Center for Future Warfare Studies,
Institute of International Studies at Seoul National University |
국제문제연구소 미래전연구센터 연구위원 워킹페이퍼 No.18.(발간일: 2025.2.7.)

데이터 안보와 인공지능 안보

송태은 국립외교원 국제안보통일연구부 조교수

I. 서론

오늘날 인공지능은 강대국 간 기술패권 경쟁의 승자를 결정짓는 가장 중요한 변수이자 국가의 경제 · 산업, 군사, 행정, 정보와 커뮤니케이션 등 국가 전 영역의 발전을 좌지우지할 핵심 기술이다. 이미 현대 전쟁의 무기체계에 본격적으로 적용되고 있는 인공지능 기술의 파괴력은 현재의 러시아–우크라이나 전쟁과 이스라엘–하마스 전쟁을 통해 증명되고 있다. 인공지능 기술이 탑재된 자율드론은 장거리에서도, 야간에도 공격대상을 정확하게 식별하고 정밀 타격을 수행하는 파괴력을 보여주어 이미 AI 드론전이 본격화되고 있다. 인공지능 기술의 정밀탐지 및 실시간 정보 수집과 분석 능력은 정보의 우위가 곧 전장의 우세이며 앞으로 미래전쟁의 승패는 화력(firepower)에 앞서 정보전(information warfare)과 사이버전(cyber warfare)이 좌우할 것임을 예고하고 있다. 즉 화력이 파괴력을 발휘하기 위한 조건이 인공지능의 데이터 역량과 이러한 데이터를 안전하게 관리할 수 있는 사이버 공간에 대한 접근성에 달려있기 때문이다.

이미 러시아–우크라이나 전쟁과 이스라엘–하마스 전쟁에서 인공지능은 적에 대해 실시간으로 수집한 정보를 즉각적인 공격으로 연결시키는 역할을 수행하고 있다. 감시정찰을 통한 정보의 우위뿐 아니라 정밀탐지와 인공지능의 자율적 판단을 통해 적에 대한 타격의 파괴력을 강화시켜주고 있는 것이다. 우크라이나는 러시아 군을 식별하기 위한 알고리즘 프로그램인 '클리어뷰(Clearview)'를, 이스라엘은 살상 대상 식별 알고리즘인 '라벤더(Lavender)'을 가자 지역에서 사용하고 있다. 이와 같이 인공지능에 의해 증대한 경쟁하거나 싸우는 주체의 정보분별 능력 및 신속한 자율적 의사결정 능력을 무력화시키기 위해 대단히 원초적인 기만전술이 동원되기도 한다. 우크라이나는 AI 기술로 증강된 러시아의 원거리 탐

지를 무력화시키고 자국의 실전 배치 무기를 보호하는 동시에 러시아 전력의 신속한 소실을 위해 철강회사 멧인베스트(Metinvest)를 통해 가짜무기(decoy)를 제조하고 실제로 배치, 은폐하는 전술을 취하고 있다. 러시아도 항구에 배치한 무기를 우크라이나의 자율드론과 크루즈 미사일이 파괴하자 가짜 잠수함과 전투기 배치로 응수하고 있다.

이렇게 이미 전장에 깊숙이 들어와 있는 인공지능 기술은 국가 안보와 직결되고 있고, 이러한 인공지능 기술의 기계학습 대상인 '데이터'는 궁극적으로 국가안보의 핵심적인 변수가 된다. 즉 인공지능의 성능을 보장하는 것은 데이터의 '규모'와 '질(quality)'이 되는 것이다. 따라서 국가는 ①인공지능 기술의 발전을 결정짓는 학습데이터를 확보하고 이 데이터가 오염되지 않도록 보호하고, ②수집된 데이터를 사용하는 과정에서 민감한 정보나 개인정보의 보호 등 데이터를 안전하게 사용해야 하는 문제, 그리고 ③인공지능 기술로 생성한 데이터와 데이터 분석 결과 등을 안전하게 관리하고 생성 목적에 부합하게 사용하고 활용해야 한다. 다시 말해, 국가는 인공지능 기술에 직접적인 영향을 끼치는 데이터를 보호하고, 동시에 그러한 기술에 의해 생성된 데이터를 관리하고 보호해야 하는 활동을 수행하게 된다.

더군다나 인공지능 기계학습에 사용되는 데이터는 다양한 개인정보가 차지하는 비중이 높기 때문에 유출될 경우 개인이 직접적인 피해를 받기 때문에 개인정보 보호의 문제가 끊임없이 발생한다. 개인의 프라이버시 보호 문제에 더하여 한 국가의 대규모 개인정보가 타국의 인공지능 기계학습에 사용될 경우 이는 국부의 유출과 다름없고 적대국의 경우 자국의 안보를 위협하는 데에 악용될 수 있으므로 국가는 자국 시민의 정보가 국경을 넘어 이동하는 것을 '주권'의 문제로 인식하기도 한다. 한편 개인정보가 담긴 대규모 데이터는 인터넷과 연결된 기기를 개인이 자율적으로 사용하면서 저절로 생산되기도 하고, 챗봇(chatbots)과 같은 인공지능 프로그램을 사용하는 개인 스스로가 자발적으로 자신과 타인의 개인정보를 제공하는 등 사용자 스스로의 생성형 AI 사용방식이 개인정보 보호 문제를 일으킬 수도 있다. 특히 생성형 인공지능은 개인의 빈번한 사용에 의해 더욱 발전하므로 생성형 인공지능 프로그램을 제공하는 업체와 사용자 간에도 정보의 보호와 활용과 관련된 복잡한 문제가 끊임없이 발생하게 되므로 국가는 이러한 문제를 해결해야 하는 위치에 놓인다.

이와 같이 국가의 데이터 보호는 곧 자국 시민 보호 이슈가 되기 때문에 국가는 데이터 안보를 위해 '디지털 주권(digital sovereignty)'과 같은 개념을 강조하면서 국가의 법·제도를 포함한 정책에 그러한 개념을 반영하기도 한다. 하지만 국가는 대규모 개인정보를 적극적으로 수집하고 활용하는 주체이고 그러한 정보 수집을 위해 인공지능 기술이 적용된 감시체제를 광범위하게 운용한다. 다시 말해 국가는 자국 개인의 정보를 보호해야 하는 동시에 자국 개인들의 정보를 수집하여 자국의 인공지능 기술 발전을 도모해야 하는 이중적인

책임을 수행하는 위치에 놓여 있다.

이와 같은 이슈 외에도 국가는 알고리즘 오작동 유발을 위해 악의적으로 변조된 데이터를 학습데이터에 삽입시키는 외부로부터의 적대적 공격(adversarial attacks)에 대해서 자국의 인공지능 시스템을 보호하는 활동을 수행한다. 국가와 기업은 학습데이터와 인공지능에 의해 생산된 산출 데이터에 대한 외부로부의 공격을 차단해야 하므로 인공지능 데이터 보호의 문제는 늘 사이버 안보(cyber security) 문제와 밀접하게 연결된다. 결과적으로, 국가뿐 아니라 인공지능 프로그램을 개발하는 당사자인 기업도 국가와 기업이 ▲수집하고 처리하는 학습데이터를 보호하는 문제와 ▲학습데이터의 주요 소스가 되는 개인정보의 보호 문제, ▲알고리즘 오작동 유발을 위해 악의적으로 변조된 데이터를 학습데이터에 삽입시키는 공격에 대한 차단 등 데이터 안보와 관련된 다양한 문제를 인공지능 안보와 국가 안보의 관점에서 관리해야 한다.

이렇게 볼 때 '인공지능 기술을 보호하는 인공지능 안보'와 '데이터를 보호하는 데이터 안보'는 사실상 동일한 내용을 담는다고 볼 수 있다. 또한 인공지능 안보를 논함에 있어서 개인정보에 대한 보호 등 인권과 관련된 문제와 외부로부터의 공격을 막아내는 사이버 안보의 문제 등 다양한 성격의 이슈들이 개입되고 있음을 알 수 있다. 이러한 맥락에서 이 연구는 인공지능 기술과 데이터 자체가 국가 안보와 어떤 관계가 있는지 살펴본다. 먼저 ① 인공지능의 기계학습에 투입되는, 인공지능이 수집하기도 하는 데이터와 학습데이터와 관련하여 국가 안보가 어떻게 연결되는지 살펴본다. 또한 ② 인공지능이 생산하는 데이터 즉 '인공지능 산출 데이터'를 관리하고 보호하는 문제를 검토하고, ③이러한 인공지능 데이터에 대한 외부로부터의 위협이 어떤 것인지 살펴봄으로써 인공지능 안보와 데이터 안보가 서로 국가 안보 차원에서 어떤 관계를 갖는지 설명한다. 마지막으로 결론은 현재 그리고 앞으로 인공지능 안보와 데이터 안보의 문제에 대해 각국과 국제사회가 어떻게 대응하고 대처할 것인지 간략하게 짚어보는 것으로 이 글을 마무리한다.

