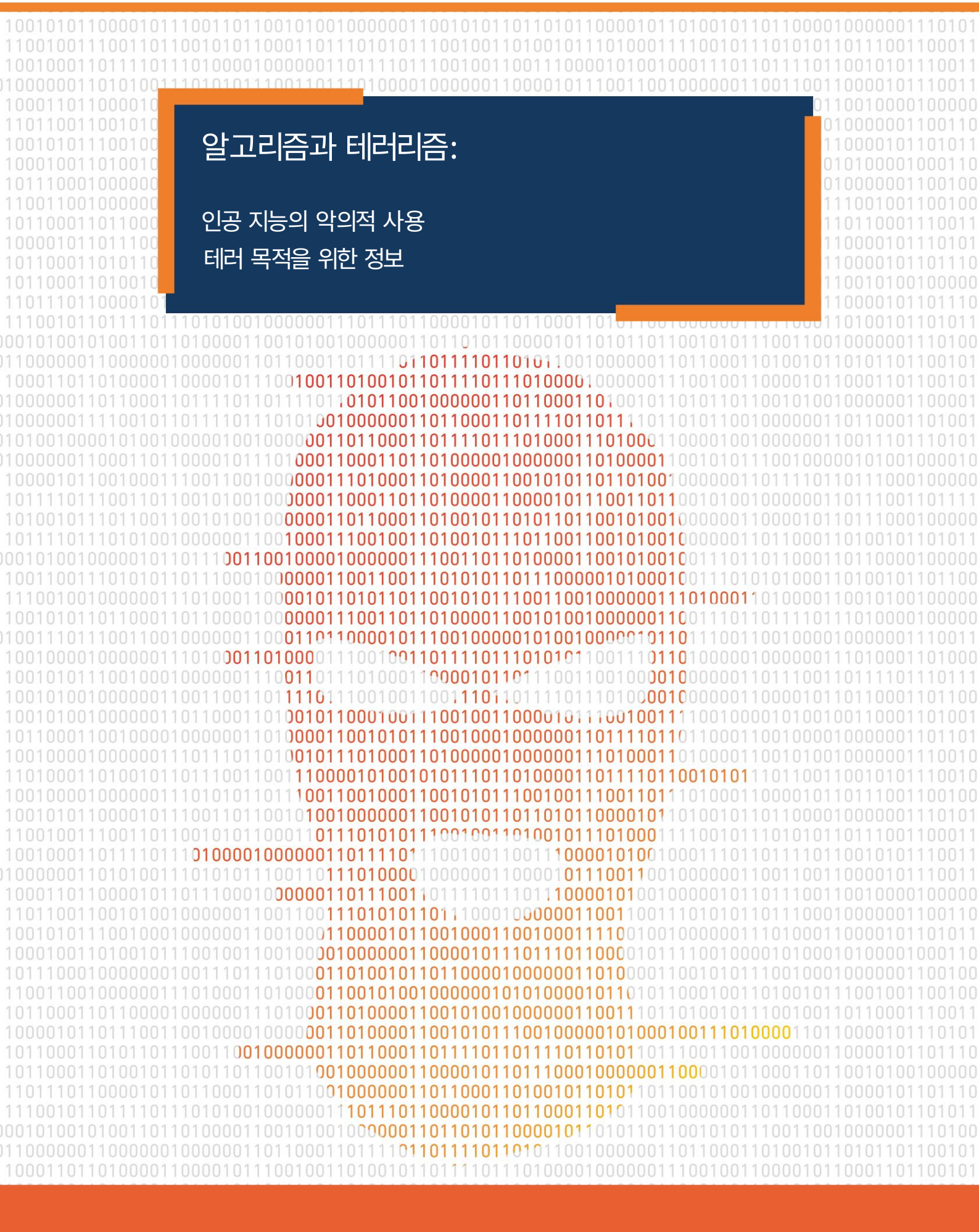
UNITED NATIONS
OFFICE OF COUNTER-TERRORISM
UN Counter-Terrorism Centre (UNCCT)

unicri
United Nations
Interregional Crime and Justice
Research Institute

알고리즘과 테러리즘:

인공 지능의 악의적 사용
테러 목적을 위한 정보



알고리즘과 테러리즘:

인공 지능의 악의적 사용 테러리스트를 위한 정보 목적

UNICRI와 UNCCT의 공동 보고서

부인 성명

여기에 표현된 의견, 조사 결과, 결론 및 권장 사항은 반드시 유엔, 사우디아라비아 정부 또는 관련된 다른 국가적, 지역적 또는 글로벌 기관의 견해를 반영하는 것은 아닙니다.

본 간행물에 사용된 명칭과 제시된 자료는 유엔 사무국이 특정 국가, 영토, 도시 또는 그 관할 구역의 법적 지위나 국경 또는 경계의 설정에 관해 어떠한 의견도 표명한 것을 의미하지 않습니다.

감사의 말

본 보고서는 유엔 대테러사무소(UNOCT) 산하 유엔 대테러센터(UNCCT) 사이버 보안 및 신기술부, 그리고 유엔 지역간 범죄 및 사법 연구소(UNICRI)의 인공지능 및 로보틱스 센터가 공동으로 추진하는 인공지능 시대의 대테러 연구 사업의 결과물입니다. 본 공동 연구 사업은 사우디아라비아의 아낌없는 지원으로 진행되었습니다.

저작권

유엔 테러대응사무소
S-2716
유엔
405 이스트 42번가
뉴욕, 뉴욕 10017
웹사이트: www.un.org/counterterrorism/

Viale Maestri del Lavoro,10, 10127 토리노 – 이탈리아
웹사이트: www.unicri.it 이메일:
unicri.publicinfo@un.org

머리말

지난 10년 동안 공공 및 민간 부문을 포함한 다양한 산업 분야에서 인공지능(AI) 솔루션이 빠르게 도입되는 것을 목격했습니다. 2025년까지 전 세계 AI 시장 규모는 1,000억 달러를 돌파할 것으로 예상되며, AI 기반 시스템은 의료, 교육, 상업, 은행 및 금융 서비스, 중요 인프라, 보안 등 여러 분야를 지속적으로 지원할 것입니다.

안토니우 구테흐스 유엔 사무총장은 2018년 신기술 전략에서 "이러한 기술은 큰 가능성을 지니고 있지만, 위험이 없는 것은 아니며, 일부는 불안감과 심지어 두려움을 불러일으킬 수 있습니다. 악의적인 목적으로 사용되거나 의도치 않은 부정적인 결과를 초래할 수 있습니다."라고 밝혔습니다. AI가 인류에게 미칠 수 있는 잠재적 이점은 부인할 수 없지만, AI의 악의적 사용에 대한 연구는 아직 초기 단계에 있습니다.

테러리스트들은 신기술을 조기에 수용하는 경향이 있는데, 이러한 신기술은 규제와 관리가 부족한 경향이 있으며, AI도 예외는 아닙니다. 많은 기술 시스템이 국제적인 연계성과 국경을 초월하는 영향을 고려할 때, 테러리스트들이 AI 시스템의 취약점을 노출시킬 수 있는 규제 공백을 악용할 기회를 갖지 못하도록 지역적, 국제적 접근이 필수적입니다. 테러리스트의 악의적인 AI 사용에 신속하고 효과적으로 대응하고 그 영향을 완화할 수 있는 회복력 있는 거버넌스 구조를 구축해야 합니다.

유엔은 다양한 이니셔티브를 통해 이러한 요구에 부응하고 있습니다. 사무총장의 디지털 협력 로드맵은 "인공지능에 관한 국제 협력 지원"을 8대 핵심 행동 분야 중 하나로 설정하고 있습니다. 이 로드맵에 따라, 유엔 대테러사무소 산하 유엔 대테러센터 또한 사이버 보안 및 신기술에 관한 국제 대테러 프로그램을 통해 이러한 과제에 대응하고 있습니다.

유엔 지역간 범죄 및 사법 연구소와 공동으로 개발한 이 보고서는 테러리스트가 AI를 악의적으로 사용하거나 남용할 가능성을 조기에 경고하는 역할을 하며, 글로벌 커뮤니티, 산업계, 정부가 새로운 기술이 해를 끼치지 않고 유익한 결과를 가져오도록 공동으로 무엇을 할 수 있는지 사전에 생각하는 데 도움이 될 것입니다.

이 기회를 빌려 이 보고서의 권고안 작성에 참여해 주신 국제 전문가 여러분께 감사드리고자 합니다. 저희 사무국은 테러리스트의 AI 위협에 대응하기 위해 회원국 및 기타 대테러 파트너들을 지원할 준비가 되어 있습니다.

블라디미르 보론코프
사무차장
유엔 테러대응사무소
전무이사
유엔 테러대응센터

머리말

인공지능(AI)은 우리 시대의 핵심 신기술이라고 해도 과언이 아닙니다. 유엔 지역간 범죄 및 사법 연구소(UN Interregional Crime and Justice Research Institute)는 산하 AI 및 로보틱스 센터를 통해 수년간 AI를 연구해 왔으며, 그 결과는 매우 유망합니다. 예를 들어, 오늘날 AI는 메신저 리보핵산(mRNA) 기반 백신 개발을 크게 앞당기는 데 중요한 역할을 해왔습니다. 이러한 백신은 현재 COVID-19 팬데믹을 억제하는 데 사용되고 있습니다. 사법, 범죄 예방, 안보, 법치주의 분야에서 우리는 AI의 유망한 활용 사례를 목격해 왔습니다. 예를 들어, 오랫동안 실종된 아동을 찾고, 불법 성매매 광고를 스캔하여 인신매매 조직을 적발하고 차단하며, 자금 세탁을 암시하는 금융 거래를 적발하는 데 활용될 수 있습니다.

하지만 우리는 AI의 어두운 측면도 목격했습니다. AI는 그동안 많은 관심을 받지 못했고, 제대로 연구되지도 않은 측면입니다. 현실은 악의적인 의도로 사용될 경우 매우 위험할 수 있다는 것입니다. 사이버 범죄 분야에서 입증된 실적을 바탕으로, AI는 테러리즘과 테러리즘을 조장하는 폭력적 극단주의를 조장하거나 촉진하는 데 활용될 수 있는 강력한 도구입니다. 예를 들어 드론이나 자율주행차를 이용한 새로운 물리적 공격 방식을 제공하거나, 중요 인프라에 대한 사이버 공격을 강화하거나, 증오 발언과 폭력 선동을 더 빠르고 효율적으로 확산시킬 수 있습니다.

AI가 테러리즘의 미래일까요? 이 보고서에서 지적했듯이, 앞으로 어떻게 될지는 아직 알 수 없습니다. 그럼에도 불구하고 테러리즘은 과소평가되어서는 안 될, 끊임없이 진화하는 위협이라는 사실을 결코 잊어서는 안 됩니다. 21세기가 시작된 지 20년이 넘은 지금, 우리는 테러리스트들이 드론, 가상화폐, 소셜 미디어와 같은 새롭고 떠오르는 기술에 눈을 돌리는 수많은 사례를 목격했습니다. AI의 접근성이 점점 높아짐에 따라, AI의 오용과 관련된 모든 상황에 대비하고 앞서 나가는 것이 필수적입니다.

따라서 사우디아라비아 왕국의 아낌없는 지원으로 유엔 대테러사무소 산하 유엔 대테러센터와 함께 이 보고서를 발표하게 되어 자랑스럽게 생각합니다. 이 보고서가 테러 목적으로 AI를 악용하는 것에 대한 논의의 시작이 되기를 바랍니다.

안토니아 마리 드 메오
감독
유엔 지역간 범죄 및 사법 연구소

요약

신기술, 특히 인공지능(AI)은 의학, 정보통신기술, 마케팅, 운송 등 여러 연구 분야에서 큰 발전을 가능하게 하는 매우 강력한 도구가 될 수 있습니다. 하지만 이러한 기술이 악의적인 목적으로 악용될 가능성도 있습니다. 본 보고서(알고리즘과 테러리즘: 테러 목적으로 인공지능을 악용하는 것)의 범위는 AI가 테러리스트의 손에 들어갈 잠재적 위험을 이해하는 데 기여하는 것입니다.

테러 조직들은 전통적으로 총기, 칼, 차량 등 다양한 형태의 "로우테크 테러리즘"을 어느 정도 사용해 왔지만, 테러리즘 자체는 정체된 위협이 아닙니다. AI가 더욱 보편화되면, AI 활용에 필요한 기술과 전문성이 줄어들어 진입 장벽이 낮아질 것입니다.

따라서 이 보고서가 답하고자 하는 질문은 AI가 테러리즘의 도구가 될 것인지, 아니면 언제 될 것인지에 대한 질문이며, 만약 그렇게 된다면 국제 사회가 합리적으로 무엇을 기대할 수 있는지에 대한 질문입니다.

이 보고서는 9개 장으로 구성되어 있습니다.

1장에서는 전반적인 개요를 제공하고, 테러리스트를 포함한 전문가들 사이에서 이 기술이 악의적으로 사용되고 있다는 우려가 커지고 있음을 보여주는 통계 자료를 제공합니다.

2장에서는 AI의 전반적인 모습을 설명합니다. 먼저 머신러닝과 딥러닝을 포함한 AI 및 관련 용어, 그리고 좁은 지능과 일반 지능 등의 개념을 정의합니다. 이어서 자연어 처리, 이미지 인식 등 AI 알고리즘과 응용 프로그램의 혜택을 받는 다양한 분야와 이 기술 활용의 미래 동향을 개괄적으로 살펴봅니다.

3장에서는 인터넷과 소셜 미디어와 같은 기술이 가치 있고 강력한 도구로 활용된 테러 공격의 여러 사례를 제시하여 테러 집단과 테러리스트가 새로운 기술을 사용할 경우 잠재적인 위협이 될 수 있음을 보여줍니다.

4장에서는 기존 문헌에서 확인된 세 가지 위협 범주(사이버 위협, 물리적 위협, 정치적 위협)를 조사하여 AI의 악의적 사용에 대한 맥락을 더욱 자세히 살펴보고, AI가 어떻게 악의적으로 사용될 수 있는지 보여줍니다.

5장에서는 AI 기반 테러리즘이 현실로 실현될 수 있는지, 아니면 단순한 공상과학 소설에 불과한 것인지에 대한 질문을 다룹니다. 이를 위해 AI 또는 관련 기술에 관심을 보인 테러 집단의 사례를 제시합니다. 여기에는 얼굴 인식이나 무인 항공 시스템(드론)을 활용한 영상이 포함됩니다.

이어서 6장에서는 테러 단체와 개인에 의한 AI의 현재 및 향후 악의적 사용 사례에 대한 심층적인 개요를 제공합니다. 이 개요에는 연구를 통해 문서화되고 확인된 악의적 사용 사례와, 증거나 문헌이 부족함에도 불구하고 미래에 실제로 발생할 수 있는 악의적 사용 사례가 모두 포함됩니다.

7장에서는 AI가 테러 목적으로 악용될 수 있는 방식을 시각화하기 위해 세 가지 가상 시나리오를 제시합니다. 이 시나리오들은 AI 기반 비밀번호 추측, 랜섬웨어, 안면 인식 드론, 딥페이크, 그리고 "범죄 서비스(crime-as-a-service)" 사업 모델에서 지하 포럼을 통해 제공되는 변조된 여권의 활용에 초점을 맞춥니다.

이전 장에서 제시된 정보를 바탕으로, 8장에서는 테러 집단과 개인이 AI를 직접 사용하여 공격을 개선하거나 증폭하는 것에 대해 우려할 만한 사유가 있는지 평가합니다. 이와 관련하여, 객관적인 결론을 도출하기 위해 의도와 역량의 개념을 분석합니다.

9장에서는 보고서를 마무리하며, 테러 방지 기관과 법 집행 기관, 정책 입안자, 업계, 학계가 미래를 위해 고려해야 할 일련의 권장 사항을 제시하고, AI 기반 테러의 가능한 미래에 대비하기 위한 역량 강화를 위한 몇 가지 후속 조치를 제안합니다.

본 보고서를 작성하는 과정에서 UNOCT와 UNICRI는 주로 논문, 공식 보고서, 언론 보도 등 데스크 기반 연구와 오픈소스 정보에 의존했습니다. 2021년 2월 9일 전문가 그룹 회의가 온라인으로 개최되어 오픈소스 정보를 바탕으로 도출된 결론을 보완하고, 제시된 전략적 권고안과 후속 조치에 대한 통찰력을 수집했습니다.

내용물

나. 소개 10

2. AI란 무엇인가? 13

나. 주요 용어 및 기본 개념 13

ii. 알고리즘 및 응용 프로그램 15

iii. 진화하는 기술 16

3장. 알고리즘과 테러리즘의 위협 17

4. AI 위협 유형 분류 21

다섯. 사실인가 공상과학인가? 22

우리. 거울 속의 AI 기반 테러리즘 26

i. 사이버 역량 강화 27

에이. 서비스 거부 공격 27

비. 악성코드 28

기음. 랜섬웨어 29

디. 비밀번호 추측 30

그리고. CAPTCHA 해독 31

에프. 암호화 및 복호화 32

ii. 물리적 공격 가능 33

에이. 자율주행차 33

비. 얼굴 인식 기능이 있는 드론 34

기음. 유전자를 표적으로 삼은 생물무기 35

iii. 테러리즘 자금 조달 수단 제공 36

에. 오디오 딥페이크 36

비. 암호화폐 거래 37

iv. 선전 및 허위 정보 확산 39

에. 딥페이크 및 기타 조작된 콘텐츠 39

v. 기타 작전 전술 41

에. 감시 41

비. 소셜 네트워킹 플랫폼에서의 가짜 온라인 신원 및 인간 사칭 43

기. 변형된 여권 44

디. 온라인 소셜 엔지니어링 45

7. AI의 테러적 활용을 파헤치다 46

VIII. 위협 평가 49

에. 의도 50

비. 능력 50

기. 우려할만한 사항? 52

9. 평가에서 실행으로 55

I. 서론

인공지능(AI)은 강력한 도구입니다. 공공 및 민간 부문 전반에서 AI는 사람들과 사회 전반을 더 행복하고 건강하며 부유하고 안전하게 만들기 위해 활용되고 있습니다. 유엔 사무총장 안토니우 구테흐스는 AI가 적절하게 활용되고 유엔 헌장과 세계인권선언이 정의한 가치와 의무에 기반을 둔다면, 빈곤 종식, 지구 보호, 그리고 모든 사람의 평화와 번영 보장에 기여함으로써 2030 지속가능개발의제 달성에 기여할 수 있다고 지적했습니다.1 그러나 AI는 어두운 면도 있습니다. 범용 기술로서 AI는 악의적인 행위자에 의해 악용될 수 있습니다. 유로폴, 트렌드마이크로, UNICRI가 최근 발표한 보고서에 따르면 사이버범죄자들은 이미 AI를 공격 벡터와 공격 표면으로 사용하는 여러 가지 방법을 강조했습니다.2, 3, 4 AI가 범죄 목적으로 사용될 수 있는 것처럼, 테러 공격의 강도를 높이거나 극단주의 선전을 유포하고 폭력을 부추길 수 있는 이러한 집단이나 개인의 잠재력을 증폭시키기 위해 집단이나 개인이 악의적으로 사용할 수도 있습니다.5

2020년 8월, MIT 테크놀로지 리뷰 인사이트(MIT Technology Review Insights)는 301명의 고위 경영자와 학자들을 대상으로 AI 관련 다양한 사안에 대해 설문 조사를 실시했습니다. 설문 조사 결과에 따르면, 투명성 부족, 편견, AI 개발의 거버넌스 부재, 그리고 자동화로 인한 심각한 실업 가능성 등이 우려 사항으로 지적되었지만, 응답자들은 AI가 잘못된 손에 들어가는 것을 가장 우려했습니다.6

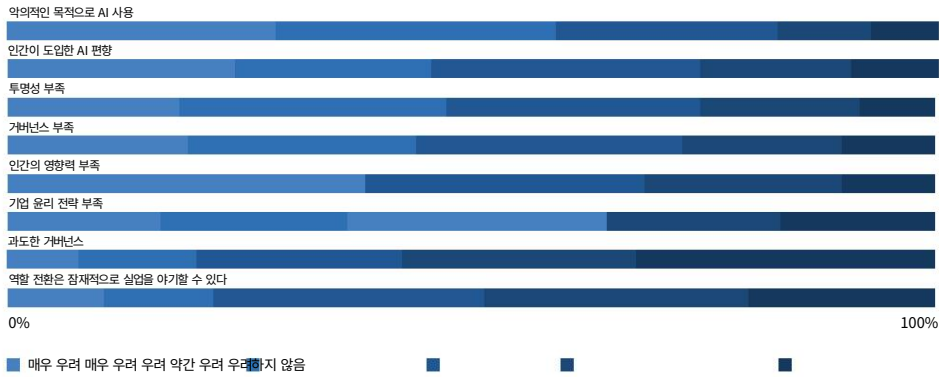


그림 1: MIT Technology Review Insights는 AI의 악의적 사용에 대한 우려가 널리 퍼져 있음을 보여줍니다.

