국가전략연구위원회 이슈브리핑 20호 (발간일: 2024.2.10.)

인공지능(AI)과 사이버안보[1)]

유지연 (상명대학교)

I. 인공지능(AI)은 사이버안보 차원에서 어떻게 접근해야 하는가?

인공지능(AI)은 우리 삶의 일부가 되었다. 인공지능(AI)는 기계 학습이나 생성 인공지능(AI)의 열기 속에서 이 시대의 중요한 이슈 중 하나가 되었고, 우리의 생산적인 구조와 우리의 삶을 변화시킬 것이다. 인공지능(AI)는 이미 기업과 시민들이 업무를 자동화하거나 의사결정을 최적화하기 위해 매일 사용하는 여러 장치와 기술에 존재한다.[2)]

이처럼 인공지능(AI) 시스템의 관련성과 기업 및 시민에 대한 잠재력이 높아짐과 동시에 인공지능(AI)의 보안 위험은 사이버안보 측면에서 중요한 문제가 되었다. 인공지능(AI) 시스템은 진화형 소프트웨어로 학습 과정에서 왜곡 또는 오류 등이 발생할 수 있으며 또한 인공지능(AI) 블랙박스 문제 등의 특수성으로 공격 표면을 더욱 복잡하게 만들고 넓힌다. 인공지능(AI) 시스템 특유의 보안 위험으로 인해 조직은 설계, 개발, 배포, 평가 및 사용과 같은 인공지능(AI) 라이프사이클의 모든 단계에서 사이버안보 및 개인 정보 위험을 관리하기 위한 전략을 수립해야 한다. 적절한 보안 제어를 구현하고 지속적인 보안 평가를 수행하여 수명 주기 전반에 걸쳐 인공지능(AI) 보안을 해결하는 것이 필수적이다.

인공지능(AI) 시스템의 라이프사이클은 일반적으로 데이터수집, 데이터전처리, 모델훈련, 모델추론, 시스템 통합의 5단계로 나눌 수 있으며, 각 단계는 다양한 보안 위협에 취약하다.

---

1) 이 글은 2024년 2월 6일 제7차 사이버 국가전략 포럼에서 발표된 발표문을 수정보완한 것이다.
2) Tarlogic, "What are the AI security risks?", 2023.09.07.
https://www.tarlogic.com/blog/ai-security-risks/

데이터 수집 단계에서의 보안 위협은 데이터 수집 수단과 밀접한 관련이 있다. 데이터 수집 방법에는 소프트웨어 기반 수집과 하드웨어 기반 수집이라는 두 가지 주요 유형이 있다. 소프트웨어 기반의 데이터 수집 방식은 주로 디지털데이터 수집을 의미하며, 그 보안 위험에는 데이터 편향, 가짜 데이터(fake data), 데이터 유출 등이 포함된다. 하드웨어 기반 수집 방식의 대표적인 공격 중 하나는 센서 스푸핑(Spoofing) 공격이다. 이는 공격자가 센서에서 제공하는 데이터에 접근하거나 변조하여 센서 공격을 수행한다.

데이터 전처리 단계에서의 대표적인 보안 위협은 스케일링 공격이다. 현재 스케일링 공격은 일반적으로 전처리단계에서 변조될 수 있는 이미지 데이터를 대상으로 하는 잠재적인 공격방식으로 추진된다. 특히, 일부 이미지 스케일링 공격(ISA)의 경우 공격자가 이미지를 변조하고 인간과 기계 사이의 (시각적)인지적 차이를 악용하여 세심한 수동 검사조차 우회하여 스푸핑 및 탈출 공격을 달성한다. 모델에 의존하는 적대적 예제 공격과 달리 ISA는 데이터전처리 단계만 대상으로 한다. 공격자는 대상 이미지와 공격 이미지 사이의 거리를 조절하여 공격 성공률을 높인다.

모델훈련 단계에서 모델에 유해 데이터를 주입하여 훈련된 모델을 변조함으로써 훈련 데이터와 훈련 프로세스에 영향을 미치는 원인공격(causative attacks)이 가능하다. 일반적으로 원인공격은 주로 데이터 중독 공격을 말하며 가용성 공격과 무결성 공격의 두 그룹으로 나누어진다. 가용성 공격의 경우 일반적으로 모델의 기울기 정보를 기반으로 중독지점을 찾거나 보조 네트워크를 사용하여 중독 데이터를 자동 생성한다. 가용성 공격은 모든 입력에 대한 모델의 전반적인 성능을 저하시킨다. 반면 무결성 공격은 정상적인 입력의 분류에는 영향을 미치지 않고 공격자가 선택한 입력에만 영향을 미친다. 백도어 공격과 클린라벨 중독 공격이 대표적인 무결성 공격이다.