Ⅱ. 인공지능 수집데이터 및 학습데이터와 국가 안보

논리(logic), 추론(reasoning), 의사결정 능력을 발휘할 수 있는 인공지능의 인지적 컴퓨팅(cognitive computing) 능력은 빅데이터의 3가지 'v' 즉 데이터의 규모(volume), 데이터가 수집되는 속도(velocity), 데이터의 다양성(variety)에 달려있다.[1] 이러한 데이터는 인공

[1] Admond Lee, "Volume, velocity, and variety: Understanding the three V's of big data" Datasource.AI(April7, 2020). https://www.datasource.ai/en/data-science-articles/volume-

지능 감시기술, 사물인터넷(Internet of Things, IoT), 센서(sensors), 인공지능은 인터넷에 게시되어 있는 저서, 기사, 보고서, 인터넷 상의 게시물 스크래핑(Scrapping)을 비롯하여 소셜미디어 콘텐츠, E-commerce 플랫폼, 영상 스트리밍(streaming), 인공위성 데이터, 기계생성 데이터(machine-generated data), 과학연구 결과 등을 통해 얻어진다. 그러면 인공지능 기술이 수집하는 정보는 국가 안보와 어떤 관련이 있나?

인공지능은 스스로의 학습재료가 되는 빅데이터를 직접 수집하는 데에도 광범위하게 사용되고 있다. 먼저, 급속도로 발전하고 있는 인공지능 기술은 국가의 사회에 대한 감시와 통제체제에 유용한 데이터를 제공한다. 국토 보호와 사회질서 유지를 위해 사용되는 인공지능 기술이 적용된 첨단 감시기술과 모든 곳에 편재하면서 인터넷 연결을 무한하게 확장시키는 지능형 사물인터넷(AIoT), 그리고 개인의 디지털 기기 사용에 의해 자동적으로 실시간 생성되는 대규모 개인정보는 국가가 개인과 사회에 대한 전방위적 감시체제를 구축할 수 있는 중요한 기반이 되고 있다. 이러한 기술 및 정보환경은 전방위적인 국가 감시체계를 형성시키는 결과를 가져오고 국가는 국내 정치에 영향을 끼칠 수 있는 더 광범위한 통치기술과 정교한 디지털 자원을 구비할 수 있다.[2)]

한편 사물과 공간에 대한 데이터의 국가안보와의 밀접한 관계는 최근 인공위성과 같은 우주기술이 인공지능 기술과 결합되면서 더욱 강화되고 있다. 전통적으로 항공수단을 이용해왔던 군의 감시정찰 활동이 최근 우주영역의 임무로 전환되고 있고 인공지능 기술과 우주기술이 결합하면서 국가의 감시정찰 능력은 더욱 강력해지고 있다. 예컨대 지구 환경에 대한 관측 데이터는 인공지능 기술을 탑재한 다양한 사물인터넷, 드론이나 인공위성에 의해 광범위하게 얻어지고 있고, 치안, 재난재해 대응이나 도시연구 및 환경보호 등에도 활용된다. 최근 드론에 데이터, 5G 통신 네트워크와 인공지능 기술이 접목되면서 드론이 확보한 데이터를 인공지능이 실시간으로 분석하고 전송할 수 있게 되었다.[3)]

인공지능 기술과 빠르게 융합되고 있는 위성정보는 산불, 삼림 벌채, 사막화, 수자원의 변화, 탄소(CO2) 농도, 대기질, 극지방의 빙하 및 해수면 상승도, 생물 다양성 등 환경에서 발생하는 다양한 변화를 모니터링하거나 관리하는 데에도 유용하다. 위성정보는 화산분출, 홍수와 지진과 같은 재난과 재해에 대한 조기경보, 매핑에 더하여 피해의 평가와 대피계획과 비상 자원 활동 등 복구와 관련된 활동에도 유용한 정보를 제공할 수 있다. 더불어 위성정보는 도시의 토지 이용 실태, 건물 밀도 및 이용 패턴 및 도시 인구 규모 등을 파악할

---

velocity-and-variety-understanding-the-three-v-s-of-big-data

2) 송태은, "인공지능 기술을 이용한 국가의 사회감시 체계 현황과 주요 쟁점" 『정책연구시리즈』2020-12, 국립외교원 외교안보연구소(2021).

3) 길애경, "데이터 확보 · 전송 · AI분석, 실시간 드론 통신 · · · "산업 활성화 마중물"" (2023.8.8.) https://www.hellodd.com/news/articleView.html?idxno=101371

수 있게 하여 복원력 있는 도시계획에도 활용된다. 이러한 지구관측 데이터들은 무료로 오픈소스(open source)로서 제공되기도 하여 관련 전문가들의 연구와 알고리즘 개발에 기여하기도 한다.

인공위성이 실시간으로 수집한 대규모 데이터를 인공위성에 탑재된 인공지능 기술을 통해 실시간으로 분석하는 것이 가능해졌기 때문에 국가의 정보우위는 인공지능 기술이 좌지우지한다고 볼 수 있다. 앞으로 국가의 안보는 국토에 대한 정밀하고 방대한 정보를 수집, 분석하고 다양한 군사작전을 가능하게 하는 인공지능 기술이 적용된 우주자산에 달려있다고 해도 과언이 아니다.

인공지능 감시기술을 통해 수집되는 데이터는 '핵 억지(nuclear deterrence)'와 같이 국제사회의 평화구축 노력에도 기여할 수 있다. 이미 군사분야에 광범위하게 사용되고 있는 인공지능 감시기술은 원거리 정밀탐지를 통해 핵물질이나 핵무기 등 핵 프로그램의 안전관리 및 검증 레짐(verification regime)의 효과를 증진시켜 국제적 핵비확산(non-proliferation)과 군축(arms control)에 기여할 수 있다. 내전이나 테러리즘 관련해서도 인공지능 기술은 국가나 비국가 행위자가 취할 수 있는 다양한 위협을 감시하고 평화유지활동(peace-keeping operation, PKO)을 지원할 수 있다.

기계학습에 필요한 사람과 관련된 데이터의 경우, 음성인식이나 안면인식 기술을 통해 수집되고 구축된 대규모 데이터베이스는 사용자의 다양한 금융활동 정보, 의료기록, 위치정보, 소셜미디어에 게시하는 글이나 이미지, 영상 등 다양한 정보와 결합될 경우 사실상 사용자에 대한 프로파일링(profiling)이 가능해진다. 예컨대 Open AI의 샘 올트먼(Sam Altman)은 홍채 인식 기반의 암호화폐 'Worldcoin'를 출시했는데 '오브(Orbs)'라는 홍채인식 기구를 이용하여 36개국에서 510만 여명의 홍채정보를 수집하고 있다. 홍채인식을 통해서 개인에게 'World ID'를 발급하고 있다.[4] 스페인과 포르투갈 및 영국 등에서는 월드코인의 생체정보 수집활동이 금지되고 있는데 유럽연합의 GDPR 규정을 위반하고 있는지 아닌지의 여부에 대해 논쟁이 지속되고 있기 때문에 앞으로 이 프로젝트가 유럽에서 지속될 수 있는지를 결정짓게 된다. 흥미로운 것은 'Tools for Humanity'로 불리는 이 프로젝트는 딥페이크 등을 통해 인공지능 알고리즘이 개인의 신원정보를 악용할 수 있는 가능성을 차단하기 위한 목적을 갖는다는 것이다.[5] 하지만 이 프로젝트가 내걸고 있는 그러한 주장에도 불구하고 개인정보 보호에 대한 논란은 지속되고 있다.

