Bốn cách tội phạm có thể sử dụng A.I

Bốn cách tội phạm có thể sử dụng AI để nhắm mục tiêu vào nhiều nạn nhân hơn (Four ways criminals could use A.I to target more victims).

Nguồn: The Conversation. (06/23)

Ngày nay, có thể thấy những cảnh báo về trí tuệ nhân tạo (Artificial Intelligence – AI) nhan nhản ở khắp mọi nơi, đi cùng với những thông điệp đáng sợ rằng A.I có thể đẩy nhân loại đến bờ vực diệt vong, kèm theo là những hình ảnh trong loạt phim Terminator. Thủ tướng Anh Rishi Sunak thậm chí đã thiết lập một hội nghị thượng đỉnh để thảo luận về sự an toàn của AI.

Tuy nhiên, nhóm của Giáo sư về An ninh mạng Daniel Prince đã sử dụng các công cụ A.I trong một thời gian dài – từ các thuật toán được sử dụng để đề xuất các sản phẩm có liên quan trên các trang web mua sắm, cho đến xe hơi có công nghệ nhận dạng biển báo giao thông và định vị làn đường. AI là một công cụ giúp tăng cường hiệu quả, giải quyết và sắp xếp khối lượng lớn dữ liệu, cũng như giảm bớt khó khăn cho quá trình ra quyết định…

Tuy nhiên, bởi vì nó chỉ là công cụ, cho nên ai cũng có thể sử dụng, kể cả người xấu. Và chúng ta đã thấy bọn tội phạm áp dụng A.I ở giai đoạn đầu. Thí dụ, công nghệ Deepfake đã được sử dụng để tạo ra các nội dung khiêu dâm giả tạo.

Công nghệ tăng cường hiệu quả của các hoạt động phạm pháp, giúp tội phạm nhắm mục tiêu vào nhiều người hơn và biến các mánh khóe lọc lừa trở nên hợp lý hơn. Quan sát cách bọn tội phạm từng thích nghi và áp dụng những tiến bộ công nghệ, chúng ta có thể tìm ra một số manh mối về cách chúng có thể sử dụng A.I.

1. Công cụ lừa đảo tốt hơn (A better phishing hook).

Thí dụ, các công cụ A.I như ChatGPT và Bard của Google cung cấp hỗ trợ kỹ năng viết, cho phép những người không có nhiều kinh nghiệm viết lách vẫn có thể tạo ra các thông điệp tiếp thị hiệu quả. Tuy nhiên, công nghệ này cũng có thể giúp tội phạm thêm phần đáng tin khi liên lạc với các nạn nhân.

Hãy nghĩ về tất cả những tin nhắn và email lừa đảo thường có nội dung dở tệ và dễ bị phát hiện. Cải trang hợp lý là chìa khóa để có thể thu thập thông tin từ nạn nhân.

Lừa đảo là trò chơi của những con số: ước tính có khoảng 3.4 tỷ email rác được gửi đi mỗi ngày. Nếu bọn tội phạm có thể cải thiện thông điệp của chúng sao cho chỉ 0.000005% số email có thể thuyết phục được ai đó tiết lộ thông tin, thì sẽ có thêm 6.2 triệu nạn nhân bị lừa đảo mỗi năm.

2. Tương tác tự động (Automated interactions).

Một trong những ứng dụng ban đầu của các công cụ A.I là tự động hóa các tương tác giữa khách hàng và nhà cung cấp dịch vụ qua tin nhắn, chat và điện thoại. Khách hàng nhận được phản hồi nhanh hơn, giúp tối ưu hóa hiệu quả kinh doanh. Lần tiếp

xúc đầu tiên của một người với một công ty, tổ chức có thể là với hệ thống AI, trước khi có thể nói chuyện với con người thật sự.

Tội phạm có thể sử dụng các công cụ tương tự để tạo tương tác tự động với số lượng nạn nhân tiềm năng ở quy mô rất lớn, vốn không thể thực hiện được nếu chỉ sử dụng sức người. Chúng có thể mạo danh các cơ quan như ngân hàng qua điện thoại và email, nhằm thu thập thông tin của nạn nhân rồi sử dụng để đánh cắp tiền của họ.