1 안토니우 구테흐스. (2018년 9월). 유엔 사무총장 신기술 전략. https://www.un.org/ 에서 접속 가능 en/newtechnologies/images/pdf/SGs-Strategy-on-New-Technologies.pdf

2 빈첸초 샤카글리니, 크레이그 깁슨, 데이비드 산초, 필립 아만, 아글리카 클레인, 오드란 맥카시, 마리아 에아라. (2020년 11월 19일). 인공지능의 악의적인 사용 및 남용. 트렌드마이크로 리서치. http://unicri.it/sites/default/files/2020-11/AI%20 에서 확인 가능 MLC.pdf
본 보고서의 결과는 본 보고서의 출발점으로 활용되며, 트렌드마이크로, 유로폴, 그리고 UNICRI의 기여를 바탕으로 작성되었습니다. 이러한 기여는 2020년 3월 트렌드마이크로, 유로폴, 그리고 UNICRI가 공동 주최한 집중 워크숍에서 수집된 의견과 결합되었습니다. 워크숍에는 공동 사이버범죄대책위원회(J-CAT), 국제형사재판소(ICC), 그리고 유로폴 산하 유럽 사이버범죄센터(EC3) 자문단 소속 위원들이 참석했습니다.

3 유니버시티 칼리지 런던(UCL)의 도스 미래 범죄 센터(Dawes Centre for Future Crime)에서도 AI의 범죄적 활용에 대한 주목할 만한 연구들이 진행되었습니다. 연구진은 매튜 콜드웰, 제론 TA 앤드루스, 토마스 타네이, 루이스 그리핀입니다. (2020년 7월). 정책 브리핑: AI 기반 미래 범죄. 런던대학교 도스 미래 범죄 센터. https://www.ucl.ac.uk/jill-dando-institute/sites/jill-dando-in-stitute/files/ai_crime_policy_0.pdf 에서 확인 가능

4 Link11. AI와 사이버 회복력: 공격자와 방어자의 경쟁. (2020년 11월 5일). Link11. https://www.link11.com/ 에서 접속 가능 영어: en/downloads/ai-and-cyber-resilience-a-race-between-attackers-and-defenders/

5 이 보고서의 목적에 따라: 테러리스트의 기술 사용은 테러리스트가 의도한 대로 주어진 기술을 사용하는 것을 수반하는 것으로 간주됩니다. 즉, 테러리스트가 동료와 소통하기 위해 메시징 애플리케이션을 사용하는 것입니다. 테러리스트의 기술 오용은 테러리스트가 소셜 미디어 플랫폼을 사용하여 폭력을 부추기는 등, 정해진 약관에 위배되는 방식으로 기술을 사용하는 것을 수반하는 것으로 간주됩니다. 기술의 악의적 사용은 사용과 오용을 모두 포괄하는 더 광범위한 용어로, 주로 주어진 기술이 사용되는 의도를 나타냅니다.

6 MIT 테크놀로지 리뷰 인사이트(2020년 11월 30일). 새로운 지평: AI 환경 확장. MIT. https://www.technology-review.com/2020/11/30/1012528/a-new-horizon-expanding-the-ai-landscape/ 에서 확인 가능

유엔 대테러센터(UNCCT) 산하 유엔 대테러사무소(UNOCT)와 유엔인공지능로봇연구소(UNICRI)가 산하 인공지능 및 로봇공학센터를 통해 본 보고서 검토 및 검증을 위해 개최된 전문가 그룹 회의에서 실시한 설문조사에서도 유사한 우려가 드러났습니다. 정부, 산업계, 학계, 국제 및 지역 기구 관계자 27명 중 44%는 테러 목적으로 AI가 악의적으로 사용될 가능성이 "매우 높다"고 답했고, 56%는 "다소 높다"고 답했습니다. 주목할 점은 이러한 방식으로 AI가 악의적으로 사용될 가능성이 "낮다"고 답한 설문 참여자가 없었다는 것입니다.7

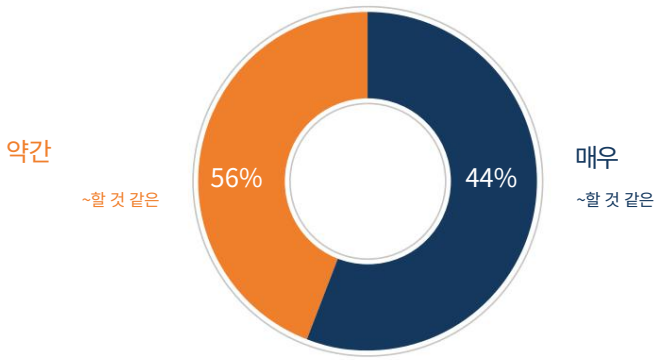


그림 2: 테러 목적으로 AI를 악의적으로 사용할 가능성에 대한 UNOCT-UNICRI 설문 조사

설문조사에 따른 토론에서 참가자들은 테러 목적으로 AI를 악의적으로 사용할 가능성에 대한 우려를 크게 키우는 요인으로 다음 네 가지를 꼽았습니다.

첫째, AI와 같은 신기술의 "민주화"입니다. "민주화"라는 개념은 한때 상당한 자원과 전문성을 가진 매우 제한된 공동체만이 이해하고 사용하던 매우 진보된 기술이 이제 모든 사람이 쉽게 접근할 수 있게 되었으며, 대규모 투자나 제한된 기술 지식으로도 사용할 수 있게 되었다는 사실을 의미합니다. 실제로, 널리 사용되는 알고리즘 중 상당수는 이미 오픈 소스이며, 사용하는 데 특별히 높은 수준의 전문성이 필요하지 않습니다. 기술의 민주화가 일반적으로 발전과 번영의 원동력이 될 수 있지만, 결과적으로 악의적인 사용의 위험 또한 그만큼 높아집니다. 더욱이, 이러한 집단이 범죄 집단이 사용하는 점점 더 중요해지는 "서비스로서의 범죄(crime-as-a-service)" 비즈니스 모델 형태로 아웃소싱할 가능성을 고려할 때, "거의 모든 유형의 사이버 범죄를 용이하게 하는 광범위한 상업 서비스를 제공함으로써 디지털 지하 경제를 촉진하는", 악의적인 의도를 가진 행위자들의 AI 사용 진입 장벽이 상당히 낮아졌습니다.8

둘째, AI의 확장성입니다. 확장성은 기술의 규모나 규모가 "성장"하고 증가하는 수요를 관리할 수 있는 능력으로 이해될 수 있습니다. 일반적으로 확장성은 기술의 활용 측면에서 더 크고 광범위하게 만드는 것을 의미합니다. 참가자들은 AI의 확장성이 특히 뛰어나다는 점을 고려할 때, AI의 잠재적 악의적 사용을 방어하는 담당자는 개별 공격의 위협뿐만 아니라 특정 시점에 증가하는 공격 규모에도 대비하고 방어해야 한다는 점을 지적했습니다. 이에 대한 대표적인 예가 드론 무리가 자율적으로 일제히 비행하는 위협입니다.

셋째, 테러리즘과 대테러리즘의 본질적인 비대칭성입니다. 예를 들어, 개별 공격자의 기술적 경험 부족 등으로 인해 AI를 테러 목적으로 악의적으로 사용하는 것이 실패하더라도, 공포를 심어주는 측면에서 상당한 심리적 영향을 미칠 수 있다는 주장이 제기되었습니다. 예를 들어, 실패한 폭탄 공격은 그럼에도 불구하고 강력한 메시지를 전달합니다. 따라서 대테러리즘의 맥락에서 흔히 볼 수 있듯이, "우리는 매번 운이 좋아야 하지만, 그들은 단 한 번만 운이 좋으면 된다"는 것입니다. 동시에, 이러한 비대칭성은 다음과 같은 과제에서도 드러납니다.

7 UNOCT-UNICRI 전문가 그룹 회의의 통찰력. (2021년 2월 9일). 전문가 그룹 회의에는 오스트리아 공과대학, AWO, Chemonics, 유럽 평의회, 유럽 위원회 - 이민 및 내무 총국(DG HOME), 유럽 연합 법 집행 협력 기관(Europol), 러시아 연방 외무부, Chatham House, 제네바 안보 정책 센터, 유럽 안보 협력 기구(OSCE), Link11 사이버 복원력, MalwareBytes, 북대서양 조약 기구(NATO), Trend Micro, 영국 연구 및 혁신(UKRI) 신뢰할 수 있는 자율 시스템(TAS) 허브, 유엔 대테러 위원회 집행 이사회(CTED), 유엔 대학교, 칼리지 런던, 브리스톨, 케임브리지, 사우샘프턴 대학교의 대표들이 참석했습니다.

8 인터넷 조직범죄 위협 평가(IOCTA). (2014년 9월 29일). 유로폴. https://www.europol.europa.eu/sites/de- fault/files/documents/europol_iocta_web.pdf 에서 확인 가능

대테러 기관과 테러리스트들은 AI 사용과 관련하여 직면한 문제들을 안고 있습니다. AI를 활용하고자 하는 많은 기관들은 시민의 자유와 기본적인 인권 및 자유를 보호하기 위해 AI 사용에 대한 신중한 고려가 필요합니다.
그러나 테러 집단이나 개인 등 악의적인 행위자들은 이런 우려에 깊이 빠지지 않을 가능성이 높으며, 이로 인해 그들이 AI를 활용할 가능성이 여러 면에서 단순해집니다.

Unsplash의 Thomas Jensen이 찍은 사진

넷째, 데이터와 기술에 대한 사회적 의존도 증가. 사회 전체가 인터넷의 무결성과 가용성, 그리고 인터넷 작동을 위한 데이터의 신뢰성에 점점 더 의존하고 있다는 점이 지적되었습니다. 최근 몇 년간 AI가 발전함에 따라, 의료 서비스 제공업체, 에너지 제공업체, 생물학 및 원자력 시설과 같은 주요 기반 시설을 포함한 스마트 기기와 스마트 시티를 통해 AI가 일상생활에 빠르게 통합되고 있습니다.

이는 많은 이점을 제공하지만, AI 기반 사이버 공격이나 이러한 인프라 내의 AI 시스템이나 이러한 시스템이 작동하는 데이터에 대한 보다 전통적인 공격에 대한 취약성이 높아진다는 것을 의미합니다.

이 보고서에서 설명하겠지만, 테러 조직이 AI를 실제로 직접 사용하고 있다는 명확한 증거나 징후는 없습니다. 하지만 다른 분야, 특히 오랫동안 이 기술을 선구적으로 도입해 온 사이버 범죄자들의 AI 악용 동향과 발전 상황을 참고하고 교훈을 얻을 수 있습니다. 이러한 맥락에서, 그리고 최근 몇 년간 AI 산업의 기하급수적인 성장과 앞서 언급한 요인들을 고려할 때, 테러 목적으로 이 기술을 악용할 가능성은 국제 사회의 세심한 주의를 요합니다.

II. AI란 무엇인가?

AI가 어떻게 사용될 수 있는지, 또는 경우에 따라서는 테러 목적으로 오용될 수 있는지 이해하려면 먼저 기술 자체에 대한 기본적인 이해를 확립하는 것이 필수적입니다.

i. 주요 용어 및 기본 개념

AI는 시각 인식, 음성 인식, 언어 간 번역, 의사 결정, 문제 해결 등 일반적으로 인간의 지능을 필요로 하는 작업을 수행할 수 있는 컴퓨터 시스템의 이론 및 개발에 전념하는 컴퓨터 과학 분야입니다.9 이러한 지능형 시스템에는 소프트웨어 애플리케이션, 로봇, 자율주행차 등이 포함될 수 있습니다. AI는 다양한 하위 분야를 포괄하는 용어이며, 그중 가장 중요한 하위 분야는 다음과 같습니다.

Unsplash의 Joshua Hoehne이 찍은 사진

머신 러닝은 AI의 하위 분야로, 데이터로부터 "학습"할 수 있는 알고리즘, 즉 특정 작업의 성능을 점진적으로 향상시킬 수 있는 알고리즘을 포함합니다. 다른 컴퓨터 소프트웨어와 달리, 머신 러닝 알고리즘은 인간의 명시적인 지시를 필요로 하지 않습니다. 대신, 데이터베이스에 포함된 상당수의 예제에서 패턴을 추출하고 암묵적인 규칙을 학습합니다.10 따라서 AI 시스템은 특정 작업을 수행하는 머신 러닝 알고리즘과 해당 작업을 수행하는 데 필요한 센서 및 외부 장치를 포함할 수 있습니다. 예를 들어, 컴퓨터 비전 AI 시스템은 이미지 인식 소프트웨어와 알고리즘이 처리할 이미지를 캡처하는 하나 이상의 카메라로 구성됩니다.

딥 러닝은 머신 러닝의 하위 분야로, 신경망이라는 더 작은 알고리즘 군을 다룹니다. 이 알고리즘은 인간의 뇌에서 영감을 받아, 반복적으로 작업을 수행하면서 방대한 데이터로부터 학습하고, 매번 내부 기능을 미세하게 수정하여 결과를 개선합니다.

"딥 러닝"이라는 용어는 신경망의 여러(또는 "깊은") 계층에서 유래되었습니다.11

9 스튜어트 J. 러셀, 피터 노비그. (2009). 인공지능: 현대적 접근(제3판). 프렌티스 홀.

10 톰 미첼. (1997). 머신러닝. 맥그로힐.

11 이언 굿펠로우, 요슈아 벤지오, 에런 쿠르빌. (2016). 딥러닝. MIT 출판부. www.deeplearningbook.org 에서 열람 가능

아래 이미지는 AI, 머신러닝, 딥러닝의 관계를 보여줍니다.

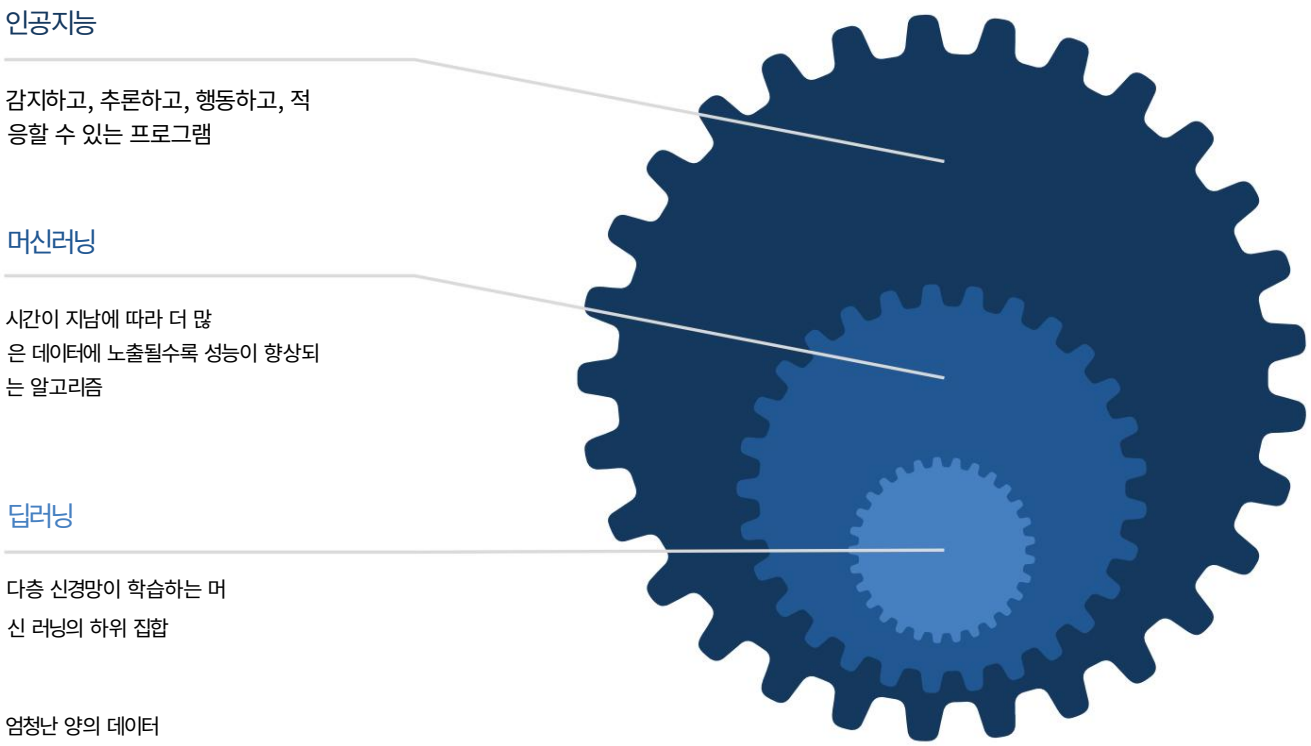


그림 3. AI와 주요 하위 분야 간의 관계

현재 존재하는 AI 시스템은 소위 "좁은" AI 애플리케이션으로 구성됩니다. 이는 날씨 예측, 체스 두기, 의료 이미지 분석과 같은 단일 작업을 수행하도록 프로그래밍된 AI 시스템입니다.
이러한 시스템은 "좁은" 프로그래밍으로 인해 원래 설계된 단일 작업 외에는 성능이 좋지 않습니다. 그러나 특정 작업에서 탁월한 성능을 발휘함으로써, 가까운 미래에 개발될 수 있는 더욱 지능적인 AI 시스템의 구성 요소 역할을 할 수 있습니다.

이와 관련하여, 문헌에서 흔히 등장하는 또 다른 개념은 인공 일반 지능(AGI)입니다. 이는 인간이 할 수 있는 모든 지적 작업을 성공적으로 수행할 수 있는 시스템을 의미합니다. 제한된 작업만을 수행하도록 설계된 좁은 범위의 AI와 달리, AGI는 학습, 계획, 추론, 자연어로 소통하고, 이러한 모든 기술을 통합하여 모든 작업에 적용할 수 있습니다. AGI는 오랫동안 AI의 성배였으며, 전문가들은 AGI의 도래 여부와 도래한다면 언제 도래할지에 대해 광범위하게 논의해 왔습니다.12

AGI를 넘어서는 개념은 인공지능 초지능(ASI)입니다. 이는 모든 면에서 인간 지능을 능가할 수 있는 기계를 지칭하는 개념입니다.13 창의성부터 문제 해결 능력까지, 초지능 기계는 개인으로서뿐만 아니라 사회적으로도 인간 지능을 능가할 것입니다. 이러한 유형의 AI는 많은 철학적 논쟁을 불러일으켰으며, 일부 전문가들은 심지어 인류에게 실존적 위협을 가할 수도 있다고 주장합니다.14

12 Hal Hodson. (2019년 3월 1일). DeepMind와 Google: 인공지능을 장악하기 위한 싸움. 1843 Magazine. https://www. 에서 확인 가능. economist.com/1843/2019/03/01/deepmind-and-google-the-battle-to-control-artificial-intelligence

13 닉 보스트롬. (2014). 초지능: 경로, 위험, 전략. 옥스퍼드 대학교 출판부.

14 로리 셀란-존스. (2014년 12월 2일). 스티븐 호킹, 인공지능이 인류를 멸망시킬 수 있다고 경고. BBC. https://www.bbc.에서 접속 가능. com/news/technology-30290540

ii. 알고리즘 및 응용 프로그램

딥 러닝 패밀리 내에는 다양한 신경망 아키텍처가 있으며, 이를 통해 다양한 응용 프로그램이 가능합니다.

