AI와 네트워크 시대, 보안의 조건

서동일 정보보호기술위원회(TC5) 부의장, 한국전자통신연구원 책임연구원

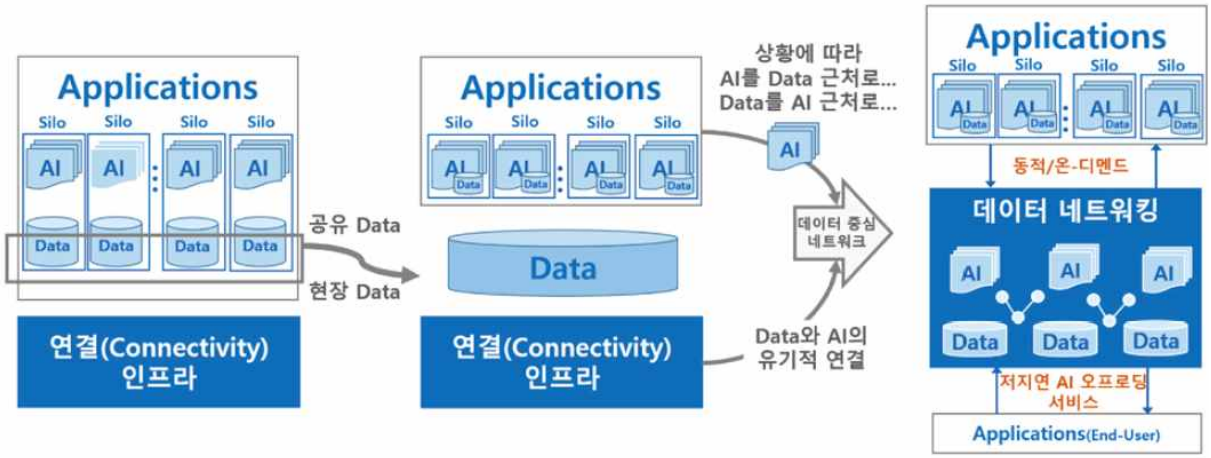
1. 머리말

최근 ChatGPT라고 하는 인공지능 채팅봇이 세계적으로 폭발적 반응을 얻고 있다. 인간이 채팅봇에게 질문하면 이에 적절한 답변을 제공하는, 어떻게 생각하면 매우 단순한 구조인데도 불구하고 세계적인 인기를 얻는 현상은 그동안의 인공지능 채팅봇 사례에 비추어 봤을 때 상당히 이례적이다. 이러한 인기는 기존 채팅봇에서 얻을 수 없던 매우 중요한 한가지 차이 때문이다. ChatGPT의 입력 및 답변의 특징은 실용적으로 사용 가능한 수준이라는 것이다. 기존 채팅봇은 미리 설정된 특정 입력 방식과 답변 혹은 매우 제한된 응용만 가능했다면, ChatGPT는 사용자 입력도 고도화된 자연어 처리 기술을 활용하여 처리하며 매우 다양한 분야에서 실제로 사용 가능한 답변을 제공한다. 이는 인공지능(AI) 중심 사회가 멀지 않다는 하나의 방증일 것이다.
가장 기본적인 정보검색을 돕는 것뿐만 아니라 창의성이 필요한 분야에도 활용할 수 있다. 주어진 주제에 맞게 새로운 글을 작성하는 등 업무 보조도구로 쓸 수 있고, 개인 자산의 관리에 대한 조언, 개인 맞춤형 요리방법이나 운동 방법 제시, 외국어 공부 도우미, 질문자가 원하는 다양한 여행에 대한 정보나 추천 제공 등등 그 활용법이 매우 다양하다. 심지어 그동안 인류에게만 허락된 영역으로 보였던 창의적 영감이 필요한 예술영역에서도 활용될 수 있다.