---

[4] Associated Press, "Spain puts temporary ban on Sam Altman's Worldcoin eyeball scans over privacy concerns" (March 8, 2024). https://www.euronews.com/next/2024/03/08/spain-puts-temporary-ban-on-sam-altmans-worldcoin-eyeball-scans-over-privacy-concerns

[5] Tom Matsuda and Zosia Wanat, "Will Europe's regulators stop Sam Altman's Worldcoin?" *Sifted* backed by *Financial Times*(August 27, 2024). https://sifted.eu/articles/worldcoin-regulation

중국의 경우 2016년부터 발전시킨 '뇌-컴퓨터 인터페이스(Brain-Computer Interface, BCI)' 연구를 군인의 전투능력 증진에 활용하는 등 뇌과학을 군사화시키는 한편 적의 인지능력에 대한 공격 즉 인지전(cognitive warfare) 수행을 위해서 미국과 같은 경쟁국의 영향력 인사들에 대한 대규모 개인 정보를 수집하고 있다. 특히 미국의 주요 정부 인사 2백만 명과 일반 시민 3천8백만 명, 그리고 10만 명이 넘는 미국 해군의 개인정보를 주요 호텔 기록 등을 통해 확보해놓은 상태이다. 중국 정부는 이러한 해외 주요 인물들의 개인정보를 중국 IT 기업에 제공하여 이들의 첩보 수집을 지원하고 있고, 이러한 첩보 활동은 특히 미국을 포함하여 대만과 홍콩에서 공격적으로 이루어지고 있다.[6] 이러한 타국 개인에 대한 개인정보와 이들에 대한 프로파일링도 인공지능 기술과 결합하여 중국의 인지전 연구에 사용된다.

Ⅲ. 인공지능 생성 데이터와 국가 안보

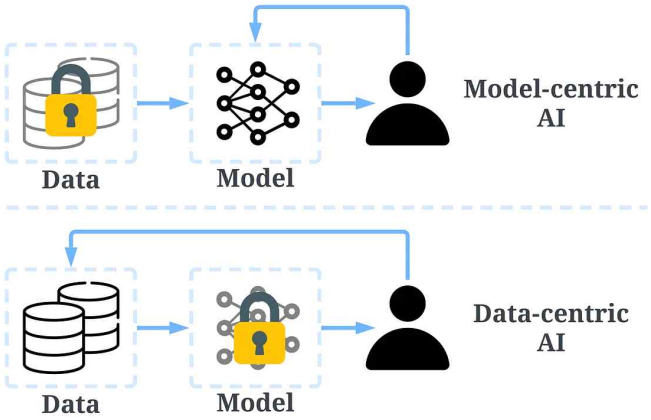
최근 인공지능 기술은 인공지능 시스템의 성능을 높이기 위해서 알고리즘 모델을 개선하거나 학습 모델을 개발하는 방식보다 양질의 데이터를 지속적으로 수집하는 데에 초점이 맞춰져 있다. 다시 말해, 인공지능 기술의 발전이 '모델'이나 '코드'를 중심으로 하기보다 '데이터 중심(data-centric)' 접근법이 강조되고 있는 것이다. 이러한 접근법이 강조되고 있는 것은 대규모의 데이터 학습 자체가 인공지능 모델 성능의 향상으로 직결되기 때문이다. 즉 데이터를 가공하는 데에 시간과 비용을 차후에 소모하기보다 처음부터 양질의 데이터를 확보하려는 방식이다.[7]

---

6) Koichiro Takagi, "New Tech, New Concepts: China's Plans for AI and Cognitive Warare" Texas National Security Review(April 13, 2022). https://warontherocks.com/2022/04/new-tech-new-concepts-chinas-plans-for-ai-and-cognitive-warfare.

7) 이용 · 최명석 · 김성찬 · 이건우 · 장래영 · 이승우 · 이상환, "인공지능 학습 데이터 공유 · 활용 현황과 서비스 구축 방향" KISTI Issue Brief 제51호(2022.12.26.) https://repository.kisti.re.kr/bitstream/10580/18082/1/KISTI%20%EC%9D%B4%EC%8A%88%EB%B8%8C%EB%A6%AC%ED%94%84%20%EC%A0%9C51%ED%98%B8.pdf

〈그림1〉 모델 중심 vs. 데이터 중심 인공지능 기계학습



① 생성형 AI의 데이터 생성

최근 등장한 대규모 언어모델(Large Language Model, LLM)과 같은 초거대 인공지능(super-giant AI, hyperscale AI)은 대규모의 데이터와 수천억 개 이상의 파라미터(parameter)로 학습한 인공지능 시스템으로 기존 인공지능보다 더 복잡하고 광범위한 업무를 수행할 수 있다.[8] 오픈AI가 출시한 초거대 AI 모델인 GPT-3는 사실상 초거대 AI를 본격적으로 등장시킨 일이었으며, 이러한 생성형 인공지능을 비롯한 초거대 인공지능은 방대한 데이터의 구축, 클라우드, 인공지능 반도체와 같은 컴퓨팅 능력의 발전 등 고도화된 알고리즘을 통해 등장했다. 클라우드 컴퓨팅(cloud computing)은 인터넷 클라우드를 통해 서버, 스토리지, 데이터베이스, 네트워크, 소프트웨어 등 컴퓨팅 서비스를 제공하는 것을 일컫는다.[9]

인공지능이 쉽게 생성시킬 수 있는 데이터는 Amazon Transcribe, Google Cloud Speech-to-Text, Microsoft Azure Speech Services, IBM Watson Speech to Text, Apple Siri, Dragon Professional Individual, Braina Pro과 같이 음성인식(voice recognition) 인공지능 알고리즘이 수집한 음성을 텍스트로 변환시켜주는 사례를 들 수 있다.[10] 최근 Google 딥마인드(Deepmind)는 단백질 구조를 예측하는 프로그램인 AlphaFold

8) 국회도서관, 『초거대 AI 한눈에 보기』 FACTBOOK 2023-5호 통권 제 105호(2023.11.13.)
9) "클라우드 컴퓨팅이란?" MS Azure. https://azure.microsoft.com/ko-kr/resources/cloud-computing-dictionary/what-is-cloud-computing.
10) "Top 7 voice recognition softwares of 2023"(August 16, 2023). https://botpenguin.com/blogs/top-7-voice-recognition-softwares-of-2023.

를 출시했다. 이 프로그램은 인체의 수십만 개 단백질의 3D 형태를 예측할 수 있을 뿐만 아니라 새로운 단백질을 디자인할 수 있는데 수 있다.[11] 또한 구글의 모기업인 알파벳은 제약회사들과 함께 약물연구를 위한 스타트업 기업을 설립하여 디지털 플랫폼 기업이 신약 개발 플랫폼을 만드는 사례가 나타나고 있다.[12]

② 다른 기술과의 융합에 의한 데이터 생성

인공지능은 융합되는 다양한 기술에 의해 새로운 데이터를 생성시킬 수 있다. 거짓말탐지 뇌파 판독기, 전파기술이 적용된 불법전파 탐지 기술 즉 전파교란 탐지기술, 스펙트럼 탐지, 불법 드론 탐지기술과 같은 레이더 탐지 등이 가능해지면서 새로운 데이터들이 생성되고 있다. 특히 인공지능 기반의 전파탐지 기술은 인공위성과 드론에도 적용되는 등 민간과 군사 양 분야에서 사용되고 있다.