일반적으로 모델추론 단계에서 수행되는 회피공격은 적대적인 예제를 만들어 모델의 예측성능을 저하하거나 방해한다. 이는 일반적으로 입력에 대한 사소하고 의미론적으로 일관된 변경을 통해 이루어지지만 대상 모델은 변경되지 않는다. 이 공격은 이미지 분류, 음성 인식 및 악성코드 탐지에서 광범위하게 연구되었다. 최근에는 FGSM(Fast Gradient Sign Method), JSMA(Jacobian-based Saliency Map Attack), DeepFool 등 다수의 적대적 예제 생성 전략이 개발되었다. 주로 최적화 검색이나 그래디언트 기반 정보를 통해 구현된다.

인공지능(AI) 시스템 통합 단계에서는 보안 문제가 다소 복잡해진다. 실제 응용 시나리오에서 인공지능(AI) 응용 프로그램의 시스템 통합에는 인공지능(AI) 기술 자체의 보안 위험뿐만 아니라 탑재된 시스템, 네트워크, 소프트웨어 및 하드웨어의 결합지점에서 발생하는 문제도 포함된다. 이러한 위협에는 인공지능(AI) 데이터 및 모델의 기밀성, 인공지능(AI) 편견, 코드 취약점 등이 포함된다.

II. 인공지능(AI) 사이버 위협 또는 방어 사례는 있는가?

악의적인 공격자가 인공지능(AI)를 사용한 사례로는 Youtube 동영상을 제작, 악성코드 감염을 실시하는 사례가 있다.[3] 2022년 11월 이후 사이버 공격자가 인공지능(AI) 기술을 활용해 생성한 YouTube 동영상을 악용하여 인포스틸러(infostealer) 악성코드 감염을 실시하고 있는 것으로 밝혀졌다. 인포스틸러는 컴퓨터에서 중요한 정보를 훔치도록 설계된 악성 소프트웨어이다. 그리고 단순히 동영상을 생성 및 게시하는 것에서 더 나아가 100K+ 구독자를 가진 인기 있는 YouTube 계정을 피싱 사기나 스틸러 로그 등에 의해 탈취하여 신뢰도를 높이고 가짜 튜토리얼 동영상을 업로드한다.[4] YouTube에서 검색을 했을 때, 어떤 탈취를 받은 계정이 가짜 동영상을 업로드하고 있을 가능성이 있어, 주의가 필요하다.

그리고 가족, 친지의 목소리를 인공지능(AI) 기술로 재생해 보이스피싱에 악용한 사례가 증가하고 있다. 2022년 미국에서 36,000명 이상이 해당 기술로 피해를 입었다. 이 중 5,100건이 전화통화를 통해 이루어졌으며 약 1,100만 달러의 피해를 야기했다.[5] 더 나아가 이미지 또는 비디오와 같은 기타 미디어를 생산하여 사기 행각을 벌이는 사례가 증가하고 다양해지고 있다. 딥페이크는 지난 몇 년간 배우 엠마 왓슨의 초상이 페이스북과 인스타그램에 등장한 일련의 외설적인 광고 에 사용되는 등 여러 가지 중요한 사건으로 악명을 얻었다. 또한 2022년에는 우크라이나 대통령 볼로디미르 젤렌스키가 우크라이나인들에게 "무기를 내려놓으라"고 말하는 모습이 담긴 영상이 널리 공유되었으며, 이는 사실이 아님이 폭로되었다.[6]

그리고 인공지능(AI)이 활용되는 대표 사례 중에 하나는 위협 탐지이다. 인공지능(AI)은 다양한 소스에서 얻은 대량의 데이터를 분석하고 사이버 공격을 나타낼 수 있는 사용자 행동의 비정상적인 패턴을 식별할 수 있다. 예를 들어, 직원이 자신도 모르게 피싱 이메일을 클릭하는 경우 인공지능(AI)는 신속하게 직원의 행동 변화를 알아차리고 잠재적인 안보 침해에 대해 경고할 수 있다.