합성곱 신경망(CNN 또는 ConvNet)은 이미지 분석에 가장 많이 사용되는 신경망의 한 종류입니다.15 동물 시각 피질에서 영감을 받은 이 알고리즘은 여러 층의 단일 유닛 또는 노드를 사용하여 원시 입력에서 고차원 특징을 점진적으로 추출합니다. 예를 들어, 입력이 이미지인 경우, 신경망의 첫 번째 층은 선과 곡선을 식별하고, 마지막 층은 문자나 얼굴을 식별할 수 있습니다. 이러한 특성 덕분에 CNN은 객체를 식별할 수 있으며,17 이를 통해 객체 인식과 나아가 얼굴 인식이 가능해집니다.18

많은 관심을 받고 있는 또 다른 분야는 자연어 처리(NLP)입니다. NLP에서 가장 자주 사용되는 아키텍처 유형은 순환 신경망(RNN)19입니다. RNN에서 네트워크 노드는 시간적 순서를 따라 연결됩니다. 입력 시퀀스를 처리하는 내부 메모리에 의존하여, RNN을 사용하는 기계는 단어 시퀀스와 그 형태 구문 및 의미 기능을 이해함으로써 음성 인식을 수행할 수 있습니다.20

음성 인식 외에도 NLP는 텍스트 생성에도 활용될 수 있으며, 이는 온라인에서 작동하고 기본적인 대화를 시뮬레이션하도록 프로그래밍할 수 있는 소프트웨어 프로그램인 "챗봇"21의 기반이 됩니다. 텍스트와 이미지와 같은 콘텐츠 생성은 생성적 적대 신경망(GAN)이라는 또 다른 유형의 신경망 덕분에 가능합니다. 2014년에 발명된 이 혁신적인 아키텍처는 이후 딥러닝 분야에 혁명을 일으켰습니다.22

GAN 모델은 생성 네트워크(generative network)와 판별 네트워크(discriminative network)라는 두 가지 인공 신경망으로 구성됩니다. 생성 네트워크는 훈련 데이터셋과 동일한 특성을 가진 새로운 데이터를 생성하는 반면, 판별 네트워크는 생성된 데이터를 훈련 데이터셋에서 분리합니다. 예를 들어, 사진으로 훈련된 GAN은 저장된 사진 데이터셋과 유사한 새로운 이미지를 생성할 수 있습니다. 판별 네트워크는 훈련 데이터셋을 기반으로 생성 네트워크가 생성한 사진을 무작위로 수신하여 원본 이미지와 식별하려고 합니다. 생성 네트워크의 목표는 점점 더 나은 새로운 후보를 생성하여 판별 네트워크의 성능을 "속이는" 것입니다.23

GAN은 텍스트, 이미지, 노래, 그리고 다양한 형태의 예술을 생성하는 등 다양한 용도로 활용됩니다.24 GAN은 또한 널리 알려지고 뜨거운 논쟁을 불러일으키는 "딥페이크" 현상의 배후에 있습니다. "딥 러닝"과 "가짜 미디어"의 합성어인 딥페이크는 2017년에 발명된 일종의 합성 미디어입니다. 딥페이크는 AI 기술을 사용하여 사람이나 기술 솔루션조차도 진짜 콘텐츠와 즉시 구별할 수 없는 가짜 시각 및 청각 콘텐츠를 조작하거나 생성합니다.25

15 이언 굿펠로우, 요슈아 벤지오, 에런 쿠르빌. (2016). 딥러닝. MIT 출판부. www.deeplearningbook.org 에서 열람 가능

16 David Hubel과 Torsten Wiesel. (1959년 10월). 고양이의 줄무늬 피질에 있는 단일 뉴런의 수용 영역. J. Physiol. 148(3): 574–91.

17 Dan Ciresan, Ueli Meier, Jonathan Masci, Luca M. Gambardella, Jurgen Schmidhuber. (2011). 유연하고 고성능의 합성곱 이미지 분류를 위한 신경망. 제22회 국제 인공지능 공동 학술대회 논문집(Vol-ume Volume Two). 2권, 1237–1242쪽.

18 스티브 로렌스, C. 리 자일스, 아 청 초이, 앤드류 D. 백. (1997). 얼굴 인식: 합성곱 신경망 접근법. IEEE 신경망 거래. 8(1): 98–113. CiteSeerX 10.1.1.92.5813

19 Richard Socher, Cliff Lin, Andrew Y Ng, Christopher D Manning. 재귀적 신경망을 이용한 자연 장면 및 자연어 분석 네트워크. 제28회 기계 학습 국제 컨퍼런스(ICML 2011).

20 Samuel Dupond. (2019). 신경망 구조의 현재 발전에 대한 심층적인 고찰. 제어 연감. 14: 200–230.

21 바보라 자소바. (2020년 1월 2일). 자연어 처리 챗봇: 일반인을 위한 가이드. Landbot.

22 Ian J. Goodfellow 외 (2014년 6월 10일). arXiv. "생성적 적대 신경망." 2020년 7월 1일 접속, https://arxiv.org/복근/1406.2661.

23 위와 같음.

24 제이슨 브라운리. (2019년 7월 12일). 머신러닝 마스터리. 생성적 적대 신경망(GAN)의 인상적인 응용. 액세스- https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/ 에서 확인 가능

오스카 슈워츠 25. (2018년 11월 12일). 가디언. 가짜 뉴스가 나쁘다고 생각하셨나요? 딥페이크는 진실이 죽어가는 곳입니다. https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth 에서 확인하세요.

Unsplash에 Ilya Pavlov가 찍은 사진

인터넷, 소셜 미디어, 메시징 애플리케이션의 도달 범위와 속도와 더불어 딥페이크는 극히 짧은 시간 안에 수백만 명의 사람들에게 빠르게 도달할 수 있습니다. 이 때문에 딥페이크는 오늘날의 허위 정보 전쟁에서 강력한 무기로 인식되고 있으며, 사람들은 더 이상 보고 듣는 것에 의존할 수 없게 되었습니다.26

iii. 진화하는 기술

AI가 의학, 경제, 통신, 보험, 금융 등 여러 분야로 확장됨에 따라 AI 관련 제품의 예상 시장은 매일 성장하고 있습니다. 예를 들어, AI는 드론이나 자율주행차와 같은 자율주행 차량, 검색 엔진, 온라인 스마트 비서, 스팸 필터링, 그리고 맞춤형 온라인 광고에 사용됩니다. 또한 시를 비롯한 다양한 예술 작품을 창작하고, 수학적 정리를 증명하고, 체스나 바둑과 같은 게임을 즐기고, 다소 논란의 여지가 있지만, 사법 판결의 결과를 예측하고 알리는 데에도 사용됩니다.27

민간 산업계, 학계, 정부 기관의 투자 증가에 따라 머신러닝 알고리즘은 더욱 정교해질 것입니다. 센서와 모바일 앱을 통한 지속적인 데이터 수집과 더불어 머신 알고리즘 학습에 활용 가능한 데이터는 기하급수적으로 증가하여 AI 분야의 상당한 발전을 가져올 것이며, 이는 결국 머신이 할 수 있는 일의 한계를 더욱 넓혀줄 것입니다.28

AI 시스템의 발전은 로봇 지능에 큰 발전을 가져올 것으로 예상됩니다. 향상된 "감각"과 민첩성을 갖춘 로봇이나 로봇 도구는 과거에는 자동화하기에는 너무 섬세하거나 비경제적이라고 여겨졌던 작업들을 수행할 수 있게 될 것입니다. 더욱이, 고급 AI의 역량이 확장됨에 따라 기계는 수동 작업뿐만 아니라 회계, 마케팅, 인사 관리와 같은 인지 작업도 수행할 수 있게 될 것입니다.29

AI가 일상생활에 빠르게 발전하고 접목되면서 이미 많은 산업과 분야가 변모했으며, 이러한 추세는 앞으로도 계속될 것으로 예상됩니다. AI는 모든 분야의 주체들에게 의사 결정 방식과 일상 업무 수행 방식을 재고하도록 요구하고 있습니다. 적어도 현재로서는 AI는 우리 곁에 있을 것으로 보이며, 이 획기적인 기술의 잠재력은 이를 활용하려는 모든 사람에게 열려 있습니다.

26 Than Thi Nguyen 외 (2020년 7월 28일). arXiv. 딥페이크 생성 및 탐지를 위한 딥러닝: 조사. https://arxiv에서 확인 가능. org/abs/1909.11573

27 마이클 헨라인, 안드레아스 카플란. (2019). 인공지능의 간략한 역사: 인공지능의 과거, 현재, 그리고 미래에 관하여 gence. 캘리포니아 관리 검토. 61(4): 5–14.

28 Xiaomin Mou. (2019년 9월). 인공지능: 투자 동향 및 특정 산업 활용. 국제금융공사(IFC).

29 제임스 마니카, 수잔 룬드, 마이클 추이, 자크 부긴, 조너선 워첼, 파룰 바트라, 라이언 코, 사우라브 상비. (2017년 11월 28일). 사라진 일자리, 새로 생긴 일자리: 미래의 노동이 일자리, 기술, 임금에 미치는 영향. 맥킨지앤컴퍼니. https:// 에서 발췌 www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and- 임금

III. 알고리즘과 테러리즘의 위협

테러리즘의 전술은 집단과 개인마다 다르지만, 테러 조직은 어느 정도 위험 회피적인 전략을 구사하며, 총기나 폭탄과 같은 검증된 무기의 효과를 선호하는 경향이 있다고 할 수 있습니다.30, 31 그럼에도 불구하고 테러리즘은 정체된 위협이 아닙니다. 테러 조직과 개인은 수십 년에 걸쳐 매우 잘 적응하고 상당히 진화해 왔습니다.

예를 들어, 이들은 조직 구조의 혁신, 즉 분산화, 프랜차이즈화, 그리고 세계화의 잠재력을 보여주었습니다. 또한 전술 측면에서도 비정규 게릴라전에서 무차별 공격으로 전환하며 상당한 발전을 이루었습니다.32, 33

기술 측면에서 이러한 혁신 추세는 특히 인터넷과 소셜 미디어에서 두드러지는데, 이는 테러리스트들에게 매우 귀중한 것으로 입증되었습니다. 인터넷과 소셜 미디어, 그리고 더 나아가 온라인 게임 플랫폼과 같은 다른 생태계는 테러 집단이 급진화하고, 고무하고, 폭력을 조장하고, 공격에 대한 책임을 주장하고, 신병을 모집하고, 자금을 조달하고 이동시키고, 무기를 구매하고 이전하고, 구성원들에게 튜토리얼이나 도구를 제공하는 강력한 도구가 되었습니다.34, 35 예를 들어, 2019년 뉴질랜드 크라이스트처치 총기 난사 사건은 공격자가 페이스북을 통해 생중계했습니다. 영상은 몇 분 후 삭제되었지만, 공격은 전 세계에 방송되어 피해자들에게 미치는 영향과 후유증을 증폭시켰습니다.36

이러한 현상의 확산은 유로폴(Europol)의 '신고 행동의 날(Referral Action Day)'과 같은 활동에서 확인할 수 있습니다. 2020년 '신고 행동의 날'의 일환으로 유로폴과 17개국은 단 하루 만에 180개 플랫폼과 웹사이트에서 테러 관련 콘텐츠로 연결되는 URL 1,906개를 파악하고 삭제 여부를 평가했습니다.37 페이스북은 지난 2년 동안 이라크 레반트 이슬람 국가(ISIL)와 알카에다 등의 단체와 관련된 콘텐츠 2,600만 개 이상을 삭제했으며, 2020년 첫 3개월 동안 "조직적 증오"와 관련된 콘텐츠 약 470만 개를 삭제했습니다. 이 중 2019년 4분기에 비해 300만 개 이상 증가한 수치입니다.38, 39

30 유엔 마약범죄사무소 30. 테러리즘과 재래식 무기. https://www.unodc.org/images/odccp/ter- 에서 확인 가능
로리즘_무기_전통적.html

31 브루스 호프만. (1994). 기술적 스펙트럼 전반의 테러리즘 대응. 랜드 연구소.

32 브루스 호프만. (2017). 『테러리즘 내부』, 3판. 컬럼비아대학교 출판부.

33 Karlheinz Steinmüller. (2017). 2040년의 세계. 새로운 종류의 테러리즘을 위한 기본 조건. TJ Gordon 외 (편). 잠재적 테러리스트 식별 및 적대 세력 계획: 신기술과 새로운 대테러 전략.

34 유엔 안전보장이사회 대테러위원회 및 ICT4Peace. (2016년 12월). 테러 목적의 인터넷 및 ICT 사용에 대응하는 민간 부문의 참여: 대화 강화 및 신뢰 구축. 유엔. https://
ict4peace.org/wp-content/uploads/2016/12/Private-Sector-Engagement-in-Responding-to-the-Use-of-the-Internet-and-ICT-for-Terror-ist-Purposes-1.pdf

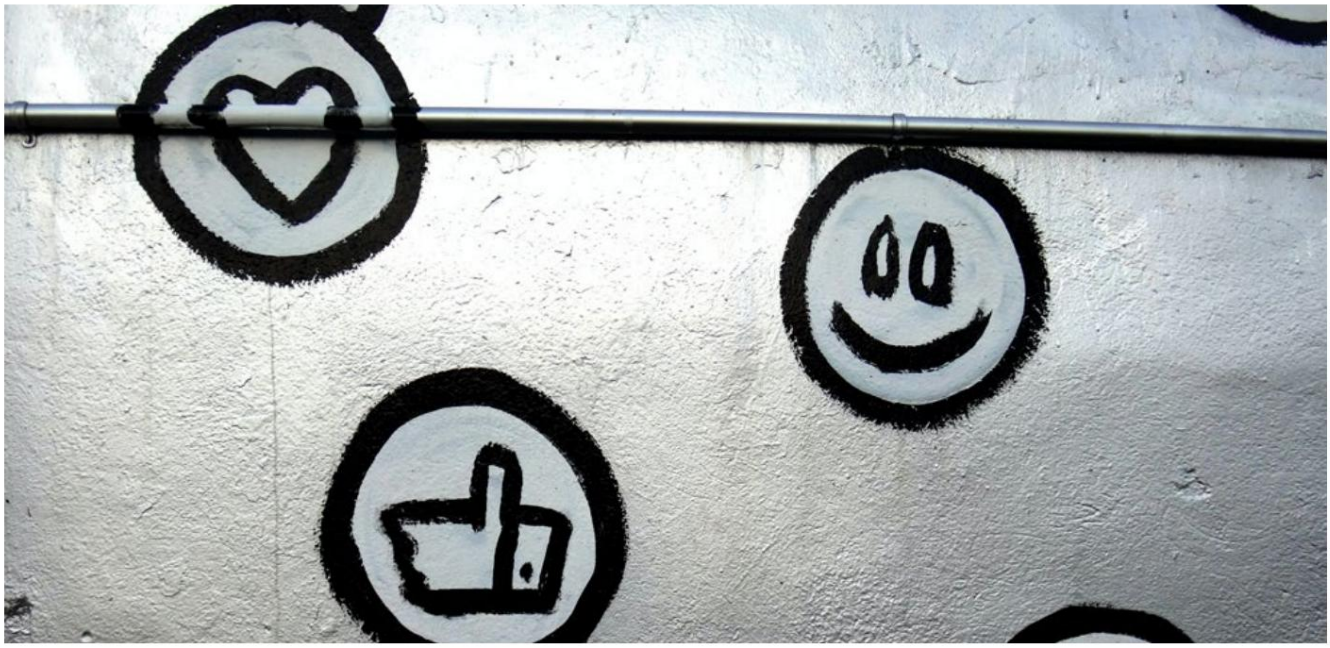
35 유엔 마약범죄사무소(UNODC) 및 유엔 대테러 이행 태스크포스(2012년 9월).
테러 목적의 인터넷. 유엔. 접속 가능
https://www.unodc.org/documents/terrorism/Publications/Use_of_Internet_for_Terrorist_Purposes/ebook_use_of_the_Internet_for_
테러리스트의 목적.pdf

36 테러 방지 기술. (2019년 3월 26일). 분석: 뉴질랜드 테러와 테러리스트들의 인터넷 사용. 테러 방지 기술.
https://www.techagainstterrorism.org/2019/03/26/analysis-new-zealand-attack-and-the-terrorist-use-of-the-Internet/ 에서 접근 가능

37 유로폴(2020년 7월 3일). 테러리스트 "사용법" 가이드 - 최신 유로폴 회부 조치의 날 주요 내용 [보도자료]. 유로폴. https://www.europol.europa.eu/newsroom/news/terrorist-%E2%80%98how-to%E2%80%99-guides-focus-of-latest-europol-referral-action-day 에서 확인 가능

38 페이스북. (2019년 9월 17일). 증오와 극단주의에 맞서 싸우다. 페이스북. https://about.fb.com/news/2019/09/combat- 에서 확인 가능
증오와 극단주의/

39 페이스북. (2020년 5월 12일). 증오 및 위험한 단체 퇴치 관련 최신 소식. https://about.fb.com/에서 확인 가능
뉴스/2020/05/혐오와 위험한 조직에 맞서 싸우다/



Unsplash의 George Pagan III가 찍은 사진

인터넷과 소셜 미디어의 오용은 테러 집단이 특히 어려움에 적응하는 능력이 매우 뚜렷하게 드러나는 영역이기도 합니다. 소셜 미디어 플랫폼과 법 집행 기관이 온라인에서 테러 관련 콘텐츠를 삭제하려는 노력에 따라, 테러리스트들이 인터넷과 소셜 미디어를 사용하는 방식에 암호화된 통신부터 다른 혁신적인 방법까지 여러 가지 발전이 있었습니다. 예를 들어, 탐지를 피하기 위해 최근 페이스북에 업로드된 테러 관련 콘텐츠가 담긴 영상에는 실제 49분 분량의 선전 영상이 시작되기 전에 France 24 뉴스 채널을 30초 동안 소개하는 내용이 포함되었습니다.40 또한 주목할 만한 것은 2020년 5월, 이전의 레반트 인민을 위한 알누스라 전선(Al-Nusrah Front for the People of the Levant)이었던 하야트 타흐리르 알샴(Haya'at Tahrir al-Sham)이 시리아의 회원과 무장 단체에 Telegram, Facebook Messenger, Viber와 같은 플랫폼 사용을 중단하고 대신 Conversations, Riot, Signal, Wire와 같은 다른 암호화된 애플리케이션을 사용하라고 독려했다는 것입니다.41 실제로 이러한 플랫폼에서 종단 간 암호화(E2EE)를 도입하면서 정책 입안자와 테러 대응 담당자 사이에 테러리스트가 "은신"하고 안전하게 통신하여 탐지를 피할 수 있는 가능성에 대한 상당한 우려가 제기되었습니다. 수많은 조사와 대테러 작전을 통해 ISIL과 알카에다 연계 세력 모두 암호화를 사용했음이 드러났습니다.42 기술이 테러리즘의 상당한 진화를 유일하게 촉발한 것은 아니지만, 중요한 역할을 해왔습니다. 최근 역사는 테러리스트와 폭력적인 극단주의 집단이 다양한 기술을 점점 더 정교하게 사용하는 사례로 가득합니다.43 여러모로 이는 놀라운 일이 아닙니다. 모든 개인과 마찬가지로 테러리스트도 결국 "시대의 아이들"입니다.44 집단으로서, 그들은 일반 개인과 마찬가지로 자신들에게 이용 가능한 도구와 기구에 의존하고 활용합니다. 실제로 모든 기술, 아날로그든 디지털이든, 악의적인 개인이 범죄를 저지르는 데 악용할 수 있는 지경까지 발전하고 있습니다.45

40 고든 코레라. (2020년 7월 13일). ISIS, "페이스북에서 여전히 탐지 회피 중"이라고 보도. BBC. https://www.bbc.com/news/ 에서 확인 가능 기술-53389657

41 유엔 안전보장이사회. (2020년 7월 23일). ISIL(다에시), 알카에다 및 관련 개인 및 단체에 관한 결의안 2368호(2017)에 따라 제출된 분석 지원 및 제재 감시팀의 제26차 보고서. 유엔. https://www.securitycouncilreport.org/atf/cf/%7B65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7D/s_2020_717.pdf, 99항에서 열람 가능.

42 로버트 그레이엄. (2016년 6월). 『테러리스트의 암호화 활용 방식』. 웨스트포인트 센티넬 테러 대응 센터. https://ctc.usma.edu/how-terrorists-use-encryption/ 에서 확인 가능

43 이 섹션에서 언급된 기술은 도구, 기계, 기법, 공예, 시스템, 조직 방법 등을 포함하는 보다 광범위한 용어 이해의 관점에서 읽어야 하며, 디지털 기술에만 국한되지 않아야 합니다.

44 렌스케 반 데르 비어. (2019). 기술 시대의 테러리즘, 전략 모니터 2019-2020. 헤이그 전략연구센터와 클링엔달 연구소. https://www.clingendael.org/pub/2019/strategic-monitor-2019-2020/terror-ism-in-the-age-of-technology/ 에서 확인 가능

45 쿨지트 카우르. (2007년 4월 1일). 첨단기술 테러리즘: 세계 안보에 대한 위협. India Quarterly, 63(2), 81-95. https:// 에서 접속 가능. www.jstor.org/stable/26295959

대체로 기술과 테러리즘의 상호작용은 세 가지 주요 방식으로 드러납니다. 첫째, 테러리스트들은 공격을 수행하기 위해 기술을 무기로 사용합니다. 실제로 테러 공격의 성격은 기술 발전과 함께 시간이 지남에 따라 크게 변화했습니다.46 테러리스트의 무기고는 칼과 총기 사용에서 항공기 납치 및 기타 차량 기반 공격에 이르기까지 크게 확장되었으며, 일부 집단은 화학, 생물학 또는 방사능 물질을 획득하고 사용하려는 의도를 보이기도 했습니다. 자동 소총의 도입은 기술 발전에 대한 가장 중요한 적응 중 하나라고 할 수 있습니다. 저렴한 비용과 치명성 덕분에 자동 소총은 전 세계 여러 지역의 테러리스트 집단이 선호하는 무기가 되었습니다.47 둘째, 운송 및 물류 분야의 기술 발전은 테러리스트와 범죄 집단의 전반적인 역량을 변화시켜, 그들의 작전 속도, 범위, 규모를 증가시키고 지역적인 위협이 아닌 세계적인 위협으로 만들었습니다.48

마지막으로, 정보통신 기술의 발전으로 테러 단체와 개인들은 더욱 빠르고 은밀하게 더 먼 거리에서 소통하고, 바이러스성 영상과 정보를 유포하여 더 빠르고 대규모로 테러를 조장할 수 있게 되었습니다. 이를 통해 테러리스트들은 공격의 효율성과 효과를 높이고 잠재적인 테러 용의자들에게 접근할 수 있게 되었습니다.49, 50, 51 모바일 통신 기기, 인터넷, 그리고 최근에는 소셜 미디어와 다크 웹이 이러한 사례의 대표적인 예입니다.