3. Giả Tạo Tinh Vi (Deepfakes).

A.I thực sự giỏi trong việc tạo ra các mô hình toán học có thể được "đào tạo" trên một lượng lớn dữ liệu trong thế giới thực, giúp các mô hình đó hoạt động tốt hơn trong một nhiệm vụ nhất định. Công nghệ giả tạo tinh vi (deepfake) trong video clip và âm thanh là một thí dụ dễ thấy nhất. Gần đây, Metaphysic đã chứng minh tiềm năng của công nghệ deepfake khi công bố một đoạn clip Simon Cowell hát opera trên chương trình truyền hình America's Got Talent.

Công nghệ này nằm ngoài tầm với của hầu hết tội phạm. Tuy nhiên, khả năng sử dụng A.I để bắt chước cách một người trả lời tin nhắn, viết email, để lại tin nhắn thoại hoặc thực hiện cuộc gọi, đều có sẵn miễn phí khi sử dụng A.I. Thí dụ, dữ liệu để đào tạo A.I có thể được thu thập từ các video trên mạng xã hội.

Phương tiện truyền thông xã hội luôn là kho tàng phong phú để tội phạm khai thác thông tin về các nạn nhân. A.I sẽ có thể được sử dụng để tạo ra một phiên bản deepfake của ai đó. Kỹ thuật Deepfake này có thể được sử dụng để tương tác với bạn bè và gia đình, thuyết phục họ cung cấp thông tin tội phạm về nạn nhân. Khi đã có được cái nhìn sâu sắc hơn về cuộc sống của nạn nhân, việc đoán mật mã hoặc gợi ý mật mã sẽ dễ dàng hơn nhiều.

4. Tấn công hàng loạt (Brute forcing).

Một kiểu tấn công khác được tội phạm ưng dùng có tên là "brute forcing," cũng có thể lợi dụng A.I. Nó sẽ thử lần lượt nhiều tổ hợp ký tự và biểu tượng để xem chúng có khớp với mật mã của nạn nhân hay không.

Đó cũng là lý do tại sao các mật mã dài, phức tạp lại an toàn hơn; chúng khó đoán hơn bằng phương pháp này. "Brute forcing" tốn nhiều tài nguyên, nhưng mọi việc sẽ dễ dàng hơn nếu có chút thông tin về nạn nhân. Thí dụ: nó sẽ cho phép danh sách các mật mã tiềm năng được sắp xếp theo mức độ ưu tiên – tăng hiệu quả của quy trình. Chẳng hạn như có thể bắt đầu với các kết hợp từ có liên quan đến tên của các thành viên trong gia đình hoặc thú cưng.

Các thuật toán được đào tạo dựa trên dữ liệu của nạn nhân có thể được sử dụng để giúp xây dựng các danh sách ưu tiên này chính xác hơn và nhắm mục tiêu nhiều người cùng một lúc – từ đó, chúng cần ít tài nguyên hơn để tấn công. Các công cụ AI cụ thể có thể được phát triển để thu thập dữ liệu trực tuyến của mọi người, sau đó phân tích tất cả để xây dựng hồ sơ về họ.

Thí dụ: nếu quý vị thường xuyên đăng bài trên mạng xã hội về Taylor Swift, thì việc ngồi rà lại từng bài đăng để tìm manh mối cho mật mã của quý vị sẽ là một công việc khó khăn. Nhưng với các công cụ tự động, việc này sẽ được thực hiện một cách nhanh chóng và hiệu quả cũng cao hơn. Tất cả thông tin thu được sẽ được sử dụng để tạo hồ sơ, giúp việc đoán mật mã và gợi ý dễ dàng hơn.

Tập thói quen cẩn trọng (Healthy scepticism).