2. AI, 네트워크 중심 사회의 도래

최근에는 인공지능으로 정보 보안을 관리하는 도구까지 출시되고 있다[2]. 기존에는 사이버 공격으로 의심되는 동작을 빠르게 탐지하기 위해 인공지능을 이용했다면, 요즘에는 사용자가 시스템 및 네트워크 환경 등을 제시하면서 특정 사이버 보안 관련 요구사항을 제시하면 AI 도구가 이를 위한 효율적 해결법을 제시할 뿐만 아니라 추후 관련 보고서도 작성해 주는 등 활용 영역이 확대되고 있다.
이러한 AI 기술에는 먼저 충분한 학습을 위한 데이터를 필요로 하며, 사용자와의 초고속 연결 상태를 상시적으로 제공해 줄 수 있는 네트워크의 발전 또한 필요하다. AI 기술을 활용하기 위해서는 AI와 사용자를 네트워크로 연결하여야 하며, 이러한 기반을 바탕으로 사회의 빅데이터를 수집 및 활용하고 사용자가 원하는 적절하고 실용적인 해답을 제공할 수 있기 때문이다.
최근 네트워크 및 AI 관련 주목받는 주요 기술로는 머신러닝 기반의 예방적 네트워크 보안기술, 네트워크 운영관리의 자동화 및 자율화 기술, 선제적 장애 대응을 통한 고신뢰 네트워크

기술, 클라우드 중심의 네트워크 운영기술, 네트워크의 가상화 및 자원 최적 배분기술, 네트워크 슬라이싱 자동화 기술, 네트워크 서비스 품질 최적화 기술 등을 꼽을 수 있으며[4, 9, 10, 11], 이러한 기술들에 AI 기술을 융합하려는 시도들이 나타나고 있다. 분산된 AI의 네트워크 연결 기술 등이 여기에 포함된다.



[그림 1] 분산 AI 네트워크 연결 기술 개념도. 공유되는 자원과 네트워크가 커질수록 보안은 어려워진다.[4]

향후 미래사회는 AI가 네트워크를 통해 사용자 환경과 밀접하게 상시적으로 연결된 초연결, 초지능화 사회로 발전할 것으로 예측되며 이러한 환경 속에서 사이버 보안의 중요성은 더욱 커질 전망이다.

3. 새로운 생태계 환경에서의 보안위협, 개인정보보호

현재의 네트워크 및 AI 생태계 환경에서는 일부 창의적이고 예술적인 분야까지 AI를 통해 제공하고자 시도 중인 것으로 보이며, 향후 미래 사회에서는 인간과 유사한 사고력과 행동력을 보여줄 수 있는 AI의 등장까지 고려하여야 할 것이다. 이러한 생태계 환경에서 나타날 수 있는 정보보안 위협은 어떠한 것들이 있을까? AI 자체에 대한 보안위협, 네트워크 및 AI 발전으로 인해 더욱 손쉽게 나타날 개인 프라이버시 노출 문제 등이 나타날 것이며, 기존 네트워크에서 계승된 전통적 사이버보안 위협도 여전할 것이다.

첫번째 고려할 보안위협으로는 AI 자체에 대한 위협을 꼽을 수 있다[3, 5, 6, 7, 12, 13, 14, 15, 16]. 예를 들어, AI 학습데이터에 잘못된 데이터를 주입하여 AI 모델의 정확도를 떨어뜨리거나, 이미 학습된 AI 모델을 활용하는 단계에서 입력되는 데이터를 위변조하여 AI가 잘못된 답변을 하도록 유도하는 등의 보안위협이다. 대표적 보안위협으로는 적대적 학습 공격(adversarial training attack)이 있다[12, 16]. 이는 주로 데이터에 대한 보안위협으로, 미래 사회에서는 데이터의 중요도가 현재와 비교하여 더욱 높아질 것이다.

AI의 개발과 활용을 위해 수집되는 데이터는 매우 방대하고 광범위할 것으로 보이며, 이로 인해 이들 데이터를 적절히 결합할 경우 특정 개인 이용자를 식별하거나 추적, 모니터링 할 수

있는 위험성이 현재보다 더욱 높아질 것으로 판단된다. 수집되는 데이터 역시 정형, 비정형을 구분하지 않을 것이며, 음성이나 영상 같은 멀티미디어 데이터 또한 기술적으로 더욱 쉽게 얻을 수 있을 것으로 보인다. 따라서, 개인 프라이버시를 위협하는 개인정보보호 문제는 새로운 AI 및 네트워크 중심사회에서 매우 관심이 큰 사항이 될 것이다.
또한 새로운 생태계 환경에서도 기존 네트워크에서 계승된 사이버보안 문제가 여전한 위협으로서 존재할 전망이다. DoS/DDoS 공격, 시스템 및 인프라 취약점 공격, 데이터에 대한 위변조 및 도감청 공격 등등이 새로운 생태계에서도 여전히 문제를 일으킬 것으로 보인다.