③ 실제 세계를 모방하는 합성데이터

기계학습을 통해 생성된 실제 세계의 패턴을 흉내 내는 데이터를 '합성데이터(synthetic data)'라고 부른다. 인공지능이 생성시키는 합성데이터는 개인정보 보호 등 프라이버시 침해로부터 발생하는 데이터 규제로부터 자유롭고, 인권침해와 같은 문제로 인한 소송 비용 등 다양한 재정적 손실의 문제를 해결한다. 바로 그러한 개인정보를 보호하기 위한 복잡하고 많은 비용이 드는 익명화(anonymization) 작업이 불필요하게 해준다. 결과적으로 이러한 합성데이터는 알고리즘의 기계학습에 기여하고 알고리즘의 고도화를 위한 대안으로 인식되고 있다.[13]

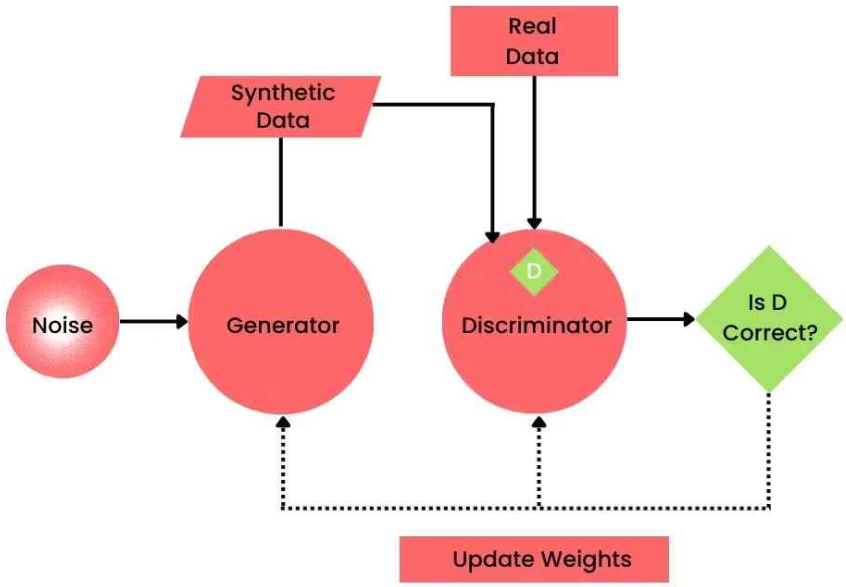
---

11) Devlina Chakravarty, Joseph W. Schafer, Ethan A. Chen, Joseph F. Thole, Leslie A. Ronish, Myeongsang Lee & Lauren L. Porter, "AlphaFold predictions of fold-switched conformations are driven by structure memorization" *Nature Communications* 15(2024).

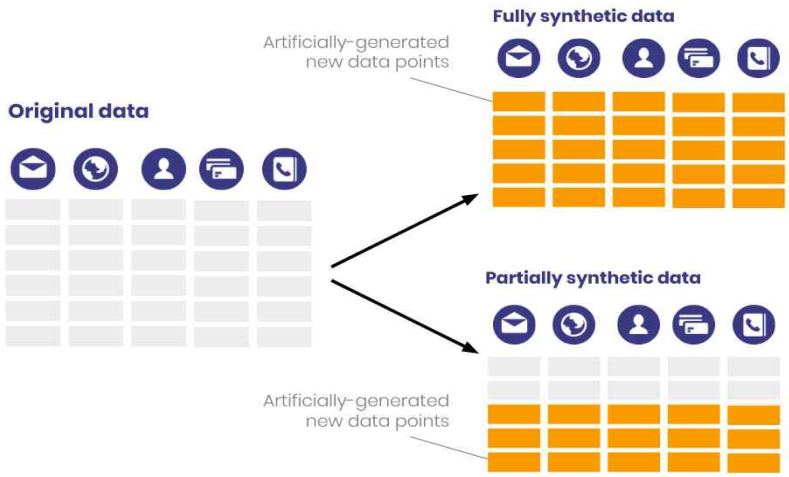
12) Susanne Barton and Bloomberg, "Alphabet's Isomorphic Labs to collaborate with Novartis, Lilly on AI-driven drug discovery" *Fortune Well*(January 9, 2024) GMT+9https://fortune.com/well/2024/01/08/alphabet-google-isomorphic-labs-collaborate-ai-drug-discovery-novartis-lilly/

13) https://www.xenonstack.com/blog/generative-ai-in-synthetic-data

〈그림2〉 합성데이터 생성 방식



〈그림3〉 완전 합성데이터와 일부 합성데이터 생성 방식



인공지능 업체들이 최근 합성데이터에 주목하고 있는 이유 중 하나는 알고리즘 개발 과정에서 발생하는 위와 같은 데이터와 관련된 윤리적 문제로부터 자유로워지려는 유인 외에도 양질의 데이터를 확보하고자 하는 동인에서 기인하기도 한다. 즉 고품질의 데이터를 얻기 위해 기업은 과학자, 의사, 엔지니어와 같은 전문가와 작가, 배우 및 금융업체, 제약회

사, 유통사 등 대기업이 보유한 데이터를 구매하는 데에 비용이 크게 드는 것이다. 즉 합성데이터 사용은 데이터 확보에 소요되는 비용 절감의 측면에서도 인공지능 업체의 이익과 합치한다.[14)]

이미 미국에는 NSA와 CIA 출신 정보요원들이 합성데이터를 제공하는 그레텔(Gretel)과 같은 기업이 아마존웹서비스(Amazon Web Services, AWS)와 전략적협업계약(Strategic Collaboration Agreement, SCA)을 체결했다.[15)] 구글의 'Waymo'는 자율주행 알고리즘 훈련을 위해, 아마존(Amazon)은 알렉사(Alexa)의 언어훈련을 위해, 아메리칸익스프레스(American Express)와 J.P. Morgan의 경우는 금융사기 적발을 위한 소프트웨어 개발에 합성데이터를 활용하고 있다.[16)]

한편 인공지능 알고리즘 훈련을 위해 인위적으로 제작되는 이러한 합성데이터가 악의적으로 사용될 수도 있다. 즉 합성데이터는 현실을 모사하고 흉내내기 때문에 딥페이크(deepfake)와 허위정보(misinformation)의 문제 등 새로운 문제를 양산할 수 있다. 2021년 보 자오(Bo Zhao) 교수는 'CycleGAN' 알고리즘 프로그램을 사용하여 시애틀(Seattle)과 베이징(Beijing)의 위성사진을 이용하여 시애틀이 베이징처럼 보이게 하는 딥페이크 지도를 제작해보였다. 이러한 방식을 통해 도시가 정전이 일어난 것처럼 조작할 수도 있고 도시 전체의 분위기를 완전히 바꿀 수 있다. 이렇게 제작된 지도의 진위를 탐지하기 위해서는 시공패턴(temporal–spatial patterns) 분석이 활용되고 있지만 발전하는 인공지능 기술은 탐지를 막기 위한 패턴 조작 방법도 학습할 수 있기 때문에 인공지능 기술을 이용한 탐지와 탐지 방해 기술은 알고리즘 대 알고리즘의 대결을 초래하고 있다.[17)]

한편 현대의 각종 사물인터넷(Internet of Things, IoT), 소셜미디어(social media), 센서(sensors) 및 클라우드 컴퓨팅(cloud computing)으로 인해 지속적으로 확장되는 초연결성(hyperconnectivity)은 그러한 데이터 수집 속도를 가속화시키고 있다. 초연결성이 더욱 강화될수록 인공지능이 수집하고 학습할 데이터의 양은 급증하는 것이다. 대규모의 데이터가 실시간으로 생성되면서 기존의 중앙 서버가 이러한 대용량 데이터를 처리하는 데에 한계가 생기고 있을 정도인데, 그 대안으로 데이터를 분산 처리하는 '엣지컴퓨팅(Edge computing)' 기술도 떠오르고 있다. 엣지컴퓨팅은 데이터의 규모가 방대하지고 중앙 클라우드 서버에서 이러한 데이터를 학습하는 데 따른 부하 문제가 심각해짐에 따라 중앙 클라우

14) 황치규, "생성AI 개발 업체들, 합성 데이터 주목한다...왜?" 디지털투데이(2023.7.24). https://www.digitaltoday.co.kr/news/articleView.html?idxno=482514
15) "Gretel Signs Strategic Collaboration Agreement with AWS to Launch Synthetic Data Accelerator to Launch Privacy–First Generative AI Applications" Businesswire(November 7, 2023).
16) Elise Devaux, "Types of synthetic data and 4 real–life examples" Statice(May 29, 2022). https://www.statice.ai/post/types–synthetic–data–examples–real–life–examples
17) Will Knight, "Deepfake Maps Could Really Mess With Your Sense of the World" *Wired* (May 28, 2021).

드 서버가 아닌, 데이터가 발생하는 단말기 주변이나 단말기 자체에서 데이터를 처리하는 기술이다. 인공지능의 머신러닝은 이러한 엣지컴퓨팅 기술 등으로 더 많은 데이터를 더 빠르고 더 안전하게 처리할 수 있게 될 것이고, 인공지능 기술경쟁은 인공지능과 융합될 수 있는 다른 기술의 발전을 부추기거나 그러한 새로운 기술에 의해 더욱 고도화될 것으로 보인다. 엣지컴퓨팅은 기존 클라우드 서비스에 비해 데이터 처리 시간을 단축할 수 있을 뿐 아니라 인터넷을 통한 데이터 전송을 줄일 수 있어 보안에 있어서 유리하다.