인공지능(AI) 모델은 딥러닝 및 머신러닝 기술을 활용하여 네트워크 동작을 분석하고 표준에서 벗어난 편차를 지속적으로 감지한다. 시간이 지남에 따라 이러한 모델은 자체 수정 및 적응을 통해 이상 현상과 잠재적인 위협을 식별하는 정확도가 향상된다. 인공지능(AI)

---

3) Tech Plus, "AIが生成したYouTube動画でマルウェアに感染させるサイバー攻撃が増加", 2023.03.14. https://news.mynavi.jp/techplus/article/20230314-2624772/

4) CloudSEK, "Threat Actors Abuse AI-Generated Youtube Videos to Spread Stealer Malware", 2023.03.13. https://www.cloudsek.com/blog/threat-actors-abuse-ai-generated-youtube-videos-to-spread-stealer-malware

5) ZDnet, "Scammers are using AI to impersonate your loved ones. Here's what to watch out for", 2023.03.09.

6) The Conversation, "AI scam calls imitating familiar voices are a growing problem – here's how they work", 2023.07.12. https://theconversation.com/ai-scam-calls-imitating-familiar-voices-are-a-growing-problem-heres-how-they-work-208221

모델의 자체 수정 특성은 조직이 새로운 사이버 위협에 신속하게 대응할 수 있는 강력하고 안정적인 사이버안보 방어 메커니즘을 제공한다.

인공지능(AI)의 자동화 기능은 탐지를 넘어 다양한 벡터의 다양한 사이버 위협에 대한 자동 대응을 촉진한다. 조직은 인공지능(AI)로 강화된 사이버안보 솔루션을 활용하여 보안 팀의 워크로드 균형을 재조정하고 사고 대응 시간을 최적화할 수 있다.

인공지능(AI) 기반의 자동화된 위협 탐지 솔루션은 매일 수십억 개의 네트워크 요청, 엔드포인트 활동, 사용자 행동 및 데이터 포인트를 처리한다. 이러한 실시간 분석을 통해 몇 분 내에 즉각적인 조치가 가능하며, 수동 방법으로는 몇 시간 또는 며칠이 걸릴 수 있는 작업이 가능하다. IBM에 따르면 인공지능(AI)는 사이버 위협을 탐지하고 대응하는 데 걸리는 시간을 무려 14주나 단축할 수 있다.

인공지능(AI) 시스템은 정확하고 상세한 IT 자산 인벤토리를 제공함으로써 침해 위험 예측에 크게 기여한다. 이러한 인벤토리에는 중요한 시스템에 대한 다양한 액세스 수준을 가진 모든 장치, 사용자 및 애플리케이션이 포함된다. 인공지능(AI)는 자산 인벤토리 데이터와 위협 노출 평가를 결합하여 사이버 침해에 가장 취약한 영역을 예측할 수 있다.

다양한 데이터 소스를 수집하고 처리하는 인공지능(AI)의 기능을 통해 보안 팀은 조직의 보안 상태에 대한 전체적인 시각을 확보한다. 이렇게 강화된 상황 인식을 통해 사전 위협 탐지, 정확한 위험 평가, 시기적절한 사고 대응이 가능하며 조직의 전반적인 사이버안보 전략이 강화된다.

III. 인공지능(AI) 보안과 관련하여 주요국에서는 무엇을 추진하고 있는가?

미국은 인공지능(AI) 안전 및 보안에 대한 새로운 표준 수립을 위해 행정명령을 발표하였다('23.10.30.)[7]. 바이든 대통령은 해당 행정명령을 통해 인공지능(AI) 시스템의 잠재적인 위험으로부터 미국인들을 보호하기 위해 취해진 가장 포괄적인 행동을 지시하였다. 또한 미국 국립표준기술연구소(NIST)는 AML(적대적 기계 학습) 분야의 용어를 정의하고 개념 분류를 개발하며 인공지능(AI) 보안 위험을 관리하기 위한 프레임워크를 발간하였다. 주요 유형의 ML 방법과 공격의 수명주기 단계, 공격자의 목표 및 학습 프로세스에 대한 공격자의 능력과 지식을 포함하는 개념적 계층 구조로 배열되어 있다.[8]

영국 NCSC는 미국 CISA와 함께 2023년 11월, 개발자가 인공지능(AI) 시스템의 설계, 개발, 배포 및 운영에 대한 정보에 입각한 결정을 내리는 데 도움이 될 보안 인공지능(AI)

7) White House, "FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence", 223.10.30. https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

8) NIST AI 100-2e2023 ipd. 2024.01.