4. 새로운 생태계 환경을 위해 준비해야 할 것들

세계 각국은 AI과 같은 새로운 생태계 환경으로 진입하면서 사이버 보안 위협에 효율적으로 대응하기 위해 전담기구 설치나 예산 배정과 같은 국가 차원의 사이버 안보 전략을 수립하여 추진하고 있다. 우리나라도 국가 사이버안보 기본계획을 수립하여 자유롭고 안전한 사이버 공간을 구축하기 위해 노력하고 있다[17]. 그러나 인공지능과 네트워크, 데이터 융합 사회로 생태계 환경이 진화하면서 사이버 보안의 국가적 전략 또한 한단계 진일보하여야 할 상황이다.
인공지능과 네트워크, 데이터 기술은 각각의 분야에서 독자적 발전을 지속하게 될 것이나, 새로운 생태계 환경으로 진입하면서 서비스의 경계는 모호해질 것이다. 네트워크 분야에서는 인공지능을 도입할 수밖에 없는 네트워크 복잡성이 매우 높아질 것이고, 사용자에게 혹은 사회적으로 큰 영향을 미칠 수 있는 정보 보안 문제가 새로운 네트워크 기술 분야에서도 더욱 빠르게 확산될 것이다[18].
사이버 보안에 있어서 현재 AI 기술은 사이버공격을 보다 정확하고 빠르게 탐지하여 대응하는데 중점을 두고 있다. 그러나 인공지능 기술과 네트워크 기술이 발전하면서 이를 더욱 광범위하고 효율적으로 사이버 보안에 적용할 수 있는 융합 기술이 필요해질 것이다. 현재 생태계 환경에서 사이버 공간의 특성을 활용한 피싱 같은 사기행각이 다양하게 이루어지고 있는데, 이에 맞서 머신 러닝과 같은 인공지능 기술을 활용하여 사기 위험을 사전에 탐지하여 대응할 수 있는 기술 등을 더욱 발전시켜야 할 것이다. 더불어, 인공지능 기술을 활용한 사이버 공격자들의 공격 기법 또한 발전할 것이기 때문에, 이에 대한 정보보안 대책 또한 준비할 필요가 있다.

5. 맺음말

인공지능 기술의 발전은 사이버 보안에 있어서도 필연적으로 패러다임 변화를 야기할 것이다. 인공지능 기술이 사이버 공격에 대한 방어 수단에 그치는 것이 아니라, 사이버 공격 기술을 고도화하는 도구로도 적극 사용될 것이기 때문이다. 또한, ChatGPT와 같은 인공지능 챗봇은 지속적으로 발전하고 사용인구도 계속 늘면서 새로운 AI 중심 생태계 진입을 촉진할 뿐만 아니라, 우리 삶에 매우 커다란 영향을 미칠 것이다[1].
그러나 새로운 생태계 환경에서 정보보호 업무가 AI에 의해 대체 가능할 것인지는 이견이 있다. 인공지능 기술이 진화 발전함에 따라 정보보호분야에서 지금보다 훨씬 더 그 활용이 촉진될 것이나, 관련 업무가 AI에 의해 대체되지는 않을 것으로 생각된다. 이는 AI 기술은 공격자

들에게도 동일하게 영향을 미칠 것이며, 정보보호 업무는 이러한 변화까지 고려하여야 하기 때문이다.

정보보호 기술 연구자는 AI로 증강된 사이버 보안 기술을 확보하기 위해 인공지능과 네트워크, 데이터 중심의 새로운 생태계 환경에 적합한 보다 광범위하고 다양한 기술적 지식이 필요할 것이며, 이들 기술들을 효율적으로 융합할 수 있는 기술과 능력을 개발하는 노력을 계속해야 할 것이다.

