nó tạo ra nội dung bất hợp pháp;

- Xuất bản tóm tắt dữ liệu có bản quyền được sử dụng cho đào tạo.

Nhóm IV. Rủi ro hạn chế

Các hệ thống AI có rủi ro hạn chế phải tuân thủ các yêu cầu tối thiểu về tính minh bạch để cho phép người dùng đưa ra quyết định sáng suốt. Sau khi tương tác với các ứng dụng, người dùng có thể quyết định xem họ có muốn tiếp tục sử dụng hay không. Người dùng nên được biết khi họ đang tương tác với AI. Điều này bao gồm các hệ thống AI tạo hoặc thao tác nội dung hình ảnh, âm thanh hoặc video, trong đó điển hình là deepfakes.

**Đạo luật AI: Quyền và Trách nhiệm**

Các nhà cung cấp mô hình nền tảng trong lĩnh vực AI sẽ phải đánh giá và giảm thiểu rủi ro có thể xảy ra (đối với sức khỏe, an toàn, quyền cơ bản, môi trường, dân chủ và pháp quyền) và đăng ký mô hình của họ trong cơ sở dữ liệu của EU trước khi phát hành trên thị trường EU. Các hệ thống AI dựa trên các mô hình như vậy, như ChatGPT sẽ phải tuân thủ các yêu cầu về tính minh bạch (tiết lộ rằng nội dung do AI tạo ra, đồng thời giúp phân biệt cái gọi là hình ảnh giả sâu với hình ảnh thật) và đảm bảo các biện pháp bảo vệ chống lại việc tạo ra nội dung bất hợp pháp. Các bản tóm tắt chi tiết về dữ liệu có bản quyền được sử dụng cho quá trình đào tạo của họ cũng sẽ phải được công khai.

Theo đó, ngay từ bây giờ các hệ thống AI tổng hợp như [ChatGPT](https://antoanthongtin.vn/hacker-malware/hon-100000-tai-khoan-chatgpt-bi-danh-cap-duoc-rao-ban-tren-cac-thi-truong-dark-web-109084) phải đáp ứng “các yêu cầu về tính minh bạch” và tiết lộ rằng nội dung được xác định là do AI tạo ra. Điều này sẽ giúp phân biệt ảnh deepfake với ảnh thật. Ngoài ra, các nhà cung cấp AI phải đảm bảo rằng không có nội dung bất hợp pháp nào được tạo ra. Các nhà khai thác AI phải công bố bản tóm tắt dữ liệu độc quyền mà họ đã sử dụng cho mục đích đào tạo. Người dùng có quyền và họ có thể gửi khiếu nại về các hệ thống AI.

*Tác giả: Nguyễn Ngoan*

Nguồn: https://antoanthongtin.vn/chinh-sach---chien-luoc/dao-luat-ai-quy-dinh-ve-tri-tue-nhan-tao-dau-tien-o-eu-109099