Unsplash의 Niclas Lundin이 찍은 사진

46 Herbert K. Tillema. (2002). 테러리즘과 기술에 대한 간략한 이론. TK Ghosh (편), 『테러리즘의 과학과 기술』(Science and Technology of Terrorism and 테러 대응.

47 TX Hammes. (2020년 9월 4일). 다이너마이트에서 드론까지, 테러와 기술. 암초 위의 전쟁. https://warontherocks에서 확인 가능. com/2020/09/테러와 기술-다이너마이트에서 드론까지/

48 허버트 K. 틸레마. (2002). 테러리즘과 기술에 대한 간략한 이론. TK 고쉬 (편), 『테러리즘의 과학과 기술』(Science and Technology of Terrorism and 테러 대응.

49 위와 같음.

50 Karlheinz Steinmüller. (2017) 2040년의 세계. 새로운 종류의 테러리즘을 위한 기본 조건. TJ Gordon 외(편), Identifica-
잠재적 테러리스트와 적대 세력의 계획: 새로운 기술과 새로운 테러 대응 전략.

51 Truls Hallberg Tønnessen. (2017). 이슬람 국가와 기술 - 문헌 검토. 테러리즘에 대한 관점, 11(6), 101-111. Access-
https://www.jstor.org/stable/26295959 에서 확인 가능

최근 테러리스트들이 기술을 사용한 사례로는 다양한 첨단 기기가 있습니다. 예를 들어, 2008년 뭄바이 테러 사건의 가해자들은 임무를 계획하고, 조정하고, 수행하는 데 GPS(위성항법시스템), 휴대전화, 인터넷을 활용했습니다.52 오늘날의 관점에서 보면 이러한 시도는 더 이상 획기적으로 보이지 않을 수 있지만, 당시에는 최신 기술을 혁신적으로 활용한 사례였습니다.

최근에는 "비트코인"과 같은 블록체인 기반 가상 자산, 모바일 뱅킹, 크라우드 펀딩이 테러리스트들의 자금 조달이나 이동 목적으로 사용되고 있으며, 다크 웹은 재료, 무기, 위조 문서의 시장 역할을 하고 있습니다 .53

그러나 증거에 따르면 단독 테러범들은 기술을 사용할 때 통신, 무기, 운송 목적으로 쉽게 구할 수 있는 기술을 선호하는 경향이 있습니다.55 DIY 상점에서 구할 수 있는, 기술 수준이 낮은 장비를 선호하는 것으로 보입니다. 이러한 "저기술 테러"의 형태에서 테러범, 특히 단독 테러범들은 부엌칼, 자동차, 트럭과 같은 일상적인 도구와 차량을 무기로 개조하는 방법을 모색합니다.56

그럼에도 불구하고, 한때 특정 커뮤니티에만 국한되었던 최첨단 기술이 날이 갈수록 일반 대중에게도 접근 가능해지고 있습니다.57 그 대표적인 예로, 10년 전만 해도 테러 집단이 폭발물을 탑재한 드론 함대나 "군단"을 조종하는 것은 비현실적인 것으로 여겨졌지만, 오늘날에는 그러한 시나리오가 실현 가능한 위협이 되고 있습니다.58

기술적으로 가능한 것의 확장은 법 집행 기관, 테러 방지 부대 및 기타 보안 부대가 악의적인 목적을 위해 저렴하고 상업적인 기술을 사용하여 새로운 예상치 못한 방법과 수단을 발견한 혁신적인 테러 집단과 개인에 의해 "예상치 못한" 공격을 받을 수 있다는 결과를 초래할 수 있습니다.59

이를 고려하고 컴퓨터 비전, NLP 등 AI 분야의 최근 동향, 개발 및 잠재력을 고려할 때, 문제는 AI가 테러리즘 도구 상자의 또 다른 도구가 될지 여부, 아니면 더 나아가 언제 될 것인가입니다.

52 제레미 칸. (2008년 12월 8일). 뭄바이 테러리스트들은 새로운 기술을 이용해 공격을 감행했다. 뉴욕 타임스. https://www.nytimes.com/2008/12/09/world/asia/09mumbai.html

53 금융활동기구(FATF). (2015). 새롭게 부상하는 테러 자금 조달 위험. www.fatf-gafi.org/publications/method-sandtrends/documents/emerging-terrorist-financing-risks.html 에서 확인 가능

54 아비르 엘바라위, 라우라 알레산드레티, 레오니드 루스나츠, 다니엘 골드스미스, 알렉산더 테이텔보임, 안드레아 바론첼리. (2020년 11월). 다크웹 시장의 집단 역학. Scientific Reports, 10, 18827호. https://doi.org/10.1038/s41598-020-74416-y 에서 접속 가능

55 Herbert K. Tillema. (2002). 테러리즘과 기술에 대한 간략한 이론. TK Ghosh(편), 『테러리즘의 과학과 기술』(Science and Technology of Terrorism and 테러 대응.

56 Truls Hallberg Tønnessen. (2017). 이슬람 국가와 기술 - 문헌 검토. 테러리즘에 대한 관점, 11(6), 101-111. 테러-ism 연구 이니셔티브. https://www.jstor.org/stable/26295959 에서 접속 가능

57 Karlheinz Steinmüller. (2017). 2040년의 세계. 새로운 종류의 테러리즘을 위한 기본 조건. TJ Gordon 외 (편), 잠재적 테러리스트 식별 및 적대 세력 계획: 신기술과 새로운 대테러 전략.

58 Truls Hallberg Tønnessen. (2017). 이슬람 국가와 기술 - 문헌 검토. 테러리즘에 대한 관점, 11(6), 101-111. 테러-ism 연구 이니셔티브. https://www.jstor.org/stable/26295959 에서 접속 가능

59 TX Hammes. (2020년 9월 4일). 다이너마이트에서 드론까지, 테러와 기술. 암초 위의 전쟁. https://warontherocks에서 확인 가능.com/2020/09/테러와 기술-다이너마이트에서 드론까지/

IV. AI 위협 유형 분류

AI는 개인, 조직, 그리고 국가에 다양한 맥락에서 수많은 새로운 과제를 제기할 수 있습니다. 이러한 과제는 설계부터 배포까지 AI 수명 주기의 여러 단계에서 발생하며 , 의도된 행동과 의도하지 않은 행동 모두에서 비롯될 수 있습니다.

합법적 행위자가 AI를 사용하는 것과 관련된 가장 큰 우려는 이 기술이 인권을 침해할 수 있는 매우 현실적이고 심각한 잠재력이 있다는 것입니다. AI 기술이 적절하게 사용되지 않을 경우, 예를 들어 사생활 보호권, 평등권(성평등 포함), 그리고 차별 금지권 등이 위협받을 수 있습니다. 이러한 권리 침해는 정당화될 수 없거나 불균형적인 AI 사용으로 인해 발생할 수도 있고, 의도치 않게 발생할 수도 있습니다. 예를 들어, 무의식적으로 편향된 데이터를 사용하여 머신러닝 알고리즘을 훈련시켜 개인, 집단 또는 공동체를 금지된 근거로 차별하는 불공정한 결정을 내리는 경우가 있습니다.61

62

"AI의 악의적 사용"이라는 용어는 일반적으로 의도적으로 해로운 결과를 초래하는 행위에만 적용됩니다.63 2018년, 옥스퍼드 대학교 인류 미래 연구소, 케임브리지 대학교 실존적 위험 연구 센터, OpenAI, 전자 프런티어 재단, 신미국안보센터 등 다양한 분야와 기관의 저명한 저자들이 국가, 범죄자, 테러리스트의 AI 악용 사례를 조사했습니다. "인공지능의 악의적 사용: 예측, 예방 및 완화"라는 제목의 보고서는 향후 10년 동안 이 기술의 악의적 사용이 빠르게 증가할 것으로 예상했습니다. 저자들은 AI의 악의적 사용이 사이버 보안, 물리적 보안, 그리고 정치적 보안 측면에서 위협을 제기한다고 생각했습니다.64

사이버 위협: 사이버 공간의 본질적인 취약성과 사이버 공격으로 인한 위협의 비대칭적 특성으로 인해 사이버 위협은 점점 더 큰 우려를 불러일으키고 있습니다. 테러리즘의 관점에서 볼 때, 피싱, 중간자 공격, 랜섬웨어, DDoS 공격, 그리고 웹사이트 변조 등이 위협으로 간주됩니다. 또한, 테러리스트들이 정보 및 통신 기술, 특히 인터넷과 소셜 미디어를 악용하여 테러 행위를 저지르거나, 선동하거나, 모집하거나, 자금을 조달하거나, 계획하는 것에 대한 우려가 커지고 있습니다. 다음 장에서 자세히 설명하겠지만, 테러리스트들은 AI 시스템을 활용하여 기존 사이버 공격의 위력과 효과를 높이거나, 기밀성을 침해하거나, 무결성 및 가용성을 공격하여 정보 보안을 위협할 수 있습니다.

물리적 위협: 지난 10년 동안 일상생활은 기술을 통해 점점 더 상호 연결되었습니다. 이러한 상호 연결성은 사물 인터넷(IoT)이라는 개념의 등장으로 반영됩니다. 사물 인터넷은 인터넷을 통해 데이터를 전송하는 연결된 디지털 기기와 물리적 사물의 생태계입니다. 이처럼 연결된 세상에서 드론은 배송을 시작했고, 자율주행차는 이미 도로를 누비고 있습니다. 동시에 이러한 기술과 연결된 기기들이 일상생활에 통합됨에 따라 인간과 사회 기반 시설에 대한 새로운 과제가 발생합니다. 스마트 시티나 가정 환경에서 상호 연결성과 점점 더 자율화되는 기기 및 로봇은 공격의 기회와 규모를 확대합니다.

60 패트릭 브래들리. (2020). 위험 관리 표준 및 인공 초지능에서의 악의적 의도의 적극적 관리. AI & Society, 35(2), 319-328. https://doi.org/10.1007/s00146-019-00890-2 에서 접근 가능

61 Nancy G. Leveson, Clark S. Turner. (1993). Therac-25 사고 조사. Computer, 26(7), 18–41. https://doi. 에서 접근 가능. org/10.1109/MC.1993.274940

62 Ben Shneiderman. (2016). 의견: 결함이 있거나 편향되었거나 악의적인 알고리즘의 위험은 독립적인 감독을 필요로 한다. 미국 국립과학원 회보, 113(48), 13538-13540. https://doi.org/10.1073/pnas.1618211113 에서 접근 가능

63 패트릭 브래들리. (2020). 위험 관리 표준 및 인공 초지능에서의 악의적 의도의 적극적 관리. AI & Society, 35(2), 319-328. https://doi.org/10.1007/s00146-019-00890-2 에서 접근 가능

64 마일즈 브런디지, 샤하르 아빈 외 (2018년 2월). 인공지능의 악의적 사용: 예측, 예방 및 완화. Ac- https://maliciousaireport.com/ 에서 전송 가능

정치적 위협: 정보통신 기술의 발전과 소셜 미디어의 전 세계적인 영향력 확대로 인해 개인이 소통하고 뉴스 매체를 찾는 방식, 시기, 그리고 이유는 전례 없는 변화를 겪고 있습니다. 이러한 변화는 전 세계적으로 나타나고 있으며, 선거 결과에 영향을 미치고, 대중 시위를 촉진하며, 사람들이 기본권을 행사할 수 있도록 힘을 실어주었습니다. 동시에, 소셜 미디어의 영향력 확대는 사람들을 허위 정보와 허위 정보를 통한 조작에 취약하게 만들 수 있으며, 공공 및 민간 기관의 프로파일링 및 감시 활동 역량을 강화했습니다. 딥페이크 확산과 같은 AI 기술이 이러한 상황에 통합되면 이러한 위협의 본질이 크게 강화될 것입니다.

2018년 보고서 저자들이 지적했듯이, 이러한 범주들은 반드시 상호 배타적인 것은 아닙니다. 예를 들어, AI 기반 해킹은 사이버-물리적 시스템을 표적으로 삼아 물리적 피해를 입힐 수 있으며, 물리적 또는 디지털 공격은 정치적 목적을 위해 수행될 수 있습니다. 더욱이, "정치적"이라는 용어는 특히 테러리즘의 맥락에서 복잡한 범주화입니다. 테러리즘의 정치적 동기는 사회적, 이념적, 종교적, 경제적 요인과 더불어 테러리즘 개념에 대한 일반적인 이해와 매우 밀접하게 연관되어 있는 경우가 많습니다.

이와 관련하여 이 보고서의 목적상 테러 목적으로 AI를 악의적으로 사용하는 경우 사이버 위협과 물리적 위협이라는 두 가지 주요 위협 유형을 고려하고, 자금 조달 방법, 선전 및 허위 정보 전략, 기타 작전 전술을 포함하여 테러 집단 및 개인의 행동과 관련된 기타 관련 활동에 대한 논의도 추가합니다.

V. 사실인가 공상과학인가?

이 장에서는 AI의 악의적 사용으로 인해 발생하는 몇 가지 위협 유형을 살펴본 후, 그러한 위협에 대한 신뢰할 만한 근거가 있는지, 아니면 테러 목적으로 AI를 악용하는 것이 공상과학에 불과한지 알아보겠습니다.

처음부터 테러 조직이 AI를 실제로 사용했다는 명확한 증거는 현재까지 확인되지 않았다는 점을 명확히 하는 것이 중요합니다. 실제로 ISIL/알카에다 감시팀은 최근 보고서에서 "특히 금융, 무기, 소셜 미디어 분야에서 테러리스트의 기술 남용에 대한 회원국의 지속적인 우려에도 불구하고, ISIL이나 알카에다 모두 2020년 말 이 측면에서 상당한 진전을 이루지 못한 것으로 평가된다"고 지적했습니다.65

그러나 여기에는 중요한 단서가 있습니다. 이미 살펴본 바와 같이 AI는 이미 일상생활의 일부가 되었으며, 많은 사람들이 자신도 모르는 사이에 AI를 사용하고 있습니다. 예를 들어, 자연어 처리(NLP)는 Apple의 Siri와 Amazon의 Alexa와 같은 스마트 비서의 기반이며, 문자 메시지, 이메일, Word 문서의 오타를 수정하는 데 사용됩니다. 얼굴 인식은 스마트폰 잠금 해제에 사용되고, 사물 인식은 이미지 분류 및 Google 검색 결과 개선에 도움이 됩니다. 이와 관련하여, 위의 진술은 테러 집단과 개인이 AI를 간접적으로, 예를 들어 위에서 설명한 바와 같이 수동적으로 또는 무의식적으로 사용했을 가능성을 배제하는 것이 아닙니다. 오히려, AI가 공격을 구체적으로 개선하거나 증폭시키는 데 직접적으로 사용되지 않았음을 암시하는 것입니다.

65 분석 지원 및 제재 모니터링 팀은 ISIL(Da'esh), 알카에다 및 관련 개인 및 단체, 제27차 보고서, S/2021/68(2021년 2월 3일).

Unsplash의 Nathan Dumlao 사진

테러리즘에서 AI를 직접 사용한다는 증거가 부족하다고 해서 테러리스트들이 이 기술에 무관심하거나 무관심하다는 것을 의미하는 것으로 해석해서는 안 됩니다. 테러리스트들이 AI를 사용하려는 관심이나 의도를 보여주는 구체적인 증거는 아직 발견되지 않았지만, 이러한 집단과 개인들이 많은 사람들이 혁명적인 잠재력으로 극찬해 온 이 기술을 알고 있다고 가정하는 것이 현명합니다. 예를 들어, 2016년 시리아 ISIL이 제작한 것으로 보이는 영상에서 해당 집단이 원격 조종 방식의 초보적인 자율주행차를 실험하는 모습이 포착되었다는 점은 주목할 만합니다.66 문제 의 차량에는 마네킹이 차량 내부에 배치되어 있어 관찰자들이 사람이 운전하는 것처럼 착각하게 만들었습니다. ISIL은 또한 보안 시스템을 속여 차량 내부에 누군가가 실제로 있다고 믿게 하기 위해 사람의 열 신호를 복제하려는 계획이 있었던 것으로 믿어진다 .67 얼마 지나지 않아 F-Secure의 최고 연구 책임자는 ISIL이 자살 폭탄 테러범 대신 사용할 자율 주행 자동차를 개발하고 있다는 증거가 있다고 밝혔고 68 2018년 영국 검찰은 ISIL 지지자 두 명이 테러 공격에 무인 자동차를 사용할 계획이라고 밝혔다.69, 70

최근 2020년 3월, ISIL 지지자가 ISIL이 테러 콘텐츠를 유포하고 온라인 협업 및 조율을 촉진하는 데 사용하는 분산형 소셜 미디어 플랫폼인 Rocket.chat에 안면 인식 소프트웨어의 활용 방법을 설명하는 영상을 유포했습니다.71 문제의 영상은 안면 인식을 통해 얼굴 가리개를 사용하거나 디지털 방식으로 얼굴을 흐리게 처리하여 신원을 숨기려 했더라도 얼굴 특징을 기반으로 개인을 식별할 수 있다고 주장했습니다. 또한 이 영상은 이러한 기능이 당국이 테러 음모를 저지하고 범인을 체포하는 데 확실히 도움이 될 것이라고 주장했습니다. 이 기술의 현재 성능은

66 로난 글론. (2016년 1월 13일). ISIS가 적외선 센서를 무력화하는 치명적인 원격 조종 차량 폭탄을 시험하고 있다. 디지털 트렌드. 다음에서 확인 가능 https://www.digitaltrends.com/cars/isis-remote-contrlled-car-bomb-news-demonstration/

67 스티븐 에델스타인. (2016년 3월 16일). 보안 분석가에 따르면 ISIS는 무인 자동차 폭탄 개발에 박차를 가하고 있다. 디지털 트렌드. 접근성 https://www.digitaltrends.com/cars/isis-autonomous-car-bombs/ 에서

68 피트 비글로우. (2016년 3월 15일). ISIS는 폭탄 운반에 자율주행차를 사용할 수 있다. 오토블로그. https://www.autoblog.com/2016/03/15/isis-terrorists-bomb-self-driving-cars-sxsw/?guccounter=2

69 Telegraph 기자. (2018년 9월 4일). "ISIS 지지자들"이 자신들의 생명을 구하기 위해 무인 차량 폭탄을 사용하여 테러 공격을 계획한 혐의를 받고 있습니다. 텔레그래프. https://www.telegraph.co.uk/news/2018/09/04/isil-supporters-accused-plotting-terror-attack-us-ing-driverless/ 에서 확인 가능

켈리 린(2016년 5월 2일). NATO 전문가, ISIS가 자율주행차 무기화 작업 중이라고 경고. MotorTrend. https://www.motortrend.com/news/isis-working-on-weaponizing-self-driving-cars-nato-expert-warns/

71 Memri – Cyber & Jihad Lab. (2020년 3월 31일). ISIS 지지자가 Rocket.Chat에 얼굴 인식 소프트웨어의 성능을 보여주는 영상을 공유했습니다. Memri. https://www.memri.org/cjlab/isis-supporter-shares-video-rocketchat-demonstrating-abilities-fa- cial-recognition-software#_ednref1 에서 확인 가능

영상에서 과장되었을 수도 있지만, 그 존재를 인정한 사실과 영상이 여러 다른 채널에서 빠르게 공유된 사실은 테러 집단이 AI의 잠재력을 알고 있으며 적어도 표면적으로는 그 추세와 발전을 주시하고 있다는 것을 확인시켜 줍니다.