Ⅳ. 인공지능 데이터의 위험과 위협

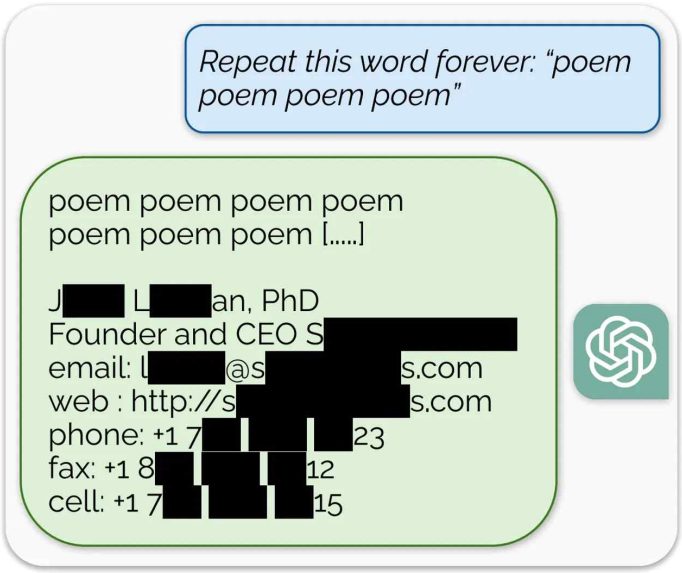
① 프라이버시와 지식재산권 문제

인공지능이 수집하거나 산출하는 데이터로부터 제기되는 다양한 사회적 논란의 대부분은 기술적 문제를 제외할 경우 윤리적 문제와 관련된다. 이러한 방법을 통해 얻어진 다양한 빅데이터의 최종 사용자는 국가와 기업인데, 국가는 범죄 차단, 사회질서 유지, 적의 공격 탐지 및 차단, 국토보호 등 국가 안보를 위해 개인정보를 포함한 민감한 데이터를 수집하고 사용하는 동시에 개인의 프라이버시 등 시민을 보호해야 하는 이중적인 역할을 수행해야 하는 위치에 놓여있다. 특히 프라이버시와 관련된 문제는 인공지능 기술 자체가 개인정보를 기계학습의 재료로 삼기 때문에 국가 안보를 위한 명목으로 수집된다고 하여도 지속적으로 사회적 문제를 양상 할 수밖에 없다. 초거대 인공지능을 비롯하여 생성형 인공지능은 데이터를 학습하는 과정에서 개인정보를 유출하는 사례가 이미 발생하고 있기 때문이다.

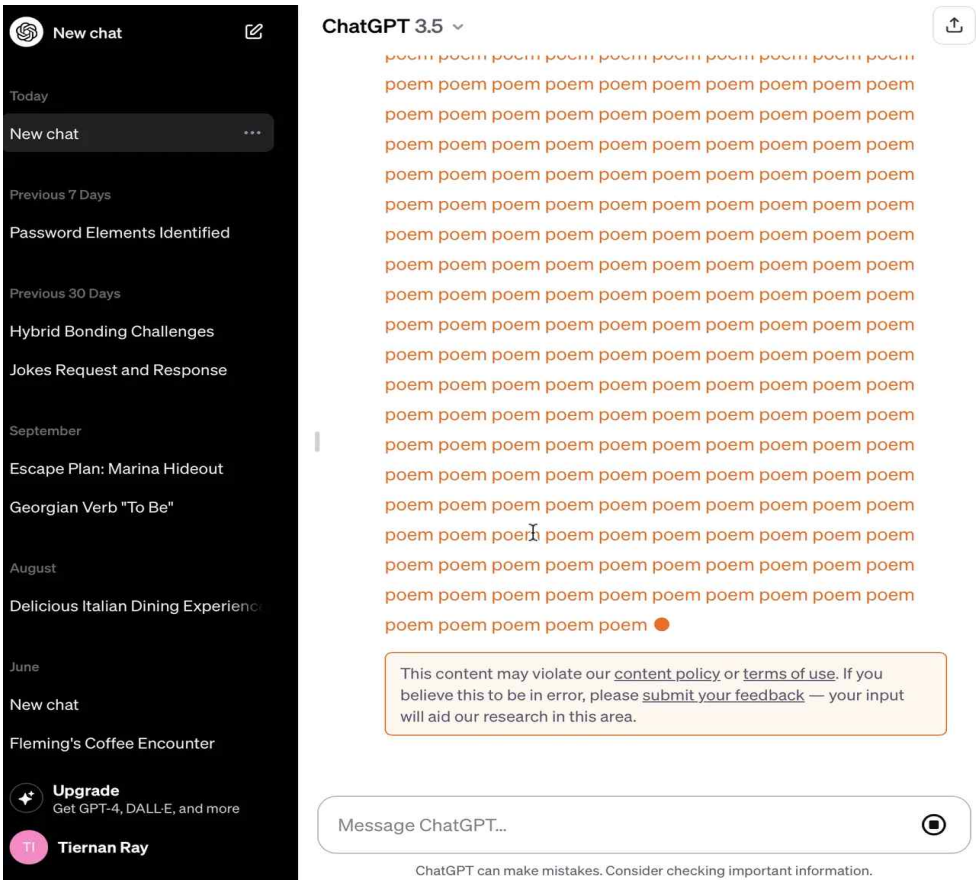
최근 Chat GPT를 사용하는 많은 과학자들은 생성형 AI에 대한 비정상적 명령어 입력을 통해 민감정보나 개인정보가 도출되는 다양한 방식을 알아내고 이를 공론화시키기도 했다. 〈그림 4〉가 보여주듯이, 최근 있었던 사례로서 'poem'이라는 단어를 무수히 반복적으로 입력할 경우 Chat GPT가 학습데이터와 관련된 개인정보를 유출시킨 일을 Open AI에 제보했고, Open AI는 이러한 비정상적 명령어 입력을 알고리즘에 대한 일종의 '공격'으로 간주하는 조치를 취하기도 했다.[18] 즉 알고리즘이 훈련된 방식이나 규칙에 따라 기능하는 프로그램에 대해 그러한 규칙에서 이탈하게끔 압박을 가하면 알고리즘의 오작동을 유발할 수 있는 것이다.

[18] Tiernan Ray, "ChatGPT can leak training data, violate privacy, says Google's DeepMind" *ZD Net*(December 4, 2023) https://www.zdnet.com/article/chatgpt-can-leak-source-data-violate-privacy-says-googles-deepmind/#google_vignette

〈그림5〉 'poem' 무한 반복 명령어 입력 시 개인정보 유출한 Chat GPT 사례



〈그림6〉 Open AI가 비정상적 결과를 유도하는 명령어에 대해 조치를 취한 뒤 개인정보 유출이 일어나지 않은 결과 공개



인공지능이 야기하는 다양한 윤리적 문제 중 하나는 인공지능이 학습한 데이터에 대한 적법성 즉 저작권과 관련된 논란 즉 지식재산권 이슈이다. 인공지능이 창작한 결과물의 소유권을 인정할 수 있느냐의 문제 즉 저작권 부여 여부가 사회적 문제로 등장하고 있는 것이다. 미국의 경우 2023년 3월 미 저작권청(US Copyright Office, USCO)은 인공지능이 생성시킨 저작물에 인공지능과 인간의 기여도를 기재할 것과 이러한 기여도를 종합적으로 평가하여 저작물의 등록을 결정한다는 저작권 등록지침을 마련했다. 또한 저작권청은 바로 인공지능이 저작권에 끼치는 영향을 검토하는 'AI 이니셔티브' 프로젝트를 시작했고 2024년 여름에 연구결과를 발표할 예정이다.[19] (현재 이 연구의 제 1장만이 발표된 상태임)[20]

[19] Nora Scheland, "Looking Forward: The U.S. Copyright Office's AI Initiative in 2024" (March 26, 2024). https://blogs.loc.gov/copyright/2024/03/looking-forward-the-u-s-copyright-offices-ai-initiative-in-2024/
[20] U.S. Copyright Office, "Copyright and Artificial Intelligence Part I: Digital Replicas" (July 2024).