가이드라인, "인공지능(AI) 개발의 핵심 요건으로서의 보안(Guidelines for secure AI system development, 2023.)"을 발표하였다. 이는 국제기관과 협력하여 인공지능(AI)를 사용하는 모든 시스템의 제공자를 위한 가이드라인으로 해당 가이드라인을 통해 모든 이해 관계자(데이터 과학자, 개발자, 관리자, 의사 결정자 및 위험 소유자 포함)가 그들의 인공지능(AI) 시스템의 배치 및 운영. 설계, 개발에 대한 정보에 입각한 의사 결정을 내릴 수 있도록 촉구한다.[9)]

또한 영국은 최첨단의 '프론티어' 인공지능(AI)에 대한 위험을 해결하기 위해 포괄적이고 협력적인 조치가 필요하다는 것을 인정하며 2023년 11월 초 진행된 인공지능(AI) 안전정상회의(AI Safety Summit) 참석국들과 함께 '블레츨리 선언(Bletchley Declaration)'에 동의하였다. 이에 기반하여 근 시일 내 인공지능(AI)로 인해 발생할 수 있는 안보 위협에 대한 평가서를 발표하며[10)] 인공지능(AI)로 인한 위협은 지속적으로 증가할 것이며 인공지능(AI)로 인한 위협 행위자를 3가지 집단으로 구분, 각각의 역량 및 수준을 평가하였다. 위협 행위자의 유형은 고도의 능력을 가진 국가행위자(Highly capable state threat actors), 유능한 국가행위자 및 국가를 대상으로 하는 기업과 조직적인 사이버 범죄집단, 고용된 미숙련 해커와 기회주의적 사이버 범죄자 및 해킹 전문가 등이다.

중국은 핵심 싱크탱크인 DigiChina에서 '인공지능(AI) 보안 백서(2018)'를 발간하였다. 인공지능(AI) 보안 위험으로 네트워크 보안 위험, 데이터 보안 위험, 알고리즘 보안 위험, 정보보안 위험, 사회보장 위험, 국가안보 위험 등을 설정, 보안 애플리케이션와 보안 관리 등 3개의 축으로 '인공지능(AI) 보안 시스템 아키텍처'를 구성하였다.

유럽연합 사이버보안청(ENISA)은 인공지능(AI) 보안 위험과 기업이 직면해야 할 과제를 다루는 다양한 프레임워크, 방법론 및 보고서를 발표했다. 인공지능(AI)의 사이버안보와 관련된 표준(기존, 초안, 고려 중 및 계획 중)의 개요를 제공하고, 적용 범위를 평가하고 표준화의 격차를 식별하기 위해 "인공지능(AI)의 사이버 보안과 표준화(Cybersecurity of AI and Standardisation, 2023.03.)"를 발표하였다.[11)] 이어 2021년 이후 인공지능(AI) 사이버안보에 대해 꾸준히 연구하며 "인공지능(AI)와 사이버안보 연구(Artificial Intelligence and Cybersecurity Research, 2023.07.)" 보고서를 발간하였다. 이는 사이버안보를 위한 인공지능(AI) 및 인공지능(AI) 안보에 대한 연구에 대한 요구를 파악하고, 관련하여 이해 관계자들과 공유하고 논의할 5가지 주요 연구 요구를 파악하여 제시하였다.[12)]

추가적으로 OWASP(2023)[13)]는 [LLM] 인공지능(AI) 모델에 대한 공격의 유형을 다음과

9) NCSC, "Introducing the guidelines for secure AI", 2023.11.27.
https://www.ncsc.gov.uk/blog-post/introducing-guidelines-secure-ai-system-development
10) NCSC, "The near-term impact of AI on the cyber threat", 2024.01.24.
https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat
11) ENISA, "Cybersecurity of AI and Standardisation", 2023.03.14.
https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardisation
12) ENISA, "Artificial Intelligence and Cybersecurity Research", 2023.06.07
https://www.enisa.europa.eu/publications/artificial-intelligence-and-cybersecurity-research
13) OWASP, "OWASP Top 10 for Large Language Model Applications", 2023.10.16.
https://owasp.org/www-project-top-10-for-large-language-model-applications/