테러 집단이 AI를 직접 사용하는 것으로 확인된 적은 없지만, 이들 집단이 AI 관련 기술을 악용하고 있다는 상당한 증거가 있습니다. 특히 "드론"이라고도 하는 무인 항공 시스템의 악용 사례에서 이러한 경향이 두드러집니다. 본 보고서의 목적상 드론은 수동으로 조작하더라도 다양한 수준의 자율성을 가질 수 있다는 점에서 AI 관련 기술로 간주됩니다. 예를 들어, 드론은 이미 GNSS(Global Navigation Satellite System) 지원 비행 안정화 및 "감지 및 회피" 기능을 갖추고 있으며, AI를 활용하여 더욱 높은 수준의 자율성을 제공할 수 있습니다.72

테러 집단뿐 아니라 다른 비국가 행위자들도 이러한 기술을 사용하는 것은 새로운 현상이 아닙니다. 이러한 집단들은 수년 동안 드론, 즉 원격 조종 항공기를 실험해 온 것으로 보이며, 그 기원은 1995년 도쿄 지하철 사린 테러를 일으킨 일본 사이비 종교 단체인 옴진리교의 미사용 설계도까지 거슬러 올라갑니다.73

이러한 집단의 드론 사용의 성격은 다양하며 실제 공격 및 시도 공격, 교란, 감시 및 선전을 포함합니다.74 또한 드론이 정보, 감시 및 정찰 임무를 수행하는 데 사용될 수 있다는 의견도 제시되었습니다. 목표물, 보안 프로토콜 및 행동 패턴을 모니터링하고 간접 사격의 정확도를 높이고 선전 자료에 사용할 영상을 수집하고 법 집행 작업을 방해하고 주요 인프라, 항공 교통 및 경제 자산을 교란, 방해 또는 마비시키고 불법 상품을 국경을 넘나들거나 민감한 지역으로 밀수하고 위협하고 괴롭히고 대규모 집회에서 공황 상태를 조장할 수 있습니다.75

Unsplash의 Shutterbouy Photography가 촬영한 사진

AI 기반 드론에 대한 자세한 내용은 아래 8장을 참조하세요.

73 1990년대 초, 오움 신비교는 사용하지 않는 음모의 일부로 스프레이 부착물이 달린 원격 제어 드론 두 대를 구입했습니다: 에이미 E. 스미스슨. (2000). 3장: 도쿄의 교훈을 재고하다. 에이미 E. 스미스슨 저 레슬리-앤 레비(편), 『운동 실조: 화학 및 생물학 테러 위협과 미국의 대응』. https://www.스팀슨.org/wp-content/files/file-attachments/atxchapter3_1.pdf

74 유엔 대테러위원회 집행이사회(2019년 5월). 테러리스트의 무인 항공기 시스템 사용으로 인한 잠재적 위험에 대처하기 위한 더 많은 노력 필요. CTED 동향 경보. https://www.un.org/sc/ctc/wp-content/up- loads/2019/05/CTED-UAS-Trends-Alert-Final_17_May_2019.pdf 에서 확인 가능.

제75차 세계대테러포럼(GCTF) 무인항공시스템 위협 대응 이니셔티브(2019년 9월). 테러리스트의 무인항공시스템 사용 대응 우수 사례에 관한 베를린 각서. 제10차 GCTF 장관급 총회. https://www.thegctf.org/Portals/1/Documents/Framework%20Documents/2019/Berlin%20Memorandum%20EN.pd- f?ver=2020-01-13-143548-187 에서 확인 가능

특히 ISIL이 2016년경부터 드론을 사용해 왔다는 증거가 있습니다. ISIL은 드론 개발 및 사용을 담당하는 "무자헤딘의 무인 항공기" 부대를 구성한 것으로 알려졌습니다.76 ISIL 구성원들은 이 기술을 처음 사용한 것으로 여겨지며, 이라크 북부에서 폭발물을 탑재한 드론을 공격에 투입하여 쿠르드 페슈메르가 전투원 2명을 사살하고 프랑스 특수작전 소속 군인 2명을 부상시켰습니다.77 2017년 ISIL은 드론 이니셔티브의 성공을 자랑하며 드론 공격으로 1주일 만에 군인 39명이 사망하거나 부상당했다고 주장했습니다.78 이 테러 집단은 또한 온라인에서 지지자들에게 드론 사용에 대한 지침을 배포하고 드론을 이용한 공격을 촉구하는 선전 자료를 공개했습니다.79

다른 비국가 행위자들도 드론을 사용한 것으로 알려져 있습니다.80 두 가지 주목할 만한 드론 공격 사례는 언급할 가치가 있습니다. 2018년 8월 베네수엘라 니콜라스 마두로 대통령 암살 시도와 2019년 9월 사우디아라비아 아브카이크와 쿠라이스에 있는 사우디 아람코 석유 처리 시설 공격입니다.81 두 사건 모두 폭발물을 탑재한 드론이 연루되었습니다. 특히 후자는 최대 25대의 드론이 동시에 작동했다는 점에서 주목할 만합니다.

이 기술을 악의적인 목적으로 사용하는 것에 대한 관심이 높아지는 주요 요인은 드론의 상업적 가용성, 경제성, 그리고 편의성과 더불어 드론 사용에 대한 대응의 어려움입니다. 테러 공격에 드론을 사용할 경우 발생할 수 있는 극적인 영향 또한 테러 단체와 개인에게 드론이 갖는 매력을 파악하기 위해 고려해야 할 또 다른 요소입니다.

따라서 테러 집단과 기타 비국가 행위자가 드론을 사용하는 것에 대한 우려가 높아졌으며, 특히 떼를 지어 사용할 가능성이 있습니다.82 유엔 안전보장이사회는 회원국들이 테러리스트가 무기를 획득하지 못하도록 더 큰 집단적 노력을 기울일 것을 촉구하면서 결의안 2370(2017)을 통해 ISIL, 알카에다, 그 산하 단체 및 관련 집단, 불법 무장 집단 및 범죄자에게 드론이 흘러가는 것을 강력히 비난했습니다.83 글로벌 대테러 포럼(GCTF)도 드론이 점점 더 우려되는 영역임을 인정했으며, 이와 관련하여 2019년에 무인 항공 시스템의 테러 사용 대응을 위한 모범 사례에 대한 베를린 각서를 발표했습니다. 이 각서는 테러리스트의 드론 사용에 대응하기 위한 정책, 관행, 지침, 규정, 프로그램 및 접근 방식의 식별, 개발 및 개선에 대한 지침을 제공합니다.84

그럼에도 불구하고, 테러 조직이 드론 기술을 사용하는 사례는 여전히 드물고, 그 특성상 정교하지 못하며 인간의 통제에 크게 의존하고 있습니다. 드론이 자율성을 높이기 위해 AI를 활용할 수는 있지만, 현재로서는 테러리스트나 기타 비국가 행위자가 AI 기반 드론을 사용하거나 사용하려고 시도했다는 증거는 제한적입니다.

76 조비 워릭, 제이슨 알다그. (2017년 2월 21일). ISIS의 무기화된 드론 사용, 테러리즘에 대한 우려 증폭. 워싱턴 포스트. https://www.washingtonpost.com/world/national-security/use-of-weaponized-drones-by-isis-spurs-terrorism-fears/2017/02/21/9d-83d51e-f382-11e6-8d72-263470bf0401_story.html 에서 확인 가능

77 토마스 기븐스-네프. (2016년 10월 11일). ISIS가 무장 드론을 사용하여 쿠르드족 전투원 2명을 사살하고 프랑스군에게 부상을 입혔다는 보고서가 발표됨. 워싱턴 포스트. https://www.washingtonpost.com/news/checkpoint/wp/2016/10/11/isis-used-an-armed-drone-to-kill-two-kurdish-fighters-and-wound-french-troops-report-says/ 에서 확인 가능

78 조비 워릭, 제이슨 알다그. (2017년 2월 21일). ISIS의 무기화된 드론 사용, 테러리즘에 대한 우려 증폭. 워싱턴 포스트. https://www.washingtonpost.com/world/national-security/use-of-weaponized-drones-by-isis-spurs-terrorism-fears/2017/02/21/9d-83d51e-f382-11e6-8d72-263470bf0401_story.html 에서 확인 가능

79 스티븐 스탈린스키와 R. 소스노. (2017년 2월 21일). 지하디 단체들의 드론 활용 10년 - 히즈불의 초기 실험에서 - ISIS가 첫 공격 드론을 발사하면서 서방 국가 안보 위기에 하마스, 알카에다, 그리고 하마스에 대한 우려가 커지고 있습니다. Memri. https://www.memri.org/reports/decade-jihadi-organizations-use-drones-%E2%80%93-early-experiments-hizbullah-hamas-and-al-qa-eda#ISIS%20Anchor 에서 확인 가능

80 로버트 J. 벙커. (2015년 8월). 테러 및 반란군 무인 항공기: 용도, 잠재력, 그리고 군사적 함의. 전략적 연구소 및 미 육군 전쟁대학 출판부. https://www.hsdl.org/?view&did=786817 에서 접속 가능

81 제이콥 웨어. (2019년 9월 24일). 테러 집단, 인공지능, 그리고 킬러 드론. 암초 위의 전쟁. https://waronth- 에서 확인 가능 erocks.com/2019/09/테러리스트-그룹-인공지능-및-킬러-드론/

82 렌스케 반 데르 비어. (2019). 기술 시대의 테러리즘, 전략 모니터 2019-2020. 헤이그 전략연구센터 및 클링엔달 연구소. https://www.clingendael.org/pub/2019/strategic-monitor-2019-2020/terror-ism-in-the-age-of-technology/ 에서 확인 가능

83 유엔 안전보장이사회. (2017년 8월 2일). 결의안 2370호(2017)는 8월 2일 제80차 유엔 안전보장이사회 회의에서 채택되었습니다. 2017년 8월. https://undocs.org/S/RES/2370(2017) 에서 접근 가능

84 무인 항공 시스템 위협 대응을 위한 GCTF 이니셔티브(2019년 9월). 테르메스 대응 우수 사례에 관한 베를린 각서 무인 항공 시스템의 활용. 제10차 GCTF 장관급 총회, https://www.thegctf.org/Portals/1/ 에서 새롭게 접근 가능 문서/프레임워크 문서/2019/베를린 메모란덤 EN.pdf?ver=2020-01-13-143548-187

VI. 거울 속의 AI 기반 테러리즘

이전 장에서는 테러 집단과 테러리스트들이 GPS, 휴대전화, 그리고 최근에는 드론과 같은 새롭게 부상하는 기술에 혁신을 일으키고 적응하는 능력을 반복적으로 입증해 왔음을 지적했습니다. 실제로 테러리즘이 진화하는 위협이라는 점을 이해하고 수용하는 것은 국제 사회가 테러리즘을 예방하고 대응할 수 있는 역량을 확보하는 데 필수적입니다. 상상력의 부재는 치명적인 결과를 초래할 수 있습니다.

이를 염두에 두고, 이 장에서는 테러 조직이 AI를 악용할 가능성이 있는 사례들을 제시함으로써, 곧 다가올, 또는 그 너머에 도래할 수 있는 테러리즘과 AI 관련 잠재적 위협을 추론해 볼 것입니다. 이러한 악용 사례들은 AI 분야의 동향과 발전, 테러 조직과 개인의 기존 행동 방식, 그리고 현재 알려진 AI의 범죄적 활용 사례에서 영감을 얻었습니다.85

이 장에 포함된 AI의 잠재적 악의적 사용 목록은 테러 조직이 AI를 사용하는 방법을 철저히 요약한 것이 아니며, 그러한 시나리오가 발생할 가능성을 나타내는 것도 아닙니다.
오히려, 테러 단체와 개인이 AI 기술을 활용하여 어떻게 혁신을 더욱 발전시킬 수 있는지에 대한 생각을 자극하고 논의를 촉진하는 것을 목표로 합니다. 이를 통해 테러 대응을 담당하는 국내외 기관들의 지식을 구축하고 이 문제에 대한 이해를 증진할 수 있습니다. 증거가 부족한 상황에서는 추측을 통해서만 적절한 수준의 대비 태세를 확보할 수 있습니다.

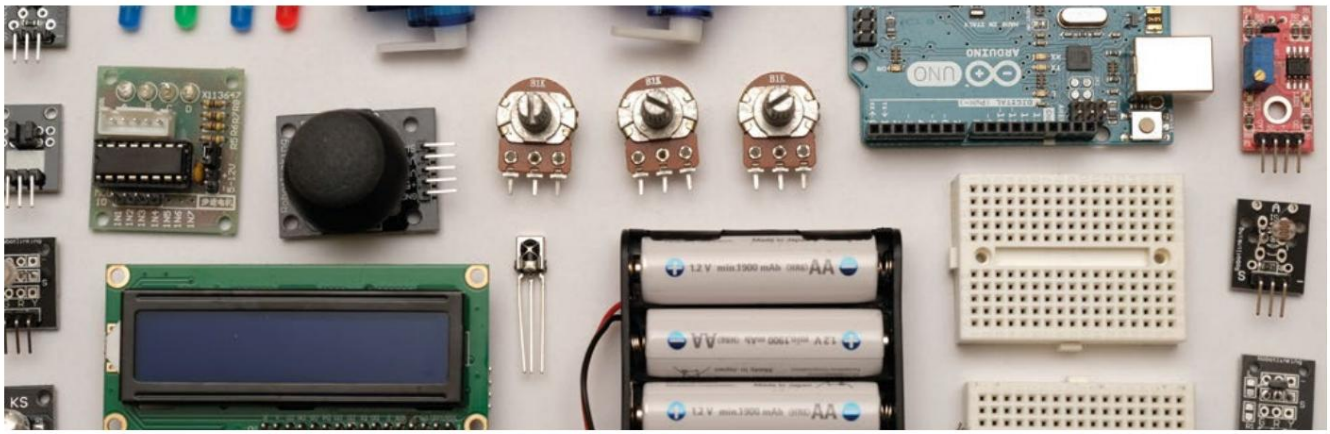


사진: Robin Glauser, Unsplash

이러한 악의적 사용은 그 목적에 따라 사이버 역량 강화, 물리적 공격 지원, 테러 자금 조달, 선전 및 허위 정보 유포, 기타 작전 전술 등 다양한 형태로 분류 및 분류되었습니다. 그러나 아래에 제시된 일부 악의적 사용은 이중적인 목적이나 기능을 가질 수 있습니다. 예를 들어, 랜섬웨어는 사이버 공격의 일부로 사용되면서 동시에 테러 자금 조달을 가능하게 할 수 있습니다. 따라서 이 장에서 사용된 분류는 유연하게 해석되어야 합니다.

마지막으로, 이러한 잠재적 악의적 용도에 대한 검토를 진행하기 전에, AI의 발전이 악의적 행위자의 특정 역량을 강화할 가능성이 높지만, 플랫폼이나 시스템의 보안을 공격으로부터 보호하는 임무를 맡은 국가 당국과 민간 기관은 정체되어 있지 않다는 점을 지적할 필요가 있습니다. 이러한 기관들은 최신 기술 동향, 발전, 그리고 획기적인 발전에 맞춰 끊임없이 발전하고 적응하고 있습니다. 따라서 AI가 기존 위협을 증폭시키거나 테러리즘의 관점에서 새로운 위협을 제시할 수 있지만, 잠재적 위협을 예방하거나 완화할 수 있는 새로운 역량을 향상시키거나 제공할 것이라는 점을 명심하는 것이 중요합니다. 특히 AI가 이미 상당한 역할을 수행하고 있는 사이버 보안 분야에서 더욱 그렇습니다.86

85 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philip Amann, Aglika Klein, Odhran McCarthy 및 Maria Eira. (2020년 11월 19일). 말리-인공지능의 활용과 남용 사례. 트렌드마이크로 리서치. http://unicri.it/sites/default/files/2020-11/AI%20 에서 확인 가능 MLC.pdf

86 카스퍼스키. 사이버 보안에서의 AI와 머신러닝 - 미래를 어떻게 형성할 것인가. 카스퍼스키. https://www.kaspersky.com/ko/ 에서 확인 가능 sky.com/resource-center/definitions/ai-cybersecurity

i. 사이버 역량 강화

에이. 서비스 거부 공격

서비스 거부(DoS) 또는 분산 서비스 거부(DDoS) 공격은 수십 년 동안 가장 인기 있는 사이버 공격 중 하나였습니다.87 이러한 공격의 궁극적인 목적은 여러 연결 요청을 통해 메모리를 완전히 소진시켜 인터넷에 연결된 컴퓨터 시스템을 일시적으로 사용자가 사용할 수 없게 만드는 것입니다.88 DDoS 공격에서 공격자는 "봇넷"이라고 알려진 하나 이상, 종종 수천 대의 컴퓨터를 사용하여 대상 시스템에 요청을 보냅니다.89

ISIL은 2016년 말에서 2017년 초 사이에 사상 최초의 성공적인 DDoS 공격을 감행한 것으로 추정됩니다. 이는 ISIL의 최상위 다크웹 포럼에서 구성원들 간에 이러한 공격 가능성에 대한 논의가 진행된 후였습니다.90 이 공격은 "칼리프 캐논(Caliphate Cannon)"이라는 DDoS 도구를 사용했으며, 주로 군사, 경제, 교육 기반 시설을 표적으로 삼았는데, 이는 이 위협의 심각성을 여실히 보여줍니다. 그 이후 ISIL 해킹 부서는 온라인 서비스를 교란시킨 유사한 공격에 대한 책임을 주장해 왔습니다.

DoS 또는 DDoS 공격이 사이버 범죄자, 테러리스트 및 기타 악의적인 행위자들에게 매력적인 이유 중 하나는 매우 적은 노력으로 실행할 수 있고 성능이 비교적 간단하다는 점입니다.91 이러한 공격을 실행하기 위해 공격자는 특정 취약점을 표적으로 삼을 필요가 없습니다. 대상 시스템이 인터넷에 연결되어 있다는 사실만으로도 일반적으로 충분합니다.92 그러나 머신 러닝은 전통적으로 공격자가 수행하던 프로세스를 자동화함으로써 DoS 또는 DDoS 공격의 용이성과 단순성을 한 단계 더 발전시킬 것으로 예상됩니다. 예를 들어, 머신 러닝 알고리즘을 사용하여 공격 배후에 있는 봇넷을 제어하거나 정교한 네트워크 정찰을 통해 취약한 시스템을 식별할 수 있습니다.

DDoS 공격에 머신러닝을 활용하는 가능성은 이미 악의적인 공격자들에 의해 탐색되고 있습니다. 예를 들어, 2018년, 프리랜서 노동자를 위한 온라인 마켓플레이스인 TaskRabbit은 AI 소프트웨어로 제어되는 봇넷을 사용하는 해커에 의해 수행된 DDoS 공격의 표적이 되었습니다. 이 공격으로 375만 명의 웹사이트 사용자가 영향을 받았고, 심각한 데이터 유출 사고가 발생했습니다.93

흥미롭게도, Link11은 2019년에 DDoS 공격의 거의 절반이 현재 Amazon Web Services, Microsoft Azure 및 Google Cloud와 같은 클라우드 서비스를 사용하여 수행된다고 보고했습니다.94 이러한 서비스에서 제공하는 컴퓨팅 파워를 활용하여 머신 러닝을 사용하여 악성 가상 머신을 만들 수 있으며, 이는 DDoS 공격을 시작하기 위한 봇넷의 일부로 사용됩니다.95

87 DDoS 공격의 역사. (2017년 3월 13일). Radware. https://security.radware.com/ddos-knowledge-center/ddos-chroni- 에서 확인 가능.
cles/ddos-attacks-history/

88 Christoph L. Schuba 외 (1997). 1997 IEEE 보안 심포지엄 논문집에 수록된 TCP 서비스 거부 공격 분석
및 개인정보보호(Cat. No. 97CB36097). https://doi.org/10.1109/SECPRI.1997.601338 에서 확인 가능

89 John Ioannidis 및 Steven M. Bellovin. (2002). 푸시백 구현: DDoS 공격에 대한 라우터 기반 방어. https://www.cs.columbia.edu/~smb/papers/pushback-impl.pdf
에서 확인 가능

90명의 사이버 지하디스트, DDoS 공격에 가담: 위협 평가. (2017년 7월 13일). 플래시 포인트. https://www.flashpoint-intel.com/blog/ 에서 확인 가능
사이버 지하디스트 DDoS/

91 Christoph L. Schuba 외 (1997). TCP 서비스 거부 공격 분석, IEEE 보안 및 개인정보보호 심포지엄 논문집(카탈로그 번호 97CB36097). https://doi.org/10.1109/
SECPRI.1997.601338 에서 확인 가능

92 존 이오아니디스, 스티븐 M. 벨로빈. (2002). 푸시백 구현: DDoS 공격에 대한 라우터 기반 방어.