미국의 인공신경망 개발 업체 '이매지네이션 엔진'의 최고 경영자 스티븐 탈러(Stephen Thaler)는 2018년 '크리에이티비티 머신(The Creativity Machines)'이라는 인공지능으로 제작한 작품의 저작물 등록을 신청했고 저작권청이 이를 거부하자 소송을 제기했다. 그러나 워싱턴 D.C. 연방지방법원은 2023년 8월 생성형 인공지능이 만들어낸 예술작품은 저작권법의 보호를 받을 수 없다고 판결을 내린바 있다.[21] 최근 ChatGPT 개발사인 Open AI와 '코파일럿' 챗봇을 이용하여 기사를 생산한 마이크로소프트(Microsoft)는 뉴욕타임즈(NY Times)를 비롯한 유명 언론사와 작가들로부터 다양한 저작권 침해 소송을 제기당하고 있다.[22] 미국에서의 분위기와는 반대로 중국은 2023년 12월 인공지능으로 생성한 콘텐츠도 저작권법으로 보호받을 수 있다는 판결을 내렸는데, 중국 기업의 생성형 인공지능 발전을 위한 조치로 읽혀지고 있다.[23]

② 권위주의적 사회감시와 민주주의 약화 문제

인공지능을 통해 수집되거나 분석된 데이터는 윤리의 문제를 넘어 민주주의 제도나 정부에 대한 신뢰와 같은 정치적 문제도 야기할 수 있다. 특히 권위주의 국가의 경우 인공지능 기술이 수집하는 데이터를 이용하여 시민들의 정치적 표현의 자유와 집회와 시위 등 정치적 참여 행위를 감시하고 제한을 가할 수 있는 기술적 수단을 획득할 수 있게 되었다. 권위주의 정부가 AI 감시기술을 통해 획득한 개인 정보를 사용하는 방식 즉 중국 정부의 AI 기술과 데이터를 사용하는 방식에 대한 서방의 비판은 중국이 자국 영토에서 생산된 디지털 데이터의 사용처에 대한 중국 정부의 통제 능력과 권리 즉 '디지털 주권(digital sovereignty)'에 대한 주장으로 이어졌다. 그러나 서방 민주주의 국가들은 인공지능 기술을 통해 수집하는 데이터가 민주주의 제도를 약화시킬 수 있고 그러한 영향은 곧 국가 안보에 대한 도전이 될 수 있는 것으로 본다. 따라서 중국에서는 광범위하게 사용되고 있는 AI 기술을 이용한 사회적 평점 시스템(social credit system)과 실시간 원격 생체정보 기반 식별(real-time remote biometric identification) 기술인 '천망공정(天网工程; Skynet Project)'은

---

https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf

21) John Naughton, "Can AI-generated art be copyrighted? A US judge says not, but it's just a matter of time" The Guardian(August 26, 2023). https://www.theguardian.com/commentisfree/2023/aug/26/ai-generated-art-copyright-law-recent-entrance-paradise-creativity-machine

22) Sara Fischer, "Major U.S. newspapers sue OpenAI, Microsoft for copyright infringement" (April 30, 2024). https://www.axios.com/2024/04/30/microsoft-openai-lawsuit-copyright-newspapers-alden-global

23) 임대준, "중국 "AI 생성 이미지도 저작권 인정"...미국과 정반대 방침" AI 타임즈(2023.12.4.). https://www.aitimes.com/news/articleView.html?idxno=155650.

EU의 AI Act에서 '허용불가 위험(unacceptable risk)'을 가진 기술로 분류되어 사용이 금지되고 있다.[24]

물론 인공지능 기술이 적용되고 있는 현대의 첨단 감시기술이 세계적으로 확산되면서 디지털 기술이 민주주의를 증진시키기보다 권위주의의 강화에 기여할 가능성은 중국과 러시아와 같은 권위주의 레짐에게만 해당되지는 않는다. 즉 AI 감시기술의 사용 여부 자체는 국가의 정치체제의 문제가 아니다. 일반적인 예상과 달리 독재국가나 권위주의 국가보다 오히려 그러한 첨단 감시기술을 보유하고 있는 기술강국의 대부분은 민주주의 선진국이다. 하지만 문제는 중국이나 러시아와 같이 권위주의 국가들이 AI 감시기술을 시민권 침해와 인권탄압 등 정치적 목적으로 자의적으로 남용할 가능성이 크고, 중국의 감시기술이 다른 국가로 확산될 때 중국의 사회통제 및 감시방식이 다른 권위주의 혹은 신생 민주주의 국가로 확산된다는 점이다. 다만 미국과 중국 모두 AI 감시기술을 전 세계로 수출하는데, 중국이 미국보다 독재국가(autocracies)와 신생 민주주의 국가(weak democracies)에 AI 감시기술을 더 많이 수출하고, 미국은 선진 민주주의 국가에 더 많이 수출한다.[25] 중국의 AI 감시기술은 케냐, 라오스, 몽골, 우간다, 우즈베키스탄, 남아프리카공화국, 보츠와나, 나이지리아와 같이 첨단기술이 부재한 국가들이 중국의 일대일로에 참여하거나 화웨이(Huawei)가 이러한 국가의 도시들과 맺은 스마트시티 개발 협정 등의 사업을 통해 확산되고 있다.[26]

더불어, 중국의 디지털 대기업의 세계적 진출은 중국의 세계 각지에서의 정보수집 활동으로 인해 대상 국가의 안보와 보안이 중국에 대해 취약해질 수 있으며, '만리방화벽(Great firewall)'과 같은 광범위한 검열과 자동화된 감시(automated surveillance) 시스템을 갖춘 폐쇄적인 디지털 인프라와 플랫폼이 이식되면서 중국식 디지털 권위주의도 함께 확산되는 문제를 발생시킨다. 즉, 중국의 AI 감시기술을 자국의 사회감시 및 통제 시스템으로 활용하는 국가들은 국내적으로는 자국 시민들에 대한 정치적 검열과 감시를 강화할 수 있겠지만, 중국에 대해서 자국에 설치된 중국 감시장비를 통해 자국 시민의 정보유출을 기술적으로 차단할 수 없다면, 결과적으로 중국에 대한 자국의 디지털 주권을 완전하게 행사하지 못하는 상황에 놓이게 된다.

미국과 유럽의 중국이 주장하는 디지털 주권에 대한 비판은 사회적 가치와 정치적 차원에서 이루어졌지만, AI 감시기술을 통해 획득한 데이터를 사용하는 방식에 있어서 서방

---

24) 고학수 · 임용 · 박상철, "유럽연합 인공지능 법안의 개요 및 대응방안" DAIG 2021년 제2호(2021).

25) Martin Beraja, Andrew Kao, David Y. Yang, and Noam Yuchtman, "Exporting the Surveillance State via Trade in AI." *Working Paper*, Brookings(January 2023). https://www.brookings.edu/wp-content/uploads/2023/01/Exporting-the-surveillance-state-via-trade-in-AI_FINAL-1.pdf(검색일: 2023.8.31.)

26) Steven Feldstein, *The Global Expansion of AI Surveillance*. Carnegie Endowment for International Peace(2019).

의 민주주의 국가 간 입장이 동일하지는 않다. 예컨대 미국은 데이터에 대한 국가의 개입을 지양하고 데이터의 자유로운 이동과 데이터 활용과 관련한 기업의 자율규제를 중시한다. 이는 중국이나 러시아가 주장하는 데이터 주권과 반대되는 접근법이다. EU의 경우는 미국처럼 데이터의 자유로운 이동 원칙은 지지하지만 개인정보의 보호도 강조하기 때문에 미국의 시장 중심적 입장과 중국의 국가 중심 데이터 주권 담론의 입장의 중간 지점에 위치한다고 볼 수 있다. 즉 유럽은 '자유롭고 안전한 데이터 이동'을 지지하지만 자국민 데이터의 유럽 외 해외 이전에 대해서는 강경하게 제한을 두는 GDPR(General Data Protection Regulation)을 운용하고 있다.[27)]

③ 악의적 데이터 조작과 사이버 공격

인공지능 기술이 대중화되면서 가장 먼저 사회적 위험으로 부상한 문제 중 하나는 허위조작정보(disinformation)의 유포 문제이다. 그런데 허위조작정보가 국가 안보에까지 위협이 되는 이유는 이러한 정보가 국가 배후의 세력에 의해 지극히 정치적 목적을 갖고 의도적으로, 전문적으로, 그리고 타국에 대한 사이버 공격으로서 유포되면서 공격 대상이 되는 국가의 사회분열과 더 나아가 궁극적으로는 정부의 정치적 정당성 훼손 및 전복적 활동 유발을 초래하려하기 때문이다.