Chúng ta không nên sợ A.I, vì nó có thể mang lại những lợi ích thực sự cho xã hội. Nhưng cũng như với bất kỳ công nghệ mới nào, xã hội cần phải hiểu rõ và thích nghi với nó. Mặc dù bây giờ chúng ta coi điện thoại thông minh (smartphone) là điều hiển nhiên, nhưng xã hội đã phải trải qua thời gian điều chỉnh để có thể thích nghi với cuộc sống có smartphone. Và dù smartphone mang lại nhiều lợi ích, nhưng vẫn còn đó những lo ngại, chẳng hạn như thời lượng sử dụng phù hợp cho trẻ em.

Điều chúng ta cần làm là chủ động tìm hiểu A.I, chứ không nên buông xuôi 'sống chung với lũ chúng.' Chúng ta nên phát triển các cách tiếp cận riêng đối với A.I, tập giữ thái độ hoài nghi, cẩn tắc như một thói quen. Chúng ta cũng cần xem xét cẩn trọng những gì mình đang đọc, nghe hoặc xem. Những việc đơn giản này sẽ giúp xã hội có thể 'tận hưởng' mọi lợi ích của A.I, đồng thời mọi người vẫn có thể tự bảo vệ mình khỏi những mối nguy hiểm chực chờ.

Các bài viết cùng tác giả hay cùng chủ đề (NnQ sưu tầm).

❖ **More Than Three Billion Fake Emails are Sent Worldwide Every Day.** Ít nhất 3,4 tỷ email giả mạo được gửi đi khắp thế giới mỗi ngày – với hầu hết các ngành công nghiệp vẫn dễ bị tấn công mạng lừa đảo và "giả mạo" đơn giản chỉ vì họ không triển khai các giao thức xác thực tiêu chuẩn ngành, theo Cảnh quan gian lận email mùa xuân 2019 của Valimail… **Đọc tiếp @ Security Magazine.**

❖ **Những kẻ tấn công sử dụng AI để tăng cường lừa đảo đàm thoại trên thiết bị di động.**

Các chuyên gia A.I, nhà báo, nhà hoạch định chính sách và công chúng đang ngày càng thảo luận về một loạt các rủi ro quan trọng và cấp bách từ AI. Mặc dù vậy, có thể khó nói lên mối lo ngại về một số rủi ro nghiêm trọng nhất của AI tiên tiến. Tuyên bố ngắn gọn dưới đây nhằm mục đích vượt qua trở ngại này và mở ra cuộc thảo luận. Nó cũng có nghĩa là tạo ra kiến thức chung về số lượng ngày càng tăng của các chuyên gia và nhân vật của công chúng, những người cũng coi trọng một số rủi ro nghiêm trọng nhất của A.I tiên tiến.

Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war (Giảm thiểu nguy cơ tuyệt chủng từ AI nên là ưu tiên toàn cầu bên cạnh các rủi ro quy mô xã hội khác như đại dịch và chiến tranh hạt nhân…) **Nguồn @ Center for A.I Safety**

❖ **A.I đang định hình cuộc chạy đua vũ trang an ninh mạng như thế nào (How A.I is shaping the cybersecurity arms race).** Một doanh nghiệp trung bình nhận được 10.000 cảnh báo mỗi ngày từ các công cụ phần mềm khác nhau mà họ sử dụng để giám sát những kẻ xâm nhập, phần mềm độc hại và các mối đe dọa khác. Nhân viên an ninh mạng thường thấy mình bị ngập trong dữ liệu họ cần sắp xếp để quản lý hệ thống phòng thủ mạng của họ… **Đọc tiếp @ The Conversation**

❖ **An Overview of Catastrophic A.I Risks.** Những tiến bộ nhanh chóng trong trí tuệ nhân tạo (AI) đã làm dấy lên mối lo ngại ngày càng tăng của các chuyên gia, các nhà hoạch định chính sách và các nhà lãnh đạo thế giới về tiềm năng của các hệ thống A.I ngày càng tiên tiến để gây ra những rủi ro thảm khốc… **Đọc tiếp @ Center for A.I Safety. (PDF File)**