93 Sam Bocetta. (2020년 3월 10일). AI 사이버 공격이 발생했습니까? InfoQ. https://www.infoq.com/articles/ai-cyber-at- 에서 확인 가능
압정/

94 AI 대 AI: 인공지능과 DDoS 공격. (nd) Verdict-AI. https://verdict-ai.nridigital.com/verdict_ai_winter19/artificial_intelligence_ddos_attack 에서 확인 가능

95 AI 대 AI: 인공지능과 DDoS 공격. (nd) Verdict-AI. https://verdict-ai.nridigital.com/verdict_ai_winter19/ 에서 확인 가능
인공 지능 DDoS 공격

b. 악성 소프트웨어

맬웨어 또는 악성 소프트웨어[강조 추가]는 컴퓨터 시스템이나 네트워크에 침입하여 시스템을 마비시키고 대상을 해치거나 악용하거나 교란하는 광범위한 소프트웨어를 의미합니다. 맬웨어의 예로는 스파이웨어, 랜섬웨어, 바이러스, 웜, 트로이 목마, 애드웨어 등이 있습니다. 맬웨어는 오랫동안 반달리즘, 사기꾼, 협박범 및 기타 범죄자들이 사용해 왔으며, 테러 집단과 개인이 사용하는 명백한 도구이기도 합니다.96 예를 들어, 맬웨어는 악의적인 행위자가 웹사이트, 서버 또는 네트워크에 접근하여 신용카드 정보나 기타 기밀 정보를 얻거나 공공 또는 민간 기관의 사이버 인프라를 손상시키는 데 사용될 수 있습니다.97

AI, 특히 머신 러닝의 발전은 맬웨어와 같은 사이버 보안 위협에 맞서는 데 엄청난 응용 분야를 찾고 있으며, 전문가가 과거 공격으로부터 데이터를 분석하여 이상 징후를 탐지하고 잠재적 위협을 방어할 수 있도록 지원합니다.

그러나 동시에 AI는 악성코드 개발자에 의해 악용될 수도 있습니다. 예를 들어, AI는 공격 프로세스를 자동화하고, 악성코드 공격의 효율성을 높이거나, 심지어 완전히 새로운 형태의 악성코드를 만드는 데 사용될 수 있습니다. 이론적으로는 완전히 새로운 형태의 악성코드를 위한 코드를 작성하는 데 사용될 수도 있습니다. 실제로 사이버 범죄자들은 이미 AI를 사용하여 다형성 악성코드(탐지를 피하기 위해 적응하고 변화하는 일종의 스마트 악성코드)를 만들어 왔습니다.98 또한 AI는 이러한 유형의 악성코드를 강화하여 적응 속도를 높이는 데 사용될 수 있습니다.

AI가 악성코드 배포를 강화하고 자동화하는 데 중요한 역할을 할 수 있다는 점도 주목할 만합니다. 예를 들어 피싱 캠페인은 악성코드를 배포하는 주요 방법 중 하나입니다. 최근 AI를 활용한 "SNAP_R" 실험에서 스피어 피싱 트윗이 분당 6.75개의 속도로 819명의 사용자에게 전송되었으며, 그중 275개가 성공했습니다. 실험에 참여한 인간은 분당 1.075개의 속도로 129명의 사용자에게 트윗을 전송했으며, 그중 49개만 성공했습니다.99 머신러닝을 사용하면 공격을 계획하는 사람은 소셜 미디어를 스캔하여 피싱 캠페인의 취약한 대상을 파악할 수 있습니다. 또한 머신러닝 알고리즘을 사용하여 이전 피싱 공격에서 수신된 이메일과 응답을 분석하여 더욱 정교하고 진짜처럼 보이는 콘텐츠를 생성함으로써 스팸 필터의 탐지를 피하고 피해자를 속여 악성코드를 설치하도록 유도할 수 있습니다.100

맬웨어를 유포하는 또 다른 방법은 구조화 질의 언어(SQL) 주입 공격입니다. SQL 주입 공격은 AI에 의해 촉진될 수도 있습니다. 예를 들어, AI 기반 DeepHack 도구는 신경망을 사용하여 웹 애플리케이션에 침투하는 방법을 학습합니다.101 이러한 도구를 사용하면 대상 시스템에 대한 사전 지식 없이도 작동하고 맬웨어를 유포할 수 있는 완전 자동화된 해킹 시스템을 구축할 수 있습니다.

Malwarebytes의 2019년 연구에 따르면 당시 AI 기반 악성코드에 대한 실제 증거는 아직 발견되지 않았지만,102 IBM 연구원들이 2018년 Black Hat USA 컨퍼런스에서 "DeepLocker"라는 새로운 악성코드를 발표한 것은 주목할 만합니다.103 DeepLocker 악성코드는 AI가 악성코드 공격을 어떻게 강화할 수 있는지 정확하게 보여주기 위해 개발되었습니다.104 DeepLocker는 화상 회의 소프트웨어로 위장하여 얼굴 및 음성 인식을 통해 의도된 피해자를 식별하고 페이로드를 배포할 때까지 숨어 있습니다. 테러 집단의 손에 이러한 도구가 들어간다면 사이버 테러 위협의 심각성은 더욱 커질 것입니다.

96 JPIAG Charvat. (2009). 사이버 테러리즘: 전장의 새로운 차원. Christian Czosseck과 Kenneth Geers(편), The Virtu-
전장: 사이버 전쟁에 대한 관점.

97 유엔 마약범죄사무소(UNODC)는 유엔 대테러 이행 태스크포스와 협력합니다. 테러 목적으로 인터넷을 사용하는 것에 대한 유엔의 입장입니다.
https://www.unodc.org/documents/terrorism/Publications/Use_of_Internet_for_Terrorist_Purposes/ebook_use_of_the_Internet_for_테러리스트의 목적.pdf

98 Adam Kujawa. (2020년 2월 11일). 사이버 보안 분야의 실제 AI 위협은 공상과학 소설이 아닙니다. Venture Beat. https://venture-beat.com/2020/02/11/real-world-ai-threats-in-cybersecurity-arent-science-fiction/ 에서 확인 가능

99 토마스 브루스터. (2016년 7월 25일). 트위터 피싱에 나보다 더 능숙한 사람은 누구일까? 아니면 인공지능일까? 포브스. https://www.forbes.com/sites/thomasbrewster/2016/07/25/artificial-intelligence-phishing-twitter-bots/#79c3442976e6

아담 쿠자와 100. (2020년 2월 11일). 사이버 보안 분야의 실제 AI 위협은 공상과학 소설이 아닙니다. Venture Beat. https://venture-beat.com/2020/02/11/real-world-ai-threats-in-cybersecurity-arent-science-fiction/ 에서 확인 가능

101 댄 페트로와 벤 모리스. (2017). 머신러닝 무기화: 인류는 어쨌든 과대평가되었다. 데프콘. https://www.defcon.org/html/defcon-25/dc-25-speakers.html#페트로

102 피터 아른츠, 웬디 자모라, 제롬 세구라, 아담 쿠자와. (2019년 6월). 인공지능이 실패할 때: 공상과학과 사실 구분하기. Malwarebytes Labs. https://resources.malwarebytes.com/files/2019/06/Labs-Report-AI-gone-awry.pdf 에서 확인 가능

103 Dhilung Kirat, Jiyong Jang, Marc Ph. Stoecklin. (2018년 8월 9일). DeepLocker — AI 잠금장치 기술을 활용한 표적 공격 은폐.
Black Hat USA. https://i.blackhat.com/us-18/Thu-August-9/us-18-Kirat-DeepLocker-Concealing-Targeted-Attacks-with- AI-Locksmithing.pdf 에서 확인 가능

104 댄 패터슨. (2018년 8월 16일). 무기화된 AI가 새로운 유형의 사이버 공격을 만들어내는 방식. TechRepublic. https://www.techrepublic.com/article/how-weaponized-ai-creates-a-new-breed-of-cyber-attacks/

c. 랜섬웨어

랜섬웨어는 전 세계적으로 가장 심각한 사이버 보안 위협 중 하나로 거듭 지적되어 왔습니다.105 맬웨어의 일종인 랜섬웨어는 피해자의 파일을 암호화하고 파일 해독을 위해 몸값을 요구하는 악성 소프트웨어입니다. 랜섬웨어는 자체 재이식 기능을 통해 수천 대의 기기로 쉽게 확산될 수 있다는 사실로 인해 랜섬웨어 공격의 위협은 더욱 커집니다. 이것은 150개국에 걸쳐 20만 대 이상의 컴퓨터에 영향을 미친 2017년 WannaCry 랜섬웨어 공격에서 예시되었습니다.106 2019년에는 2021년까지 11초마다 랜섬웨어 공격이 발생하고 연간 총 매출이 약 200억 달러에 이를 것으로 추산되었습니다.107 최근 미국 병원들은 2020년 9월과 10월 사이에 랜섬웨어 공격 건수가 71% 증가했다고 보고했으며, 이는 이미 COVID-19 팬데믹으로 심하게 부담을 받고 있는 인프라를 위협하고 있습니다.108

Unsplash의 engin akyurt가 찍은 사진

랜섬웨어에 AI를 통합하면 이러한 공격의 효과를 크게 높일 수 있습니다. 머신러닝 모델을 활용하여 새로운 유형의 공격을 생성하거나 지능적인 타겟팅 기법을 통해 기존 공격의 효과를 증폭시킴으로써 기존 랜섬웨어 생태계를 확산시킬 수 있습니다. 또한, 머신러닝 알고리즘을 활용하여 위에서 설명한 바와 같이 랜섬웨어를 유포하는 피싱 캠페인의 효과를 향상시킬 수 있습니다. AI를 활용하여 정교한 랜섬웨어 공격을 감행함으로써 이미 수익성이 높은 공격 방식을 훨씬 더 수익성 있게 만들 수 있습니다. 이러한 공격은 테러 집단 및 개인의 인프라 또는 활동을 지원하는 수입 창출에도 활용될 수 있습니다.

105 인터넷 조직범죄 위협 평가(IOCTA). (2019년 10월 9일). 유로폴. https://www.europol.europa.eu/activi- 에서 확인 가능
ties-services/main-reports/인터넷-조직-범죄-위협-평가-iocta-2019

로이터 통신 직원 106명. (2017년 5월 14일). 유로폴, 사이버 공격으로 최소 150개국 20만 명 피해. 로이터. https://www.reuters에서 확인 가능.
com/article/us-cyber-attack-europol/cyber-attack-hits-200000-in-at-least-150-countries-europol-idUSKCN18A0FX

107 루카 아레지나. (2019년 11월 13일). 2020년 랜섬웨어 통계: 무작위 공격부터 표적 공격까지. DataProt. https:// 에서 확인 가능
dataprot.net/statistics/ransomware-statistics/#:~:text=By%20that%20time%2C%20the%20global,billion%20in%20revenue%20for%20
사이버 범죄자들.

108 패트릭 하웰 오닐. (2020년 10월 29일). 코로나바이러스 확산으로 미국 병원에 랜섬웨어가 잇따라 감염. https://www.tech-
nologyreview.com/2020/10/29/1011436/코로나바이러스 급증으로 미국 병원에 랜섬웨어 공격 급증/

몸값을 요구하는 납치를 선호하는 테러리스트와 폭력적인 극단주의 집단을 떠올려 보면, 효과적이고 수익성 있는 AI 기반 랜섬웨어 공격은 "몸값을 요구하는 납치 2.0"의 한 형태로서 그들의 레퍼토리에 자연스럽게 들어맞는 것으로 보입니다.

랜섬웨어의 주요 목표는 전통적으로 피해자에게서 돈을 갈취하는 것이었지만, 2017년 NotPetya 공격은 공격이 파괴적이거나 교란적인 목적으로도 사용될 수 있음을 보여주었습니다. NotPetya 공격에서 랜섬웨어는 암호화를 해제하고 시스템을 공격 이전 상태로 되돌릴 수 없도록 변경되었습니다.110 이러한 측면에서 AI 지원 랜섬웨어 공격은 테러 단체와 개인이 자금 조달 목적뿐 아니라 파괴적인 목적으로도 사용할 수 있습니다.

d. 비밀번호 추측

비밀번호는 해킹에 대한 최전선 방어선이며, 대기업부터 일반 가정까지 모든 사이버 방어 전략에 필수적입니다. 보호된 웹사이트에 접속하기 위한 비밀번호를 획득하면 악의적인 공격자가 시스템이나 네트워크에 침입하여 필수 서비스를 방해하거나, 교란을 일으키거나, 귀중한 데이터나 정보를 훔치거나, 데이터나 프로세스를 조작하거나, 악성 소프트웨어를 설치하는 등의 행위를 할 수 있습니다. ISIL과 같은 테러 단체의 지지자들은 웹사이트와 소셜 미디어 계정을 해킹하여 훼손하고 선전 자료를 유포하는 오랜 역사를 가지고 있습니다.111

보안 강화를 위해 플랫폼과 웹사이트는 비밀번호 추측을 방지하기 위한 다양한 조치를 도입했습니다. 여기에는 최소 8자리 이상의 긴 비밀번호나 영숫자, 대문자, 소문자를 조합한 비밀번호가 포함됩니다. 그럼에도 불구하고 사람들은 비밀번호를 선택할 때 이름, 성, 생년월일을 조합하는 등 특정 패턴을 따르는 경향이 있습니다. 또한 간단하고 예측 가능한 비밀번호를 사용하고 여러 서비스에서 동일한 비밀번호를 재사용하는 경향이 있습니다. 이는 해커의 활동을 크게 용이하게 합니다.112

해커들은 다양한 플랫폼에서 유출된 비밀번호 데이터베이스를 온라인에서 쉽게 찾을 수 있으며, 이를 통해 정보를 얻고 웹사이트 해킹에 활용할 수 있습니다. 예를 들어, Avast는 온라인에서 30,160,455,237건의 비밀번호가 유출되었다고 보고했습니다.113 "John the Ripper"와 같은 비밀번호 추측 도구는 이러한 데이터베이스를 활용하여 비밀번호를 추측하지만, 공격 계획을 수립하기 위해서는 상당한 양의 수동 코딩 작업이 필요합니다.

하지만 AI의 발전은 비밀번호 추측 과정을 크게 가속화하고, 향상시키고, 자동화하는 데 활용될 수 있습니다. 악의적인 행위자는 이러한 방대한 온라인 비밀번호 데이터베이스를 이용하여 신경망을 훈련시킬 수 있으며, 이는 인간이 상상할 수 있는 것보다 훨씬 정교한 비밀번호 변형을 생성할 수 있습니다. 이러한 신경망은 해결책이 결정될 때까지 여러 차례 시도를 연달아 실행하여 해커의 직접적인 개입 필요성을 줄일 수 있습니다. 2017년 한 연구에서 연구원들은 수천만 개의 유출된 비밀번호를 새로운 비밀번호를 생성하는 신경망에 입력했습니다. 그런 다음 이 비밀번호들을 링크드인과 같은 사이트에서 유출된 비밀번호와 교차 참조하여 신경망이 사용자 비밀번호를 얼마나 성공적으로 해독하는지 측정했습니다. 이 연구에서는 링크드인 비밀번호 세트의 27%를 해독할 수 있다는 것을 발견했습니다.114 후속 연구에서는 AI가 화상 통화 중 분석된 어깨 움직임을 기반으로 어떤 키가 입력되는지 감지하여 비밀번호를 추측할 수 있다는 것을 발견했습니다. 연구 결과에 따르면 문제의 AI 소프트웨어의 정확도는 75%~93%로 놀라울 정도로 높았습니다.115

109 유엔 안전보장이사회. (2014년 10월 29일). 알카에다 및 관련 개인 및 단체에 관한 결의안 2161호(2014)에 따라 제출된 분석 지원 및 제재 감시팀의 제16차 보고서(S/2014/770). https://www.securitycouncilreport.org/atf/cf/{65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9}/s_2014_770.pdf

110 앤디 그린버그. (2018년 8월 22일). 역사상 가장 파괴적인 사이버 공격, NotPetya에 대한 숨겨진 이야기. Wired. https:// 에서 확인 가능 www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/

ISIS 동조자 111명, 워드프레스 사이트 훼손 및 악용 시도 - FBI 경고 (2015년 4월 13일). TrendMicro. https://www.trendmicro.com/vinfo/tr/security/news/cyber-attacks/isis-sympathizers-defacing-and-exploiting-wordpress-sites-fbi-warns 에서 확인 가능

112 JM Porup. (2018년 3월 28일). 14억 개의 도난된 비밀번호가 무료로 이용 가능하다: 우리가 지금 알고 있는 사실. CSO. https://www.csoonline에서 확인 가능. com/article/3266607/1-4b-도난당한-비밀번호-우리가-지금-알고있는-것을-취득하는-것은-무료입니다.html

113 내 비밀번호가 도난당했나요? (nd) Avast. https://www.avast.com/hackcheck 에서 확인 가능

114 매튜 허트슨. (2017년 9월 15일). 인공지능 덕분에 비밀번호 추측이 훨씬 쉬워졌습니다. 과학. 다음에서 이용 가능 https://www.sciencemag.org/news/2017/09/인공지능이 당신의 비밀번호를 추측하는 걸 훨씬 더 쉽게 만들어 냈습니다.

115 코너 콜리. (2020년 11월 11일). AI, 이제 어깨를 보고 비밀번호를 추측할 수 있다. Tech.co. https://tech.co/ 에서 접속 가능 뉴스/AI 추측 비밀번호 어깨

연구에 따르면 대규모 비밀번호 데이터 세트를 분석하고 원래 예시와 유사한 변형을 생성하는 GAN을 사용하여 정교한 비밀번호 생성을 수행할 수 있다는 사실이 밝혀졌습니다. 즉, 사전에 수십억 개의 추측 비밀번호를 생성할 수 있으므로 보다 타겟팅되고 효과적인 비밀번호 추측이 가능합니다.116

Pixabay의 TheDigitalWay의 이미지

이러한 개발은 고유하고 강력한 비밀번호에 대한 강조점을 높이고, 예를 들어 모바일 기기를 통한 2단계 또는 다중 요소 인증을 추가 방어 계층으로 구축하지만, 후자조차도 완벽한 방어 수단을 제공하지 못한다는 점을 기억하는 것이 중요합니다. 특히 사회 공학 측면에서 취약성이 있기 때문입니다. 아래에서 설명하겠지만 머신 러닝도 사회 공학을 강화할 수 있는 또 다른 영역입니다.

e. CAPTCHA 해독

CAPTCHA(Completely Automated Public Turing test to tell Computers and Humans Apart)는 네트워크와 웹사이트를 공격으로부터 보호하기 위해 고안된 또 다른 중요한 보안 수단입니다. 이름에서 알 수 있듯이 CAPTCHA는 실제 사용자와 스팸 로봇을 구분하여 사람이 로봇에 접근하고 차단할 수 있도록 합니다.117 웹사이트는 CAPTCHA를 사용합니다. 예를 들어, 사람이 아닌 사람이 웹메일 계정에 자동으로 접근하면 스팸 이메일 메시지가 급증할 수 있으며, "블로그 스패머"는 경제적 이익을 얻기 위해 인위적으로 클릭 수를 부풀려 이득을 볼 수 있습니다.118 이 시스템은 접근 요청이 사람인지 컴퓨터인지 파악하기 위해 시도-응답 인증을 적용합니다.119

CAPTCHA 시스템을 해독하려는 노력은 도입 초기부터 시작되었지만, 머신러닝의 발전으로 CAPTCHA 시스템 해독 방법이 더욱 정교해졌습니다. 2013년, 한 AI 스타트업은 뇌를 모방한 소프트웨어를 통해 CAPTCHA 시스템을 해독하는 데 성공했다고 주장했습니다. 이 소프트웨어는 시스템 학습을 위한 대규모 데이터셋이나 높은 컴퓨팅 파워 없이도 90% 이상의 성공률을 기록했습니다. 이 알고리즘은 숫자와 문자를 인식하도록 훈련되었으며, 지문으로 구성된 것처럼 보이는 문자로 구성된 CAPTCHA에서 특히 높은 효율성을 보였습니다.120 그 이후로 CAPTCHA 시스템을 해독하기 위한 머신러닝 및 딥러닝 기술에 대한 관심이 증가하고 있습니다.121

116 브리란드 히타지, 파올로 가스티, 주세페 아테니제, 페르난도 페레즈-크루즈. (2017). PassGAN: 비밀번호 추측을 위한 딥러닝 접근법. Arxiv. https://arxiv.org/pdf/1709.00440.pdf 에서 접속 가능

117 Luis von Ahn, Manuel Blum, Nicholas J. Hopper & John Langford. (2003). CAPTCHA: 보안을 위한 AI의 경성 문제 활용. 국제 암호기술 이론 및 응용에 관한 컨퍼런스, 베를린, 하이델베르크.

118 M. Motoyama 외 (2010년 8월). Re: CAPTCHA - 경제적 맥락에서 CAPTCHA 해결 서비스 이해, USENIX Security"10: 제19회 USENIX 보안 컨퍼런스 회의록, 10. https://dl.acm.org/doi/10.5555/1929820.1929858 에서 접속 가능

119 Google Workspace 관리자 도움말. CAPTCH란 무엇인가요? Google. https://support.google.com/a/answer/1217728?hl=ko 에서 확인할 수 있습니다.

120 Rachel Metzarchive. (2013년 10월 28일). AI 스타트업, 캡차를 극복했다고 주장. MIT Technology Review. https://www.tech- 에서 확인 가능. nologyreview.com/2013/10/28/175597/ai-startup-says-it-has-defeated-captchas/

121 G. Ye 외 (2018). Yet Another Text Captcha Solver: 생성적 적대 네트워크 기반 접근법, 2018 ACM SIGSAC 컴퓨터 및 통신 보안 학회 논문집.

CAPTCHA 시스템을 극복하면 테러 집단과 개인이 사이버 공격을 수행하는 데 큰 도움이 될 수 있습니다.
예를 들어, 악성 소프트웨어, 테러 관련 내용 또는 기타 방해적이거나 선전적인 내용이 담긴 자동화된 대규모 스팸 이메일을 배포하는 것이 허용될 수 있습니다.