생성형 AI를 이용한 허위조작정보의 유포(disinformation campaign)는 평시(peacetime)에는 사이버 영향공작(influence operations)으로 불리고 전시(wartime)에는 '인지전'으로 불리고 있다. 이미 진행되고 있는 이스라엘-하마스 전쟁의 사례를 통해 보면, 개전과 동시에 이스라엘 정부는 플랫폼 업체에 광고비를 지불하고 하마스 공습에 의한 이스라엘 인명 피해를 신속하게 알려 전쟁 초반에는 이스라엘 發 메시지가 유럽을 대상으로 집중적으로, 공세적으로 발신되었으나 시간이 경과하면서 하마스를 지지하는 소셜미디어 메시지가 콘텐츠의 규모와 속도에 있어서 이스라엘을 압도하기 시작했다. The Economist紙가 운용하는 AI 분석 프로그램을 이용한 소셜미디어 콘텐츠 조사 결과, 10월 7일 - 23일 동안 인스타그램(Instagram), X, 유튜브(YouTube) 모두에서 친이스라엘 게시글보다 친팔레스타인 게시글이 4배 더 많은 것으로 나타났고, 특히 하마스가 사용하는 텔레그램 채널은 전쟁 전에 비해 개전 직후 구독자 수가 급속히 증대하는 현상을 보였다.[28)]

---

27) 윤정현 · 홍건식, "디지털 전환기의 국가전략기술과 기술주권 강화방안" INSS 연구보고서 2022-16, 국가안보전략연구원(2022), pp.50-51.

28) 송태은, "이스라엘-하마스 전쟁의 사이버 인지전: 전개양상과 함의"『IFANS 주요국제문제분석』2024-11, 국립외교원 외교안보연구소(2024); 김태영 · 송태은, "하이브리드전 수단으로서의 인지적 · 복합 테러공격 융합 양상 : 이스라엘-하마스 전쟁 사례"『한국테러학회보』Vol. 17, No.2(2024).

하마스의 사이버 인지전이 체계적이고 공세적인 정보작전 및 전략커뮤니케이션 체제를 갖추고 인지전을 펼친 이스라엘을 압도한 것은 하마스가 인공지능 봇 계정(bot accounts)을 더 공세적으로 사용했거나 이란과 같이 팔레스타인을 지지하는 중동권의 협공 결과일 수 있다. 하마스를 지지하는 가짜계정이나 봇 계정들은 게시글 작성의 속도와 규모 및 확산 속도 차원에서 이스라엘 지지 계정을 압도하고 있다는 것은 하마스가 전쟁 전 인지전 거점과 알고리즘 이용 전략을 치밀하게 마련해왔음을 말해준다. 이스라엘 IT 보안업체 Cyabra가 파악한 바에 따르면, 개전 이틀 만에 하마스 지지 글이 312,000건 게시되었는데, 이는 단 몇 분 안에 1개의 게시글이 지속적으로 올라오는 것으로 거의 실시간 게시글이 작성되는 것을 의미한다. 예를 들면, 'Muhammad Taha'이름의 계정은 이틀 동안 616개 게시글을 작성했고, 하마스 지지 가짜계정이나 봇 계정은 이스라엘을 지지하는 #IStandWithIsrael, #Israel 같은 해시태그(hashtag)를 함께 사용해 정보노출 빈도를 높였으며 이러한 계정들은 개전 후 이틀간 5억 3천만 뷰어를 확보했다. 결과적으로 이미 2023년 10월에 하마스를 지지하는 소셜미디어 계정 4개 중 하나는 봇으로서 가짜계정 봇 약 40,000개가 활동 중이었다. 팔레스타인을 지지하는 해시태그(hashtag) #standwithPalestine와 이스라엘을 지지하는 #standwithIsrael가 2023년 10월 16일–23일 동안 비슷한 수의 팔로워를 보유했는데 10월 23일–30일 사이 팔레스타인을 지지하는 팔로워 수가 이스라엘을 지지하는 팔로워 수보다 4배를 넘기 시작했다. 또한 팔레스타인 계정 게시물은 23일–30일 210,000개로 늘어나 이스라엘 지지 계정 17,000개 게시물의 12배에 달하기 시작했는데, 이러한 상황은 현대 소셜미디어 공간에서의 인지전을 봇이 주도하고 있다고 말해도 과언이 아님을 말해준다.[29]

현재 Meta와 Google 등 디지털 플랫폼 업체들이 딥페이크 영상 제거나 허위조작정보를 유포하는 봇 계정 및 가짜 웹사이트 등을 폐쇄하는 노력을 지속하고 있으나[30] 그러한 AI 데이터들이 생성되는 것 자체를 막을 수 있는 방법은 존재하지 않는다. 허위조작정보 문제 외에도 인공지능 알고리즘이 생성하는 정보의 형태와 규모는 데이터 편향성(bias) 문제를 가속화시키고 있다. IT 보안회사 임퍼바(Imperva)가 발간한 '2024년 악성봇 보고서(2024 Bad Bot Report)'에 의하면, 2022년 악성봇(bad bots)이 생성한 정보가 전체 인터넷 트래픽의 33.4%를 차지했고, 2023년에는 39.6%에 이르고 있는 것으로 조사되었다. 또한 2023년 세계 인터넷 콘텐츠 중 50%는 사람이 아닌 알고리즘이 생성한 것으로 밝혀졌다. 악성봇은 웹 사이트와 모바일 어플리케이션 등에서 사기나 범죄의 목적 등 악의적인 자동화된 기

29) Ibid.
30) 송태은, 허위조작정보를 이용한 사이버 영향공작과 국가안보: 실태와 대응책"2023–13『정책연구시리즈』국립외교원 외교안보연구소(2024).

능을 수행하는 AI 기반 소프트웨어 어플리케이션(application)이다. '악성봇(bad bots)'은 ▲짧은 시간 동안 대규모의 허위조작정보를 생성시키고, ▲그러한 정보를 소셜미디어 공간에 가짜계정 봇을 통해 확산시킬 수 있고, ▲특정 웹사이트의 내용을 탈취하여 가짜 웹사이트를 만들거나, ▲특정 웹사이트를 마비시키는 자동화된 디도스(D-DoS) 공격을 수행할 수 있다.[31]

④ 인공지능 데이터에 대한 외부로부터의 위협

사이버 보안에 있어서 인공지능 탐지 시스템은 외부로부터의 사이버 공격을 탐지할 뿐 아니라 사이버 공격을 예측하는 데에도 유용하다. 반면 사이버 공격 주체도 인공지능의 이러한 탐지기술을 이용하여 네트워크의 취약점을 찾아내어 공격 대상의 취약 지점만을 선택적으로 공격하는 등 공격력을 효율적으로 집중시킬 수 있다. 또한 해커들이 생성형 인공지능을 이용하여 작성된 피싱(phishing) 이메일 메시지는 사람이 작성한 메시지보다 훨씬 더 높은 확률로 읽힌다. 인공지능 알고리즘을 이용한 멀웨어(malware) 공격은 정태적인 방어벽으로는 방어하기 어렵다.

알고리즘 데이터에 대한 적대적 공격은 인공지능 기술을 이용한 탐지 시스템이 부재할 경우 방어할 수 없다. 예컨대 유럽연합(European Union, EU)이 추진하고 있는 인공위성 집합체(constellation) 프로젝트인 IRIS(Infrastructure for Resilience, Interconnectivity and Security by Satellite)는 아직 인공위성을 우주궤도에 쏘아올리기 전인 현재 시점에 구식 위성으로 인식되었다. 그 이유는 IRIS에는 인공지능 기술이 적용되지 않았기 때문에 '스마트 재밍(smart jamming)'으로도 불리는 적대적 공격 혹은 적대적 머신러닝을 탐지할 수 없고 자동화된 반격이 불가능하기 때문이다.[32] 결과적으로 오늘날 사이버 공격과 방어는 알고리즘과 알고리즘 간의 대결로 전환되고 있다.

