Một nhóm đã khai thác AI: Tội phạm

Từ phát tán thông tin sai lệch đến tăng cường bẻ khóa mật khẩu, tin tặc đang phát hiện ra sức mạnh của AI để khai thác độc hại.

Các nhà lập pháp vẫn đang tìm ra cách tốt nhất để sử dụng trí tuệ nhân tạo. Những kẻ vi phạm pháp luật cũng đang làm như vậy.

Việc sử dụng trí tuệ nhân tạo để làm hại mọi người đang gia tăng. Các quan chức đang cảnh báo chống lại các cuộc tấn công sử dụng công nghệ phát tán thông tin sai lệch (deepfake), các chiến dịch "lừa đảo" do AI tăng cường và phần mềm đoán mật khẩu dựa trên phân tích dữ liệu lớn.

"Theo dõi và trừng phạt tội phạm là nhiệm vụ của chúng tôi. AI là một thành phần của dịch vụ này," Philipp Amann, người đứng đầu chiến lược tại Trung tâm tội phạm mạng châu Âu (Europol của cơ quan thực thi pháp luật EU) cho biết. "Chúng tôi cũng sẽ có AI-for-crime-as-a-service." (Sử dụng AI để ngăn chặn tôi phạm như một dịch vụ)

Mối quan tâm lớn nhất đối với các quan chức an ninh mạng là công nghệ deepfake - sử dụng hàng loạt ảnh và video để phát triển những hình ảnh kỳ lạ hoặc hình đại diện hoàn toàn mới. Công nghệ này có sức mạnh tạo ra hình ảnh và video đánh lừa mọi người nghĩ rằng họ đang nhìn vào thực tế và đó chính xác là điều mà các chuyên gia an ninh mạng lo lắng.

Nếu tội phạm mạng "quản lý để đưa ra các cách giả định danh tính của bạn hoặc danh tính của tôi, hoặc ai đó để tạo ra những gì không tồn tại và sau đó họ quản lý để vượt qua các quy trình xác minh trực tuyến, đó là một rủi ro rất lớn," Amann nói.

"Một khi bạn đã phá vỡ quy trình, bạn có thể nhanh chóng tạo ra một số lượng lớn tài khoản", ông nói thêm rằng điều này sẽ giúp rửa tiền dễ dàng hơn và giúp bọn tội phạm thực hiện gian lận trên các nền tảng trực tuyến.

Trong một trường hợp, công nghệ deepfake được cho là đã được sử dụng để mạo danh một giám đốc điều hành cho các vụ lừa đảo lớn của công ty. Trong các trường hợp khác, các kỹ thuật đã được sử dụng để tạo hồ sơ giả mạo như một phần của các trò gian lận lừa đảo phức tạp.

Nỗi sợ bị deepfake lừa là rất hiện hữu, đôi khi nó đã được sử dụng làm vỏ bọc cho việc bị lừa theo những cách khác. Các chính trị gia trên khắp châu Âu đã đổ lỗi cho deepfake khi họ bị lừa tham gia các cuộc họp với một người đàn ông đóng giả là chánh văn phòng leonid Volkov của nhân vật đối lập Nga Alexei Navalny. Những kẻ chơi khăm Nga cho biết pha nguy hiểm là của họ và không liên quan đến công nghệ deepfake.

"Công nghệ nào được sử dụng ở mức độ nào, chúng tôi thường không biết. Nhưng chúng tôi liên tục tự hỏi bản thân và đặt câu hỏi về việc chúng tôi có thể tin tưởng ai", Agnes Venema, một nhà nghiên cứu công nghệ và an ninh quốc gia tại Đại học Malta cho biết.

"Đôi khi đó chỉ đơn giản nhằm mục đích để truyền bá những nghi ngờ và đặt ra câu hỏi," cô nói.

Mối đe dọa vượt ra ngoài deepfakes. Việc sử dụng trí tuệ nhân tạo một cách độc hại có thể bao gồm từ phần mềm độc hại được hỗ trợ bởi AI, khai thác các tài khoản mạng xã hội giả mạo được hỗ trợ bởi AI, các cuộc tấn công từ chối dịch vụ phân tán được hỗ trợ bởi AI, các mô hình tạo sâu để tạo dữ liệu giả mạo và bẻ khóa mật khẩu được hỗ trợ bởi AI, theo một báo cáo của cơ quan an ninh mạng của EU được công bố vào tháng 12.

Europol, cùng với công ty an ninh mạng Trend Micro và viện nghiên cứu UNICRI của Liên Hợp Quốc, đã tìm thấy phần mềm đoán mật khẩu dựa trên phân tích được hỗ trợ bởi AI về 1,4 tỷ mật khẩu bị rò rỉ, cho phép tin tặc truy cập vào hệ thống nhanh hơn.

Họ cũng tìm thấy các dịch vụ phần mềm giá rẻ có thể đánh lừa các nền tảng như dịch vụ phát trực tuyến và mạng truyền thông xã hội để tạo tài khoản bot thông minh. Ở Pháp, một nhóm các hãng âm nhạc độc lập, các hiệp hội sưu tầm và nhà sản xuất đang phàn nàn với chính phủ về "các luồng giả mạo", theo đó các bản nhạc được hiển thị là do bot phát hoặc người thật được thuê để tăng lượt xem một cách giả tạo, mang lại lợi ích cho nghệ sĩ có bản nhạc được phát.

Những kẻ lừa đảo khác đang phát triển các công cụ AI để tạo ra nội dung email "lừa đảo" giả mạo tốt hơn để lừa mọi người cung cấp thông tin đăng nhập hoặc thông tin ngân hàng.

"Càng ngày các công cụ này trở nên tốt hơn về chất lượng mỗi lần thực hiện. Ngoài ra còn có nhiều công cụ hơn có sẵn để phát hiện và phân tích [sử dụng AI một cách độc hại], nhưng câu hỏi đặt ra là liệu doanh nghiệp bình thường đã có quyền truy cập vào [những công cụ này]" khi họ chuẩn bị chống lại mối đe dọa mới, Marietje Schaake, giám đốc chính sách tại Trung tâm Chính sách Mạng tại Đại học Stanford và là cựu thành viên của Nghị viện châu Âu cho biết.

Tội phạm mạng "nhắm đến quy mô rộng lớn hơn", cô nói. Ví dụ, trong trường hợp email lừa đảo, "họ có thể gửi hàng triệu email và có thể thu lợi từ chỉ một trong số những email này."