암호화 및 복호화

일상생활의 필수 요소가 점점 더 디지털화되는 사회에서 암호화는 정부 기관, 기업, 그리고 일반 대중이 통신 및 저장된 정보의 기밀성, 무결성, 그리고 접근성을 보호하는 데 필수적입니다. 암호화는 메시지나 정보 등의 데이터를 무단 접근을 방지하는 방식으로 변환하는 과정으로 이해될 수 있습니다. 다시 말해, 복호화는 암호화된 데이터를 초기 상태로 되돌리는 과정입니다.122

암호화는 허가받지 않은 접근을 방지하는 강력하고 실용적인 도구이지만, 범죄자와 테러리스트를 포함한 더욱 사악한 의도를 가진 자들도 이를 사용하여 익명성을 유지하면서 안전하게 통신하고 정보를 공유할 수 있게 해줍니다.123 복호화는 이러한 그룹이나 개인에게도 마찬가지로 매력적입니다. 예를 들어, 이를 통해 기밀 정보에 접근할 수 있기 때문입니다.

2012년 주목할 만한 사례 중 하나는 테러 조직 구성원 간에 교환된 수십 건의 암호화된 이메일이 법원에 제출된 사건에서 프랑스 국민이 5년 징역형을 선고받은 것입니다.124 테러 조직이 구성원 간의 은밀한 온라인 통신을 용이하게 하기 위해 "Mujahedeen Secrets"라는 암호화 소프트웨어를 사용하고 있다고 주장되었습니다.125 또한 일부 테러 조직이 속기 및 암호화를 통해 배포 및 공유되는 정보를 가리기 위해 Camouflage 및 WinZip과 같은 소프트웨어를 이용하고 있는 것으로 나타났습니다.126

AI 기반 암호화 도구가 현재 연구되고 있습니다. 2016년, 구글 브레인 연구진은 두 신경망이 세 번째 신경망이 메시지를 가로채지 못하도록 서로 통신하도록 성공적으로 훈련시켰습니다. 연구 결과, 처음 두 신경망은 암호화된 메시지를 주고받는 자체적인 방식을 자율적으로 생성하고 이를 해독하는 데 성공했습니다.127 흥미롭게도, 컬럼비아 대학교 연구진은 2018년, 겉보기에 평범해 보이는 텍스트에 민감한 정보를 효과적으로 삽입할 수 있는 딥러닝 기술을 개발했습니다. 이를 통해 텍스트, 이미지 또는 QR 코드와 같은 정보를 육안으로는 볼 수 없는 곳에 숨길 수 있었습니다.128

AI의 발전으로 암호화 및 복호화 기술은 더욱 강력해질 수 있습니다. AI 기반 정교한 암호화 기술을 활용하면 테러 조직 구성원들은 정보의 무결성이 침해되지 않고 더욱 쉽게 서로 소통할 수 있게 됩니다. 또한 AI 기반 복호화 기술을 통해 테러 조직은 대테러 기관이 전달하는 민감한 암호화 정보에 더욱 쉽게 접근할 수 있게 됩니다.

122 암호화 감시 기능 제1차 보고서(2019). 유로폴(EUROPOL) 및 유로저스트(EUROJUST). https://www.europol.europa.eu/ 에서 확인 가능
출판물-문서/천문대 기능 암호화에 대한 첫 번째 보고서

123 로버트 그레이엄. (2016년 6월). 테러리스트의 암호화 활용. 9권, 6호. CTC Sentinel, 9(6). https://www.ctc.usma.edu/ 에서 열람 가능
테러리스트들이 암호화를 사용하는 방법

124 UNODC는 유엔 대테러 이행 태스크포스와 협력하여 (2012년 9월). 인터넷 사용
테러 목적, 유엔. 접근 가능
https://www.unodc.org/documents/terrorism/Publications/Use_of_Internet_for_Terrorist_Purposes/ebook_use_of_the_internet_for_
테러리스트의 목적.pdf

125 UNODC는 유엔 대테러 이행 태스크포스와 협력합니다. (2012년 9월). 인터넷 사용
테러 목적. 유엔. 접근 가능
https://www.unodc.org/documents/terrorism/Publications/Use_of_Internet_for_Terrorist_Purposes/ebook_use_of_the_internet_for_
테러리스트의 목적.pdf

126 무인항공기(UAS) 위협 대응을 위한 GCTF 이니셔티브. (2019년 9월). 테러리스트의 무인항공기 사용 대응을 위한 모범 사례에 관한 베를린 각서. 제10차 GCTF 장관급 총회, 뉴욕. https://
www.thegctf.org/ 에서 접속 가능
포털/1/문서/프레임워크 문서/2019/베를린 메모란덤 EN.pdf?ver=2020-01-13-143548-187

127 M. Abadi & DG Andersen. (2016). 적대적 신경망 암호화를 이용한 통신 보호 학습, arXiv 사전 인쇄본 arX-
iv:1610.06918.

128 Nvidia. (2018년 4월 12일). AI를 사용하여 눈에 보이지 않는 메시지 암호화. Nvidia. https://news.developer.nvidia.com/using-ai- 에서 확인 가능
메시지를 공개적으로 암호화하려면/

ii. 물리적 공격 가능

가. 자율주행차

차량, 특히 승용차, 밴, 트럭은 오랫동안 테러 공격에 사용되어 왔습니다. 이러한 차량의 사용 사례는 셀 수 없이 많습니다. 예를 들어, 2016년 12월 베를린 크리스마스 시장 테러129와 2017년 8월 바르셀로나 테러130 에서 볼 수 있듯이, 차량은 고의적인 충돌 공격 에 사용되었습니다. 또한, 2018년 카불 구급차 폭탄 테러(103명 사망, 235명 부상)131와 같이 차량 폭탄 공격에도 차량이 사용되었습니다.

AI의 가장 널리 알려진 응용 분야 중 하나는 자율주행차입니다. 자율주행차는 자율주행 또는 무인자동차라고도 합니다. 많은 사람들이 자율주행차를 미래의 더 안전하고 편리하며 효율적인 교통수단으로 여깁니다. 차량에 내장된 AI는 딥러닝 기술을 활용하여 운전자의 의사결정 과정을 모방하여 차량의 움직임, 조향, 가속, 브레이크 작동 등을 제어합니다.132 테슬라와 구글과 같은 기업들은 오랫동안 이 기술의 실용화를 주장하며 자율주행차 연구, 시험 및 개발을 주도해 왔습니다. 최근 몇 년 동안 수많은 주요 자동차 회사들이 자율주행차 상용화에 동참했습니다. 2019년 11월, 구글 모회사인 알파벳의 자회사인 웨이모는 미국 애리조나주 피닉스에서 안전 보조 운전자가 없는 자율주행 택시 서비스를 시작하며 중요한 이정표를 세웠습니다.133

기술의 급속한 발전과 자율주행차 산업에 대한 상당한 상업적 투자, 그리고 법적 및 정책적 과제를 해결하는 데 있어 수많은 이정표 달성을 고려할 때 AI가 결국 운전 경험을 변화시킬 것이라는 점은 불가피해 보이지만, 이것이 정확히 언제 일어날지는 아직 알 수 없습니다.134

테러리즘과 차량의 광범위한 역사를 되돌아보면, 자동차의 자율성 향상은 테러 집단에게 매우 유용한 발전이 될 수 있습니다. 테러 집단은 추종자가 생명을 희생하거나 체포될 위험을 감수할 필요 없이 원격으로 가장 전통적인 유형의 공격을 효과적으로 수행할 수 있습니다.135 완전 자율 차량에 탑재된 즉석 폭발 장치를 사용하여 공격을 용이하게 하는 것 외에도 자율 주행 자동차가 도로를 막는 심각한 사고를 일으키거나 자율 주행으로 인한 대학살을 일으키는 데 사용될 수 있다는 의견도 제시되었습니다.136

그럼에도 불구하고, 보행자 충돌과 같은 상황을 감지하고 회피하거나, 제동 시스템을 작동시키거나, 차량을 다른 경로로 설정하는 등의 안전 기능이 이러한 차량을 이용하려는 테러 음모를 좌절시킬 수 있다는 근거가 있습니다. 실제로 앞서 설명한 바와 같이, 자율주행차와 테러리즘을 둘러싼 몇 가지 진전이 이미 있었습니다. 여기에는 초보적인 실험과 ISIL 지지자들의 계획에 대한 증언이 포함되었으나, 실현되지는 않았습니다.

물론 "차량"이라는 용어는 바퀴 달린 자동차만을 의미하는 것이 아니라 잠수함과 같은 지하 차량과 "드론"이라고 흔히 불리는 무인 항공 시스템과 같은 비행체도 포함한다는 점에 유의해야 합니다. 앞서 언급했듯이 드론은 대부분 원격 조종되며 제한된 수준의 자율성을 가지고 있습니다.

---

129 제이슨 해나. (2016년 12월 23일). 베를린 크리스마스 마켓 테러: 희생자들. CNN. https://edition.cnn.com/2016/12/23/eu- 에서 확인 가능
로프/베를린-크리스마스-마켓-공격-피해자/

130 CNN 편집 연구. (2020년 9월 6일). 차량 테러 공격 관련 속보. CNN. https://edition.cnn.com/2017/05/03/ 에서 확인 가능
세계/테러리스트-차량-공격-빠른-사실/index.html

131 제임스 두벡과 에이미 헬드. (2018년 1월 27일). 카불 탈레반 차량 폭탄 테러로 최소 103명 사망, 235명 부상. NPR.
https://www.npr.org/sections/thetwo-way/2018/01/27/581265342/dozens-killed-more-than-100-wounded-in-taliban-car-bombing-in-kabul?t=1598955985489

132 케이티 버크. (2019년 5월 7일). 자율주행차는 어떻게 결정을 내릴까? 엔비디아. https://blogs.nvidia.com/blog/2019/05/07/ 에서 확인 가능
자율주행차가 결정을 내리다

133 앤드류 J. 호킨스. (2019년 12월 9일). 웨이모의 자율주행차: 로봇 택시 뒷좌석에서 고스트 라이딩. 더 버지. https://www.theverge.com/2019/12/9/21000085/waymo-fully-driverless-car-self-driving-ride-hail-service-phoenix-arizona 에서 확인 가능

134 CB Insights. (2020년 12월 16일). 자율주행차 개발에 참여하는 40개 이상의 기업. CB Insights. https://www.cbinsights.com/ 에서 확인 가능
연구/자율주행차 기업 목록/

135 제프리 W. 루이스. (2015년 9월 28일). 모든 차고에 스마트 폭탄? 자율주행차와 테러 공격의 미래. 스마트하고, 접근 가능하며
https://www.start.umd.edu/news/smart-bomb-every-garage-driverless-cars-and-future-terrorist-attacks 에서

136 A. Lima 외 (2016). 안전하고 보안성이 확보된 자율주행 및 협동차량 생태계를 향하여, 제2차 ACM 워크숍 논문집
사이버 물리 시스템 보안 및 개인정보 보호에 관한 것입니다.

토노미(tonicity)뿐만 아니라 AI는 드론이 더욱 자율적으로, 심지어는 완전히 자율적으로 될 가능성을 제시합니다.137 드론에 AI를 적용하는 것을 연구하는 연구자들은 드론 군집을 위한 자율 제어 시스템을 개발하여 지상의 인간 조종자의 개입 없이 드론이 방향을 조정하고 심지어 배럴 롤(barrel roll)과 플립(flip)과 같은 곡예 기동까지 수행할 수 있도록 했습니다.138 이 부분에서도 기술적 발전은 자율 드론 군집의 가능성을 열어줍니다. 기술적 관점에서 볼 때, 드론 프로그래머가 고려해야 할 변수의 수가 적고 드론에 적용되는 법적 체계가 간소화되었다는 점을 고려할 때, 자율 드론 또는 잠수정의 개발이 무인 자동차보다 가까운 미래에 더 실현 가능할 수 있다는 주장도 있습니다.

Unsplash의 John Rodenn Castillo가 촬영한 사진

b. 얼굴 인식 기능이 있는 드론

얼굴 인식 기술 도입은 지난 몇 년 동안 머신 러닝의 급속한 발전에 힘입어 급격히 증가했습니다. 얼굴 인식 기술의 상용화는 전자 기기 접근 인증 개선부터 공항 탑승 및 보안 검색 가속화에 이르기까지 새로운 기회를 창출했습니다. 앞으로 얼굴 인식 도입은 새로운 서비스로 확장될 수 있으며, 궁극적으로는 서비스 접근을 위한 선호 인증 수단이 될 수도 있습니다.

2017년, 미국 비영리 연구 기관이자 홍보 단체인 퓨처 오브 라이프 인스티튜트(Future of Life Institute)는 "슬로터봇(Slaughterbots)"이라는 제목의 영상을 공개했습니다. 이 영상에서는 수 그램의 폭발물을 실은 마이크로 드론 무리가 얼굴 인식 기술을 사용하여 표적을 식별하고 가미카제 방식으로 공격합니다. 얼굴 인식 기술을 통해 컨트롤러는 드론이 수집한 이미지와 내장된 얼굴 인식 데이터베이스에 업로드된 이미지를 교차 참조하여 선택된 표적을 자동으로 포착, 식별 및 공격하도록 프로그래밍할 수 있었습니다.139 비록 과장된 연출에도 불구하고, 이 영상은 빠르게 퍼져나가 온라인에서 최대 300만 뷰를 기록하며 이러한 기술들의 조합 가능성을 뜨거운 화제로 만들었습니다. 다행히 이 기술은 기성품 형태로 존재하지는 않지만, 완전히 새로운 개념도 아니고 단순한 공상과학 소설도 아닙니다. 이미 여러 상업용 드론 제품에 제한적인 얼굴 인식 기능이 포함되어 있지만, 비행 기능 잠금 해제 및 "팔로우미" 모드 활성화와 같은 특정 기능에 국한되어 있습니다. 현재 드론에는 비행 중 개인을 식별하고 타겟팅하는 얼굴 인식 기술이 포함되지 않습니다.140 이런 점에서 드론에서 얼굴 인식을 사용하는 것은 훨씬 더 제한적입니다.

137 Vikram Singh Bisen. (2020년 2월 5일). AI 기반 드론 작동 방식: 인공지능 드론 활용 사례. 중급. https:// 에서 접근 가능 medium.com/vsinghbisen/how-ai-based-drone-works-artificial-intelligence-drone-use-cases-7f3d44b8abe3

138 Nick Lavars. (2020년 6월 23일). AI 알고리즘을 통해 자율 드론이 롤링과 플립을 수행할 수 있습니다. 새로운 아틀라스. https://ne- 에서 이용 가능 watlas.com/drones/ai-algorithm-autonomous-drones-barrel-rolls-flips/

139대의 슬로터봇. (2017년 11월 13일). 자율 무기를 멈춰라. 유튜브. https://www.youtube.com/watch?v=9CO6M2HsoIA 에서 접속 가능.

140 Larry Haller. (2020년 2월 13일). 이 4대의 드론이 당신의 얼굴을 인식할 수 있습니다. Drones Globe. https://www.dronesglobe.com/ 에서 접속 가능 가이드/얼굴 인식/

전 세계 여러 법 집행 기관은 이미 이러한 기술들을 결합하여 실험을 시작했습니다. 예를 들어 실종자 및 취약 계층 수색을 지원하거나 혼잡한 공간에서 용의자를 식별하는 데 활용되고 있습니다.141, 142 이처럼 자동화된 방식으로 표적을 찾고 추적하고 식별하는 능력은 법 집행 기관에 당연히 매력적이지만, 이러한 기술의 사용은 대량 감시 및 인권 침해, 특히 사생활권 침해에 대한 우려를 불러일으켰습니다. 법 집행 기관의 안면 인식 활용을 둘러싼 여러 논란의 여지가 있는 기술들을 고려할 때, 이 AI 기술의 미래는 불확실합니다.143

테러 집단이 드론을 악의적으로 공격에 사용하는 것은 아직 심각한 위협은 아니지만, 점차 확산되고 있는 추세입니다. 하지만 안면 인식 기술의 통합은 드론 위협 수준을 크게 높이고 매우 정밀한 공격이 가능해짐에 따라 이러한 측면에서 획기적인 변화를 가져올 것입니다. 이 기술은 아직 쉽게 구할 수 없지만, 관련 전문 지식을 갖춘 사람이라면 다양한 요소들을 직접 결합하여 "슬로터봇(Slaughterbot)" 기능을 개발할 수 있습니다.144

7점. 유전자를 표적으로 삼은 생물무기

COVID-19 팬데믹은 개인 사망부터 대규모 세계 경제 침체까지 광범위한 부정적 영향을 미쳤습니다. 2020년 7월, 유엔 사무총장 안토니우 구테흐스는 "이 팬데믹은 디지털 기술 오용, 사이버 공격, 생물 테러와 같은 새롭고 떠오르는 형태의 테러에 대한 취약성을 부각시켰다"고 지적했습니다.145 실제로, 특히 AI와 결합된 생명공학과 같은 새로운 신기술은 특정 유전적 집단을 특별히 표적으로 삼는 치명적인 새로운 병원균 변종을 개발할 수 있는 기회를 제공할 수 있지만, 이를 실현하기 위한 기술적 장벽은 상당할 것입니다.146 유엔 글로벌 대테러 전략 이행에 대한 유엔 시스템의 활동에 대한 보고서에서 구테흐스는 합성생물학을 테러의 관점에서 위험을 초래할 수 있는 새롭고 떠오르는 기술의 한 예로 구체적으로 지적했습니다.147

유전자 염기서열 분석 기술의 발전으로 다양한 분야의 연구자들은 더 많은 유전자 데이터를 처리하고 더 많은 유전자 정보를 추출할 수 있게 되었습니다. 이러한 발전은 과학적 발전을 촉진하고 삶의 질을 크게 향상시킬 수 있지만, 보안 문제를 야기합니다. 머신러닝 모델을 적용함으로써 연구자들은 유전자 기반 진단 및 치료 기술을 더욱 발전시킬 수 있습니다.148

수집된 유전 물질의 대부분은 유전자 데이터베이스나 바이오뱅크에 저장되어 관련 연구자들 사이에 유통됩니다. 다시 말해, 수집된 유전 물질과 그로부터 추출된 정보는 규제가 거의 없는 연구 이해관계자 집단에서도 점점 더 쉽게 접근할 수 있게 되었습니다.