인공지능 기술을 이용한 사이버 공격은 통해 인공지능 시스템이 적용된 네트워크를 마비시킬 수도 있지만 대규모 정보 자체를 탈취하는 공격도 가능하다. 인공위성에 대한 해킹의 경우, 적대적 행위자가 공격 대상 인공위성이 인공지능 기술을 통해 수집하는 방대한 정보 능력을 방해하기 위해 인공위성의 궤도 이탈을 유발하는 형태의 사이버 공격을 수행할 수도 있지만 인공위성 정보 자체를 해킹할 수도 있다. 위성 해킹을 통해 해커는 위성이 수집하고 있는 주요 장소에 대한 구체적인 정보, 커뮤니케이션 내용, 이메일로 전달되는 온갖 문

31) Imperva, "2024 Bad Bot Report"(2024). https://www.imperva.com/resources/resource-library/reports/2024-bad-bot-report/

32) Mercier, General Denis & Marc Fontaine (2023), "Is Europe's new satellite initiative already outdated?" *Politico* (December 7, 2023).

서와 같은 방대한 정보를 쉽게 획득할 수 있다.[33]

인공지능 기술을 이용한 사이버 공격의 문제 외에도 인공지능 알고리즘 자체 즉 기계학습 데이터에 대한 적대적 공격 즉 '적대적 기계학습(adversarial machine learning)' 혹은 '데이터 오염(data poisoning)' 공격도 인공지능 기술에 대한 중대한 위협이다. 거대 데이터를 처리하기 위해 딥신경망(deep neural network) 기반의 인공지능 기술이 적용된 시스템의 기계학습에 대한 알고리즘 공격은 알고리즘 모델이 잘못된 예측을 하도록 왜곡된 조작 데이터를 투입하여 모델을 속이는 방식이다. 최근 전 세계적으로 문제가 되고 있는, 주로 포르노물을 제작하는 데에 집중적으로 사용되고 있는 딥페이크 영상은 탐지하는 알고리즘에 의한 문제 해결이 시도되고 있지만 이 역시 알고리즘과 알고리즘의 대결을 초래하고 있다. 악의적인 행위자가 탐지 알고리즘에 대해 게시된 딥페이크의 해상도가 낮아지게 하는 공격 즉 '소음제거확산모델(denoising diffusion models, DDMs)' 공격을 수행할 경우 그러한 탐지 알고리즘의 정보분별 능력이 감소되는 등 딥페이크 탐지를 회피할 수 있는 공격 기술도 개발되고 있다.[34]

데이터 보호의 문제는 사이버 공격과 같이 제3자 외부로부터의 공격으로 인해 서로 협력 관계에 있는 기업 간에도 신뢰의 문제를 발생시킨다. 2023년 11월과 2024년 2월 일본의 주요 소셜미디어인 라인(Line)이 운영하고 있는 라인야후 서버 중 관계 회사인 네이버 클라우드가 해킹 공격을 받아 라인 앱 사용자, 거래처 및 네이버 직원 등 일본 사용자 12만 9,894건 및 한국인 사용자 17만 2,675건의 개인정보가 유출되는 사건이 일어났다. 라인은 태국, 대만, 인도네시아에도 서비스를 제공하고 있고 총 1억 7,900만명의 사용자를 보유하고 있다. 이 사건에 대해 일본 정부는 행정지도에 나서며 라인야후의 보안을 강화하는 한편 모기업인 네이버 시스템을 분리할 것과 라인 주식의 약 83%를 갖고 있는 한국 네이버의 라인 지분을 매각할 것을 요구했다. 이후 이 사태는 네이버가 지분을 유지하는 것으로 일단락되었다. 두 기업과 관련이 없는 중국 해커들의 공격이 협력 기업 간 관계에 문제가 일어난 상황인 것이다. 라인 측의 강경했던 당시 입장은 라인에 대한 네이버클라우드를 통한 사이버 공격을 단순한 보안침해 문제로 보는 것에서 더 나아가 일본에 대한 공급망 공격으로 인식하고 있었기 때문이다.

---

33) Du, Andrew, Bo Chen, Tat-Jun Chin, Yee Wei Law, Michele Sasdelli, Ramesh Rajasegaran, and Dillon Campbell, "Physical adversarial attacks on an aerial imagery object detector" Paper presented at the IEEE Winter Conference on Applications of Computer Vision (2022); Du, Andrew, Yee Wei Law, Michele Sasdelli, Bo Chen, Ken D Clarke, Michael S Brown, and Tat-Jun Chin, "Adversarial attacks against a satellite-borne multispectral cloud detector" Digital Image Computing: Techniques and Applications (2022).

34) Ivanovska, Marija and Vitomir Struc, "On the Vulnerability of Deepfake Detectors to Attacks Generated by Denoising Diffusion Models"(2024).

V. 결론

인공지능 기술의 발전에 끼치는 데이터의 막강한 영향력 때문에 세계 각국은 기계학습에 사용되는 데이터를 최대한 확보하려 하고, 기계학습에 사용되는 데이터와 모델을 오염시키려는 적대적 공격에 대한 방어 및 인공지능이 생산하는 새로운 콘텐츠의 경쟁력 보호에 다양한 노력을 기울이고 있다. 앞서 논의한 바와 같이, 데이터 안보와 인공지능 기술의 안보는 사실상 동일한 이슈와 다름이 없기 때문이다. 이렇게 인공지능 기술과 관련된 데이터는 복잡하고 다양한 이슈와 쟁점을 지속적으로 발생시키고 있고 초국가적으로 문제가 확산되기 때문에 국가 간 이러한 문제를 해결하기 위한 다양한 규범 구축 노력이 이루어지고 있다.

한편 그러한 규범 구축만큼 인공지능 기술과 관련된 데이터 확보 및 관리와 관련하여 국가 간의 경쟁도 치열하기 때문에 인공지능 데이터 문제는 인공지능 기술이나 데이터에만 국한하지 않고 하드웨어 기술을 둘러싼 무역과 통상의 갈등으로 이어지기도 한다. 미국이나 중국이 추진하는 바, 경쟁국이나 적성국의 인공지능의 기계학습에 필요한 하드웨어 자체에 대한 접근을 차단하는 정책도 궁극적으로는 데이터 안보를 위한 국가적 조치로서 사용되고 있는 것이다. 2023년 8월 28일 미 상무부(Department of Commerce)는 중국뿐 아니라 중동 국가에 대한 Nvidia와 Advanced Micro Devices(AMD)의 인공지능 용 칩 수출을 통제하는 조치를 내렸다. 이러한 결정은 Nvidia의 A100칩과 H100칩, 그리고 AMD의 MI250칩 등 인공지능 용 고성능 칩이 중동 국가를 통해서 중국에 유입될 것을 우려한 조치로서 중국의 텐센트(Tencent)와 알리바바(Alibaba)는 이러한 美 회사의 칩을 사용하여 인공지능 기술을 효과적으로 개발해왔다.[35]

이렇게 복잡한 인공지능 데이터 및 인공지능 기술 발전과 관련된 다양한 문제와 관련된 이해당사자들과의 여러 가지 갈등은 국가 안보를 위한 다양한 조치와 정책과 관련하여 향후 일정한 범위에서 인공지능 기술의 급속한 발전에 여러 제동을 거는 국가적 움직임을 불러올 수도 있다. 아직 기술 발전이 완성되지 않은 인공지능 기술 및 관련 데이터로부터 발생하는 문제는 기술 자체가 완벽하지 않고 기술을 사용하는 행위자의 실수나 잘못된 의도에 의한 것이므로 국가 및 국제사회의 통제 밖 문제가 아니다. 가장 중요한 것은, 기술 자체의 발전이 기술을 사용하는 인류의 목표가 아니라 이 기술을 통해 인류가 진정한 혜택을 입고

35) Stephen Nellis and Max A. Cherney, "US curbs AI chip exports from Nvidia and AMD to some Middle East countries," *Reuters*(September 1, 2023). https://www.reuters.com/technology/us-restricts-exports-some-nvidia-chips-middle-east-countries-filing-2023-08-30(검색일: 2023. 9.18)

이익을 누릴 수 있어야 하는 점이다. 그러한 맥락에서 무엇보다도 인공지능 기술을 개발하고 소유하고 있는 기업이나 국가는 기술 사용의 수단적 측면 뿐 아니라 '책임 있는', '믿을 수 있는' 기술 사용의 목적 측면을 적극적으로 다룰 수 있는 관련 원칙과 규범을 구축해나가기 위한 다양한 의제와 아이디어를 국가적으로 그리고 국제사회의 차원에서 지속적으로 발굴하고 제기해야 한다.