같이 정의하였다. 데이터 중독(Data poisoning)은 학습 데이터가 변경되면 모델의 동작이 조작될 수 있음을 의미한다. 이러한 변화를 통해 인공지능(AI) 시스템을 방해하거나 공격자가 원하는 결정을 내리게 할 수 있다. 입력 조작(Input manipulation)은 오해의 소지가 있는 입력 데이터로 모델을 조작하려고 하는 공격이다. 신속한 주입(Prompt injection)은 입력 조작 유형의 공격의 전형적인 예시이다. 회원 추론(Membership inference)은 데이터 기록과 모델에 대한 블랙박스 액세스를 통해 문서가 학습 데이터 세트에 있는지 여부를 확인할 수 있다. 이는 적대 행위자가 특정 질병을 앓고 있는지, 정당에 속해 있는지, 특정 조직에 가입했는지를 알 수 있음을 의미한다. 모델 반전 또는 데이터 재구성(Model inversion or data reconstruction)은 모델과 상호작용함으로써 학습 데이터가 추정되는 문제이다. 데이터가 민감한 경우 위험 수준은 더욱 커진다. 가장 대표적인 예시는 개인정보이다. 모델 도용(Model theft)은 모델과 상호 작용하면 모델의 동작을 파악하고 이를 복사하여 다른 모델을 교육하는 공격을 의미한다. 이는 지적 재산의 도용 문제를 야기한다. 모델 공급망 공격(Model supply chain attack)은 예를 들어 공개된 기본 모델을 오염시키고 전이 학습을 사용하여 해당 모델을 개선하는 딥러닝 모델을 손상시키는 등 모델의 수명 주기를 조작할 수 있는 공격이다.

IV. 인공지능(AI)과 사이버안보 차원에서 우리는 무엇을 준비해야 하는가?

인공지능(AI)는 다른 기술과 마찬가지로 좋은 목적과 악의적인 목적으로 사용된다. 사이버 보안 위험에 인공지능(AI) 기술이 적용되면서 국가에서 숙련되지 않은 개인까지 다양한 위협 행위자가 등장하고, 그에 따른 사이버 범죄 및 위험의 종류 또한 확장되었다. 인공지능(AI) 기술을 통해 등장한 새로운 위험으로는 인공지능(AI)로 인해 최적화된 사이버 공격이나 자동화된 멀웨어, 자율주행자동차·무인항공기와 같은 소프트웨어를 중심으로 작동하는 물리 장비에 대한 위협, 활용되는 인공지능(AI) 모델에 대한 도난, 인공지능(AI)를 통한 사칭 및 사기·피싱 등의 정교화 등이 있다.[14)]

새로운 사이버 보안 위험의 등장은 새로운 사이버안보 위험으로 이어지므로 이에 대한 추가적인 대응체계가 마련되어야 한다. 전술한 사이버 보안 위험으로 발생할 수 있는 사이버안보 위험은 국가 위험 행위자의 공격 기술의 정교화 및 자동화, 위협 행위자의 신원에 대한 사칭으로 정확한 대적 식별의 어려움, 기반시설에 대한 물리적 피해, 국가 시스템에 적용된 인공지능(AI) 모델의 탈취 등이다.

따라서 마찬가지로 사이버안보 위험의 대응을 위해 인공지능(AI)가 적용되어야 한다. 인공지능(AI)로 인해 확장된 위험은 데이터 처리 능력이나 지속성, 정확성 등의 영역에서 인

14) Malwarebytes, "AI in Cyber Security: Risks of AI", https://www.malwarebytes.com/cybersecurity/basics/risks-of-ai-in-cyber-security

간이 수행할 수 있는 범주를 넘어서기 때문에, 인공지능(AI) 기술의 활용으로 대응해야 한다.

그리고 국가안보에서의 사이버 피해는 기존의 사이버 보안 위험 수준과 비교할 수 없이 파급력이 크기 때문에 신속하고 정확한 대응과 함께 적극적이고 선제적인 대응이 필요하다. 기존 국가 사이버안보 차원에서 적극적 사이버 방어(ACD)가 함께 추진되었듯이 인공지능(AI)를 활용한 적극 대응이 필요하다.

그리고 인공지능(AI) 시스템이 실현하는 비즈니스 어플리케이션은 다양하며, 모두에서 완전한 대책을 목표로 하는 것은 비현실적이다. 인공지능(AI) 시스템의 위험 분석을 수행함으로써 대책과 비용의 균형을 객관적으로 검증해야 한다.

