HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
KHOA AN TOÀN THÔNG TIN

Học phần: Quản lý an toàn thông tin

Họ và tên: Hoàng Trung Kiên
MSV:B20DCAT098

Câu 1:

- Tấn công mạng dưa trên AI là các thuật toán AI độc hại có thể làm giảm hiệu suất và phá vỡ các chức năng bình thường của các thuật toán AI lành tính, đồng thời cung cấp các kịch bản tấn công biên công nghệ trong cả không gian mạng và không gian vật lý.

Tấn công dựa trên AI khác so với tấn công thông thường là:

- Tự động hóa các nhiệm vụ lặp đi lặp lại: AI có thể được sử dụng để tự động hóa các nhiệm vụ lặp đi lặp lại, chẳng hạn như phân tích dữ liệu hoặc gửi email lừa đảo. Điều này có thể giúp tin tặc thực hiện các cuộc tấn công với tốc độ và quy mô lớn hơn.
- Tăng cường hiệu quả của các cuộc tấn công: AI có thể được sử dụng để tăng cường hiệu quả của các cuộc tấn công, chẳng hạn như phát hiện các lỗ hổng bảo mật mới hoặc tạo ra các mã độc tinh vi hơn. Điều này có thể khiến các cuộc tấn công trở nên khó phát hiện và ngăn chặn hơn.
- Tốc độ: AI có thể tự động hóa các nhiệm vụ lặp đi lặp lại, giúp tin tặc thực hiện các cuộc tấn công với tốc độ nhanh hơn.

Câu 2:

- Tấn công phishing: Tin tặc sử dụng AI để tạo ra các email lừa đảo trông giống như được gửi từ một nguồn đáng tin cậy. Điều này có thể khiến người dùng dễ bị lừa cung cấp thông tin cá nhân hoặc tải xuống phần mềm độc hại.
- Tấn công ransomware: Tin tặc sử dụng AI để mã hóa dữ liệu của nạn nhân và yêu cầu tiền chuộc để giải mã. Điều này có thể gây ra tổn thất tài chính đáng kể cho nạn nhân.
- Tấn công từ chối dịch vụ (DDoS): Tin tặc sử dụng AI để gửi một lượng lớn lưu lượng truy cập đến một trang web hoặc dịch vụ nhằm làm cho nó không thể truy cập được. Điều này có thể gây gián đoạn nghiêm trọng cho hoạt động kinh doanh hoặc dịch vụ của nạn nhân.

Có một số công cụ hiện có có thể được sử dụng để thực hiện các cuộc tấn công mạng dựa trên trí tuệ nhân tạo. Một số công cụ phổ biến bao gồm:

- Deepfake: Deepfake là một kỹ thuật sử dụng trí tuệ nhân tạo để tạo ra video hoặc âm thanh trông giống như một người thật đang nói hoặc làm điều gì đó mà họ không bao giờ thực sự làm. Deepfake có thể được sử dụng để tạo ra các thông tin sai lệch hoặc truyền bá tin đồn.
- Machine learning: Machine learning có thể được sử dụng để phát hiện các lỗ hổng bảo mật, tạo ra mã độc tinh vi hoặc thực hiện các cuộc tấn công DDoS quy mô lớn.

- Natural language processing: Natural language processing có thể được sử dụng để tạo ra các email lừa đảo trông giống như được gửi từ một nguồn đáng tin cậy hoặc để phát hiện các mối đe dọa mạng dựa trên ngôn ngữ.

Câu 3:

Phần 6.2 của bài báo này nói về các cuộc tấn công mạng dựa trên AI đối với dữ liệu hình ảnh. Một số ví dụ và phương pháp tấn công được mô tả như sau:

- *Tấn công nhận dạng khuôn mặt*: Các nhà nghiên cứu đã sử dụng mạng sinh đối kháng có điều kiện để tạo ra các hình ảnh khuôn mặt giả, chẳng hạn như thay đổi tuổi, giới tính, hoặc biểu cảm của một người. Các hình ảnh giả này có thể được sử dụng để lừa qua các hệ thống xác thực dựa trên khuôn mặt hoặc tạo ra các tin đồn hoặc bằng chứng giả.
- *Deepfake*: Deepfake là một kỹ thuật sử dụng AI để tạo ra video hoặc âm thanh trông giống như một người thật đang nói hoặc làm điều gì đó mà họ không bao giờ thực sự làm. Deepfake có thể được sử dụng để tạo ra các thông tin sai lệch hoặc truyền bá tin đồn.
- *Face morphing*: Face morphing là một kỹ thuật sử dụng AI để kết hợp các đặc điểm khuôn mặt của hai người thành một khuôn mặt mới. Face morphing có thể được sử dụng để tạo ra các nhân vật hư cấu hoặc để tạo ra các hình ảnh của những người không tồn tại.
- Phương pháp tấn công:
    + Tuyên truyền và thông tin sai lệch: Deepfake có thể được sử dụng để tạo ra các video khiêu dâm hoặc bạo lực, hoặc để lừa đảo người dùng tin vào các tuyên bố sai lệch.
    + Lừa đảo và trộm cắp: face morphing có thể được sử dụng để tạo ra các hình ảnh giả mạo của người khác, chẳng hạn như để lừa đảo người dùng cung cấp thông tin cá nhân hoặc để sử dụng thẻ tín dụng của họ.
    + Face morphing có thể được sử dụng để tạo ra các hình ảnh của những người không tồn tại, chẳng hạn như để tạo ra các nhân vật hư cấu hoặc để tạo ra các hình ảnh có tính chất bạo lực hoặc khiêu dâm.