[참고문헌]

[1] 뉴시스, https://v.daum.net/v/20230412111337966, 2023년 4월

[2] 마이크로소프트/Zdnet Korea, https://zdnet.co.kr/view/?no=20230329074125, 2023년 3월

[3] 이태진, "인공지능(AI) 기반의 정보보호 기술 동향," KISA Report, 2020 Vol.12

[4] 김태연, 고남석, 양선희, 김선미, "네트워크와 AI 기술 동향," 전자통신동향분석 제35권 제5호, 2020년 10월

[5] 김호원, "컴퓨터 비전 분야에서 AI 보안에 대한 연구 동향," 2021 KISA Report, Volume 07

[6] 유진호, 민경식, 박진상, 김관영, "AI 중심사회의 도래와 보안 이슈 분석," KISA Insight, 2022 Vol.3

[7] 김민진, "인공지능: 사이버보안 패러다임의 전환," 정보통신정책연구원, AI Trend Watch, 2021-3호

[8] 민경식, 김관영, 장한나, "2030 미래사회 변화 및 ICT 8대 유망기술의 사이버 위협 전망," KISA Insight, 2022 Vol.1

[9] Ali Hussein et al., "Machine Learning for Network Resilience: The Start of a Journey," 2018 Fifth International Conference on Software Defined Systems(SDS).

[10] Cisco, "Machine Learning And AI for Networking," https://www.ciscolive.com/c/dam/r/ciscolive/us/docs/2019/pdf/BRKSDN-2260.pdf, June 2019

[11] Guosheng Zhu et al., "A Supervised Learning Based QoS Assurance Architecture for 5G Networks," IEEE Access, Vol 7, 2019

[12] I. Goodfellow et al., "Explaining and Harnessing Adversarial Examples," International Conference on Learning Representations, 2015.

[13] N. Papernot et al., "Practical black-box attacks against machine learning," In Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, ASIACCS'17, pp. 506–519, 2017.

[14] P.Y. Chen et al., "ZOO: Zeroth Order Optimization Based Black-box Attacks to Deep Neural Networks without Training Substitute Models," In Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security (AISec), 2017.

[15] M. Alzantot et al., "Genattack: Practical black-box attacks with gradient-free optimization," In Proceedings of the Genetic and Evolutionary Computation Conference, pages 1111–1119, 2019.

[16] A. Raghunathan et al., "Certified defenses against adversarial examples," International Conference on Learning Representations, 2018.
[17] 관계부처 합동(청와대 국가안보실), "국가 사이버안보 전략,"
https://ccdcoe.org/uploads/2018/10/South-Korea_%EA%B5%AD%EA%B0%80%EC%82%AC%EC%9D%B4%EB%B2%84%EC%95%88%EB%B3%B4%EC%A0%84%EB%9E%B5%EA%B5%AD%EB%AC%B8National-Cybersecurity-Strategy_2019_original-2.pdf, 2019년 9월
[18] CISCO, "2020 글로벌 네트워킹 트렌드 보고서,"
https://www.cisco.com/c/dam/m/ko_kr/solutions/enterprise-networks/networking-report/files/GLBL-KO_NB-06_0_NA_RPT_PDF_MOFU-no-NetworkingTrendsReport-NB_rpten018612.pdf, 2019
[19] Marcus Comiter, "Attacking Artificial Intelligence,"
https://www.belfercenter.org/sites/default/files/2019-08/AttackingAI/AttackingAI.pdf, August 2019
[20] Alexander Chistryakov, Alexey Andreev, "AI under Attack,"
https://content.kaspersky-labs.com/se/media/en/businesssecurity/enterprise/machine-learning-cybersecurity-whitepaper.pdf, 2019
[21] Blessing Guembe, Ambrose Azeta, Sanjay Misra, Victor Chukwudi Osamor, Luis Fernandez-Sanz & Vera Pospelova, "The Emerging Threat of Ai-driven Cyber Attacks: A Review," Applied Artificial Intelligence, Mar. 2022
[22] Matti Aksela, Samuel Marchal, Andrew Patel, Lina Rosenstedt, WithSecure, "The security threat of AI-enabled cyberattacks,"
https://www.traficom.fi/sites/default/files/media/publication/TRAFICOM_The_security_threat_of_AIenabled_cyberattacks%202022-12-12_en_web.pdf, Dec. 2022
[23] Ning Yu, Zachary Tuttle, Carl Jake Thurnau, Emmanuel Mireku, "AI-Powered GUI Attack and Its Defensive Methods," https://arxiv.org/ftp/arxiv/papers/2001/2001.09388.pdf, 2019

※ 출처: TTA 저널 제206호