인공지능(AI) 시스템이 침해될 확률과 그 때 발생하는 피해의 크기를 곱하여 위험을 평가할 수 있다. 인공지능(AI) 시스템은 이용 환경이나 액세스 경로에 의해 외부로부터의 위협에 노출되는 정도가 다르다. 또, 인공지능(AI) 모델이 담당하는 태스크나 취급하는 데이터에 따라서 피해의 종류・성질에 차이가 있어, 이것을 고려한 영향평가를 실시하는 것이 포인트가 된다. 평가 축의 설정에 있어서는 인공지능(AI) 원칙이나 가이드라인 등이 참고가 된다. 예를 들어, 자율주행차는 실제 환경에서 오픈 액세스가 가능한 다양한 위협에 노출되는 인공지능(AI) 시스템이지만, 적대적 인공지능(AI) 공격에 의해 인체나 생명에 위기가 있을 가능성이 있기 때문에, 철저한 안보 대책이 요구되고 있다. 한편, 사내 이용의 챗봇 등의 폐쇄된(closed) 환경에서 한정적 태스크를 실행하는 인공지능(AI) 시스템은 공격자의 동기가 부족하고 공격을 실행하기도 어렵기 때문에 최소한의 대책으로 끝낼 수 있다는 판단이 들기 마련이다.

그리고 보다 구체적인 안보 조치를 구현하려면 인공지능(AI) 시스템의 특성을 이해하고 설계해야 한다. 지금까지 보았듯이, 인공지능(AI) 모델은 교육과 프로덕션 운영 중에 각각의 위협이 존재했으며, 인공지능(AI) 모델에 고유한 대책과 기존 IT 시스템을 위한 대책을 결합하는 것이 중요하다.

인공지능(AI) 시스템에 대한 정찰, 공격 수단의 준비, 모델 정보에 대한 무단 액세스(물리·디지털), 공격(중독·회피)의 실행 등 스테이지별 수법과 사례가 정리되어 있다. 공격 시나리오를 읽고 대책을 설계할 때는 공격자 시선의 사고방식과 기술을 이해할 수 있는 사이버안보 전문가가 필요하다. 인공지능(AI) 안보 대책의 구현은, 인공지능(AI)의 전문가 혹은 사이버안보의 전문가가 어느 쪽인가 단독으로 실시하는 것은 불가능하다. 팀으로서 인공지능(AI) 안보 능력을 가져야 한다.

마지막으로 인간과 마찬가지로 인공지능(AI) 시스템도 훈련받은 데이터의 편견에 영향을 받을 수 있다. 훈련 데이터가 편향된 경우 인공지능(AI) 시스템은 차별적인 결과를 생성하여 사이버안보 의사 결정에 영향을 미칠 수 있다. 선도적인 인공지능(AI) 플랫폼은 시스템의 편견을 최소화하고, 보다 공정한 결과를 보장하기 위해 지속적이고 사려 깊은 머신러닝 교육에 투자해야 한다.

<참고문헌>

CloudSEK, "Threat Actors Abuse AI-Generated Youtube Videos to Spread Stealer Malware", 2023.03.13. https://www.cloudsek.com/blog/threat-actors-abuse-ai-generated-youtube-videos-to-spread-stealer-malware

ENISA, "Artificial Intelligence and Cybersecurity Research", 2023.06.07. https://www.enisa.europa.eu/publications/artificial-intelligence-and-cybersecurity-research

ENISA, "Cybersecurity of AI and Standardisation", 2023.03.14. https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardisation

NCSC, "Introducing the guidelines for secure AI", 2023.11.27. https://www.ncsc.gov.uk/blog-post/introducing-guidelines-secure-ai-system-development

NCSC, "The near-term impact of AI on the cyber threat", 2024.01.24. https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat

NIST AI 100-2e2023 ipd. 2024.01.

OWASP, "OWASP Top 10 for Large Language Model Applications", 2023.10.16. https://owasp.org/www-project-top-10-for-large-language-model-applications/

Malwarebytes, "AI in Cyber Security: Risks of AI", https://www.malwarebytes.com/cybersecurity/basics/risks-of-ai-in-cyber-security

Tarlogic, "What are the AI security risks?", 2023.09.07. https://www.tarlogic.com/blog/ai-security-risks/

Tech Plus, "AIが生成したYouTube動画でマルウェアに感染させるサイバー攻撃が増加", 2023.03.14. https://news.mynavi.jp/techplus/article/20230314-2624772/

The Conversation, "AI scam calls imitating familiar voices are a growing problem – here's how they work, 2023.07.12. https://theconversation.com/ai-scam-calls-imitating-familiar-voices-are-a-growing-problem-heres-how-they-work-208221

White House, "FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence", 223.10.30. https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

ZDnet, "Scammers are using AI to impersonate your loved ones. Here's what to watch

out for", 2023.03.09.

※ 이슈브리핑에 수록된 내용은 필자 개인의 의견이며, 한국사이버안보학회의 공식적인 입장이 아님을 알려드립니다.