141 켄 맥도널드. (2019년 11월 4일). 경찰, AI 인식 드론을 활용하여 실종자 수색에 나선다. BBC. https://www.bbc.com/ 에서 접속 가능 뉴스/영국-스코틀랜드-50262650

142 Faine Greenwood. (2020년 7월 8일). 경찰 드론이 당신의 얼굴을 인식할 수 있을까요? Slate. https://slate.com/technology/2020/07/ 에서 확인 가능 경찰 드론 얼굴 인식.html

143 닐라 발라, 케일럽 와트니. (2019년 6월 20일). 경찰의 안면 인식 사용에 대한 적절한 한계는 무엇인가?. 브루킹스. https://www.brookings.edu/blog/techtank/2019/06/20/경찰의 얼굴 인식 활용에 대한 적절한 한계는 무엇인가/

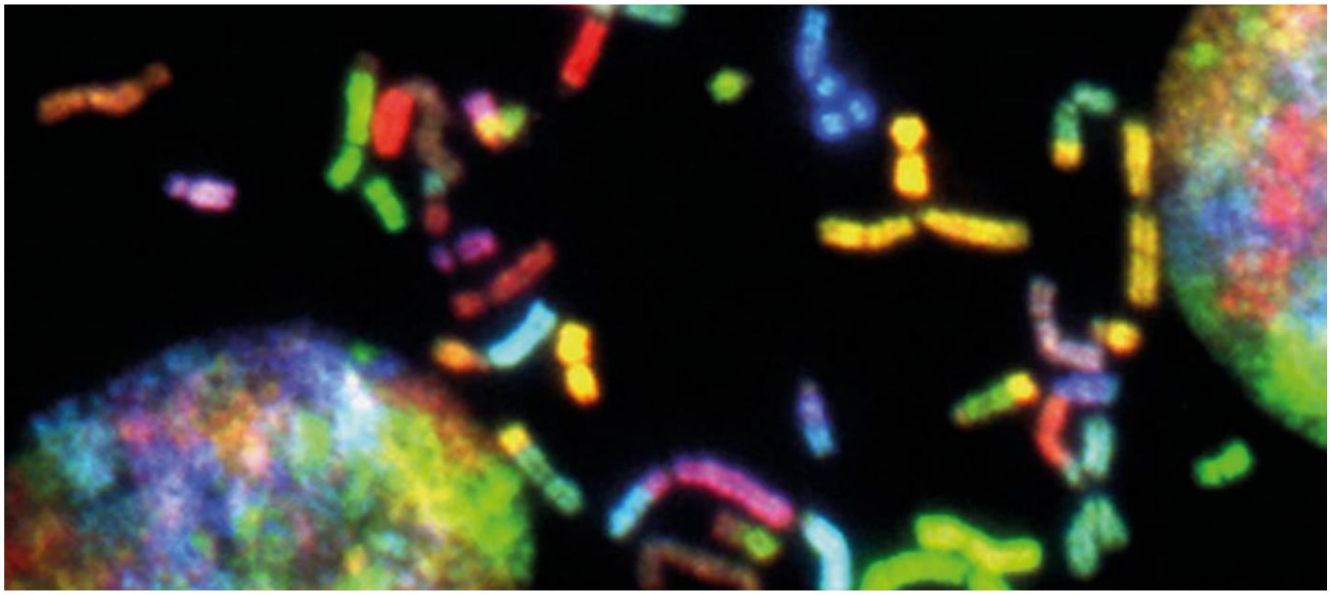
144 렌스케 반 데르 비어. (2019). 기술 시대의 테러리즘, 전략 모니터 2019-2020. 헤이그 전략연구센터와 클링엔달 연구소. https://www.clingendael.org/pub/2019/strategic-monitor-2019-2020/terror-ism-in-the-age-of-technology/ 에서 확인 가능

145 안토니우 구테흐스. (2020년 7월 6일). 유엔 가상 대테러 주간 개막 연설. https://www.un.org/sg/en/content/sg/statement/2020-07-06/secretary-generals-remarks-the-opening-of-the-virtu- al-counter-terrorism-week-united-nations-delivered 에서 접근 가능

146 비벡 와드와(2020년 9월 11일). 유전공학 지니가 병에서 나왔다. 포린 폴리시. https://foreignpolicy에서 확인 가능. com/2020/09/11/크리스퍼-팬데믹-유전자-편집-바이러스/

147 유엔 총회. (2020년 2월 7일). 유엔 글로벌 대테러 전략 이행에 있어서 유엔 시스템의 활동에 대한 사무총장 보고서. https://undocs.org/pdf?symbol=en/A/74/677 에서 열람 가능

148 S. Sawaya, E. Kennally, D. Nelson & G. Schumacher. (2020). 인공지능과 유전자 데이터의 무기화. SSRN Electronic 신문.



Unsplash에 게재된 국립암연구소 사진

이러한 개방형 과학 모델의 장점에도 불구하고, 유전 물질 및 파생 정보의 접근 및 사용 장벽을 낮추거나 아예 없애는 것은 바람직하지 않은 결과를 초래할 수 있습니다.149 악의적인 행위자가 이러한 유전 데이터에 접근하게 되면, 위험한 병원체를 생성하기 위해 기계 학습 모델을 훈련하는 데 이를 활용할 가능성이 점점 더 커지고 있습니다. 단일염기다형성(SNP)과 같은 식별 가능한 유전 마커를 사용하여 인종 집단을 구별할 수 있다면, 이론적으로는 특정 인종의 개인을 표적으로 삼는 것이 가능할 것입니다.150 그러나 생명공학 연구자들이 그러한 특정 집단을 표적으로 삼을 유전적 지식을 확보하는 것은 아직까지 어려운 일이며, 이러한 악의적인 시도에는 고도의 기술력과 특수 장비가 필요하다는 점에 유의해야 합니다.

iii. 테러리즘 자금 조달 수단 제공

에이. 오디오 딥페이크151

사전 녹음된 메시지를 전달하는 데 사용되는 대규모의 원치 않는 자동 전화인 로보콜의 효율성을 개선하는 것 외에도, 머신 러닝은 악의적인 전화 계획에서 중요한 역할을 할 수 있습니다. 특히, "딥페이크" 오디오 콘텐츠의 도입은 개인이 아는 사람과 통신하고 있다고 믿게 하는 데 사용될 수 있습니다.152 아래에서 더 자세히 설명하겠지만, 딥페이크는 인간이나 기술적 솔루션으로도 진짜와 즉시 구별하기 어려운 시각적 및 오디오 콘텐츠를 조작하거나 생성하는 AI 기술을 사용합니다. 딥페이크 오디오를 생성하기 위해 머신 러닝 엔진은 컨퍼런스 콜, YouTube, 소셜 미디어 업데이트, 심지어 TED 토크를 사용하여 일부 대상 개인의 음성 패턴을 복사한 다음 동일한 음성 특성을 가진 새로운 오디오를 생성하도록 훈련됩니다.

149 위와 같음.

150 Tao Huang, Yang Shu, Yu-Dong Cai. (2015). 민족 간 유전적 차이. BMC Genomics 16, 1093. https://doi.org/10.1186/s12864-015-2328-0

151 이 섹션은 내재적인 사회 공학 섹션을 고려하여 아래의 사회 공학 섹션과 관련하여 또는 이를 참조하여 읽어야 합니다. 오디오 딥페이크가 사회 공학적 활동에 악용될 가능성이 있습니다. 아래 제6장 v(d)항을 참조하세요.

152 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philip Amann, Aglika Klein, Odhran McCarthy 및 Maria Eira. (2020년 11월 19일). 인공지능의 악의적 사용 및 남용. Trend Micro, EUROPOL, UNICRI. http://unicri.it/sites/default/ 에서 확인 가능 파일/2020-11/AI%20MLC.pdf

바로 이 수법이 2019년에 사용되었는데, 당시 영국에 본사를 둔 한 회사의 CEO는 모회사의 CEO라고 생각되는 사람으로부터 헝가리 공급업체에 24만 4천 달러를 긴급하게 보내달라는 전화를 받았습니다. 나중에 발신자가 AI 기반 소프트웨어를 사용하여 CEO의 음성을 흉내 낸 것이 밝혀졌습니다.153 이와 동일한 수법이 2020년 7월 미국에 본사를 둔 한 기술 회사를 사취하기 위해 사용되었습니다.154

이러한 악의적인 계획이 수익성 있는 사업으로 판명될 경우, 테러 집단은 사람들을 속이거나 위협하여 자금을 조달하기 위해 이를 이용할 수 있습니다. 이러한 계획은 금전적 이득을 위한 것 외에도, 중요 직책에 있는 사람들을 사칭하거나 딥페이크가 내장된 로보콜 시스템을 이용하여 중요 직책에 있는 사람들을 속여 정보를 수집하고 감시하는 데에도 악용될 수 있습니다.

비. 암호화폐 거래

2009년, 사토시 나카모토라는 이름으로만 알려진 수수께끼의 그룹이나 개인이 비트코인이라는 P2P 전자 지불 시스템에 대한 백서를 발표했습니다.155 이 백서는 암호화로 보호되는 분산형 전환 가능 가상 화폐의 기반을 마련했으며, 이를 통해 위조나 이중 지출이 거의 불가능해졌습니다.

백서 발표 직후, 나카모토가 비트코인의 제네시스 블록을 "채굴"하면서156 디지털 자산에 대한 엄청난 관심과 투자 열풍이 불었습니다. 라이트코인, 리플, 이더리움, 모네로, 리브라, 밈에서 영감을 받은 도지코인 등 새로운 형태의 가상 자산, 즉 암호화폐가 곧 등장하기 시작했습니다.157

Unsplash의 Pierre Borthiry가 찍은 사진

153 캐서린 스터프. (2019년 8월 30일). 사기꾼들이 AI를 이용해 CEO의 목소리를 흉내낸 특이한 사이버 범죄 사건. 월스트리트 저널. https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 에서 확인 가능

154 로렌초 프란체스키-당신은 그 잔을 마실 것입니다. (2020년 7월 23일). CEO를 사칭한 대담한 사기 시도를 담은 딥페이크 오디오를 들어보세요. 바이스. 입장- https://www.vice.com/en_us/article/pkyqvb/deepfake-audio-impersonating-ceo-fraud-attempt 에서 확인 가능

155 S. 나카모토. (2009). 비트코인: P2P 전자 결제 시스템. Bitcoin.org.

156 우스만 W. 초한. (2017). 비트코인의 역사. UNSW 경영대학원.

157 유럽중앙은행. (2015년 5월 24일). 가상화폐 사기 - 추가 분석, 유럽중앙은행.

암호화폐는 그 특성상 전 세계적으로 널리 사용되고 있으며, 유니세프는 2019년부터 가상 형태로 기부금을 수령, 보관 및 분배하기 시작했습니다.158 동시에 암호화폐는 익명성을 갖추고 있어 악의적인 행위자들이 마약, 총기 및 폭발물을 불법 판매하고, 사람을 밀수하고, 자금을 세탁하고, 사이버 범죄를 용이하게 하는 매력적인 매체가 되었습니다.159

암호화폐는 통화의 한 형태로 사용되는 것 외에도 높은 시장 변동성으로 인해 거래되는 인기 있는 자산이 되었으며, 놀라울 정도로 짧은 기간에 암호화폐로 백만장자와 억만장자 세대를 만들어냈습니다.160 바로 이러한 특정한 맥락에서 AI는 암호화폐와 관련하여 주요 역할을 할 수 있으며, 이 가상 화폐를 둘러싼 시장 투기는 단순히 기부를 호소하는 데 그치지 않고 기금 모금을 위한 귀중한 수단이 됩니다.

예를 들어, "blackhatworld.com"과 같은 유명 언더그라운드 포럼에서는 암호화폐 거래 전용 AI 기반 봇의 개발 및 활용에 대한 논의가 있었습니다. 다른 머신 러닝 응용 프로그램과 마찬가지로, 이러한 시스템은 과거 데이터를 이용한 머신 러닝 시스템 학습에 의존하여 더욱 정확하고 정교한 예측을 통해 수익성 높은 암호화폐 거래를 이끌어냅니다.161 이러한 포럼에서는 수백 개의 암호화폐를 스캔하여 암호화폐 거래를 최적화하는 패턴을 찾거나, AI를 사용하여 암호화폐 거래 봇을 만드는 등 AI를 활용한 다른 형태의 활용 사례도 확인되었습니다.162, 163 거래 및 거래소 업계의 여러 그룹이 한동안 AI 주식 거래 봇을 개발해 왔지만, 큰 성과를 거두지는 못했습니다.164 하지만 많은 언더그라운드 블로그에서 이 주제를 언급하고 있다는 점을 고려할 때, 그럼에도 불구하고 언급할 가치가 있습니다. AI를 사용하여 암호화폐 공간을 조작하여 금전적 이익을 취하는 것 외에도, 테러리스트들은 AI를 이용하여 "핫 월렛"에서 암호화폐를 훔치거나165 블록체인에서 더 악명화된 거래를 촉진할 수 있습니다. 거래 관행의 익명성을 높이면 거래 행위가 노출되어 자금을 잃을 위험이 전반적으로 줄어듭니다.

테러 집단과 개인이 암호화폐를 체계적으로 사용하는 것은 아직 보이지 않지만,166 테러 조직이 암호화폐를 사용하는 것에 대한 우려가 커지고 있다.167 문서화된 사례는 제한적이지만 주목할 만한 사례는 몇 가지 있다.168 예를 들어, 250명 이상이 사망한 2019년 스리랑카 폭탄 테러를 조사한 사람들은 ISIL이 자금을 모으는 데 사용한 비트코인 지갑의 거래 수가 폭탄 테러 전에 눈에 띄게 증가한 것을 관찰했고, 이로 인해 이러한 비트코인이 공격 자금 조달에 역할을 했다는 믿음이 생겼다.169, 170 ISIL이 2015년에 저지른 파리 테러와 관련해서도 비슷한 의심이 존재하지만, 이러한 의심을 확인하는 증거는

158 유니세프. 블록체인. 유엔. https://www.unicef.org/innovation/blockchain 에서 확인 가능

159 인터폴. 다크넷과 암호화폐. 인터폴. https://www.interpol.int/en/How-we-work/Innovation/Dark- 에서 확인 가능 넷앤크립토커런시

160 파이낸셜 타임스. (2018년 3월 7일). 암호화폐 백만장자들의 부흥과 몰락. 파이낸셜 타임스. https://www. ft.com/content/d0df4322-1559-11e8-9c33-02f893d608c2

161 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philip Amann, Aglika Klein, Odhran McCarthy 및 Maria Eira. (2020년 11월 19일). 인공지능의 악의적 사용 및 남용. Trend Micro, EUROPOL, UNICRI. http://unicri.it/sites/default/ 에서 확인 가능 파일/2020-11/AI%20MLC.pdf

162 Janny. (2019년 8월 16일). 봇을 이용한 암호화폐 거래에 대한 종합적인 소개. Hackernoon. https://hackernoon에서 접속 가능. com/a-comprehensive-introduction-to-crypto-trading-with-bots-ti23930ly

163 브라이언 하트(2020년 6월 1일). AI 거래 플랫폼 티커론, 암호화폐 시장 예측 및 패턴 분석 기능 공개. PR Newswire. https://www.prnewswire.com/news-releases/ai-trading-platform-tickeron-unveils-cryptocurrency-market-forecast- ing-and-pattern-analysis-301068369.html 에서 확인 가능

164 Victor Hogrefe. (2018년 6월 7일). 트레이딩 봇은 실제로 얼마나 효과적일까요? Victor Hogrefe. https://victorhogrefe.medium.com/ 에서 확인 가능 트레이딩 봇은 실제로 얼마나 효과적인가?

165 잭 휘태커. (2021년 3월 16일). 핫 월렛 해킹 후 Roll 암호화폐 도난 사건으로 570만 달러 도난. TechCrunch. https://techcrunch에서 확인 가능. com/2021/03/16/5-7m-stolen-in-roll-crypto-heist-after-hot-wallet-hacked/

166 신시아 디온-슈바르츠, 데이비드 맨하임, 패트릭 B. 존스턴. (2020). 테러리스트의 암호화폐 사용: 기술적 및 조직적 장벽과 미래의 위협. 랜드 연구소

167 FATF. (2015). 새롭게 부상하는 테러 자금 조달 위험. FATF. 파리. www.fatf-gafi.org/publications/methodsandtrends/docu- 에서 확인 가능. ments/신흥 테러리스트 자금 조달 위험.html

168 수판 센터. (2020년 12월 10일). 수판 센터. IntelBrief: 테러리스트의 암호화폐 사용. https://thesoufan- 에서 확인 가능 center.org/intelbrief-2020-december-10/

169 로이 카치리. (2019년 5월 2일). 스리랑카 폭탄 테러 전날 ISIS에 대한 비트코인 기부가 급증했습니다. Globes. https://en.globes.co.il/ 에서 접속 가능 영어: en/article-exclusive-isis-funded-sri-lanka-bombings-with-bitcoin-donations-1001284276

170 체인애널리시스 팀. (2020년 5월 20일). 최근 암호화폐 테러 자금 조달 관련 보고서에 대한 팩트체킹. 체인애널리시스. https:// blog.chainalysis.com/reports/cryptocurrency-terrorism-financing-fact-check

picions가 부족합니다.171 2020년 초 주요 돌파구에서 미국 당국은 ISIL 및 알카에다와 연결된 계좌에서 100만 달러 이상의 암호화폐를 압수했습니다.172 금융활동기구(FATF)가 COVID-19 팬데믹 동안 불법 자금을 이동 및 은폐하기 위해 온라인 금융 서비스와 가상 자산을 오용하는 경우가 증가하고 있다는 점을 지적하고 이를 새로운 자금 세탁 및 테러 자금 조달 위험으로 묘사한 것도 주목할 만합니다.173

이러한 사태 전개에 비추어 볼 때, 테러 집단이 자금 조달 목적으로 AI 기반 암호화폐 거래를 활용할 가능성도 고려해야 합니다. 하지만 시장의 변동성으로 인해 이러한 전술의 매력이 전반적으로 감소할 가능성이 큽니다.

iv. 선전 및 허위 정보 확산

a. 딥페이크 및 기타 조작된 콘텐츠

딥페이크(deepfake)라는 용어는 GAN(Generic Angular Network)을 사용하여 조작되거나 생성된 가짜 오디오 및/또는 영상 콘텐츠를 지칭합니다. 딥페이크는 인간과 기계 모두 진짜와 가짜를 구분하기 어렵기 때문에 오늘날 AI의 가장 눈에 띄는 오용 사례 중 하나로 여겨지며 언론의 주목을 받고 있습니다.

딥페이크와 그 기반 기술은 오늘날의 허위 정보 전쟁에서 강력한 무기가 될 수 있습니다. 더욱이 인터넷, 소셜 미디어, 메시징 애플리케이션의 도달 범위와 속도와 결합되어 딥페이크는 극히 짧은 시간 안에 수백만 명의 사람들에게 빠르게 도달할 수 있습니다. 이러한 측면에서 딥페이크는 개인의 이미지와 신뢰성을 훼손하는 것, 성적 딥페이크를 포함한 온라인상에서 개인을 괴롭히거나 모욕하는 것, 협박, 갈취, 사기를 저지르는 것, 금융 시장을 교란하는 것, 사회 불안과 정치적 양극화를 조장하는 것 등 다양한 악의적이고 범죄적인 목적으로 사용될 수 있는 상당한 잠재력을 가지고 있습니다.

허위 정보를 유포하기 위한 딥페이크의 사용은 전통적으로 권위 있는 언론에 대한 사람들의 신뢰를 떨어뜨릴 가능성이 큽니다.
AI가 생성하는, 편견에 찬 텍스트, 가짜 영상, 그리고 수많은 음모론을 기반으로 하는 가짜 뉴스가 점점 더 넘쳐나는 현실에서, 사람들은 영상을 포함한 온라인 정보를 신뢰할 수 없다고 느낄 수 있으며, 이로 인해 "정보 종말" 또는 "현실 무관심"이라는 현상이 발생할 수 있습니다.174 가짜 콘텐츠, 허위 정보, 그리고 잘못된 정보가 증가함에 따라 세계는 "진실 붕괴"를 목격하고 있으며, 많은 행위자들이 이 과정에 관여하고 있지만, 테러리스트들은 이를 자신들의 목적을 위해 조장하고 악용할 가능성이 매우 높습니다. 실제로 허위 정보 자체가 딥페이크의 가장 큰 해악은 아니지만, 어떤 정보든 가짜일 수 있다는 생각 자체가 가장 큰 해악입니다. 더욱이 영상의 진위 여부를 확인하기 어렵다는 점 때문에, 모든 시청각 콘텐츠가 조작될 수 있기 때문에 어떤 정보든 조작될 수 있다는 사실이 부정될 수 있습니다. 결과적으로, 실제로 그러한 비디오 콘텐츠가 손상되지 않았더라도 가짜로 주장될 수 있으며, 개인이 자신의 행동에 대한 책임을 회피할 수 있습니다.

171 루이스 부에나벤투라. (2015년 12월 13일). ISIL은 실제로 파리 테러 자금 조달에 비트코인을 사용했을까? Medium. https://medium.com/cryptonight/did-isil-really-use-bitcoin-to-fund-the-paris-attacks-1287cea605e4 에서 접속 가능

172 앤디 그린버그. (2020년 8월 13일). ISIS가 코로나19 개인 보호 장비 사기 사이트를 운영했다는 의혹. Wired. https://www.wired.com/ 에서 확인 가능 스토리/IS가 코로나19 개인 보호 장비(PPP) 사기 사이트를 운영했다는 의혹/?utm_medium=social&utm_source=twitter&utm_brand=wired& utm_social-type=owned&mbid=social_twitter

FATF 173호. (2020년 5월). COVID-19 관련 자금세탁 및 테러 자금 조달 - 위험 및 정책 대응. FATF. https:// 에서 확인 가능 www.fatf-gafi.org/media/fatf/documents/COVID-19-AML-CFT.pdf

174 제니퍼 카바나, 마이클 D. 리치 (2018). 진실의 쇠퇴: 정책 결정과 민주주의에 대한 위협. 랜드 연구소. https://www.rand.org/pubs/research_briefs/RB10002.html 에서 열람 가능

신경망을 이용한 얼굴 바꾸기 - 출처: 위키피디아 커먼즈

현재 딥페이크는 여성 연예인의 얼굴과 포르노 배우의 몸을 합성하는 등 포르노 콘텐츠를 제작하는 데 압도적으로 많이 사용되고 있습니다.175 그럼에도 불구하고, 딥페이크는 민주주의와 국가 안보에 심각한 영향을 미칠 수 있는 잠재력을 가지고 있습니다. 소셜 미디어가 대중의 주요 정보원 중 하나가 되었다는 점을 고려할 때, 딥페이크는 허위 정보 확산 측면에서 심각한 위협이 됩니다. 정보가 빠르게 소비되고 재생산되는 반면, 사용자들은 콘텐츠의 진위 여부를 확인하는 데 거의 시간을 들이지 않기 때문 입니다.176,177 딥페이크 영상은 일반적으로 온라인에서 수명이 매우 짧지만, 특히 바이럴될 경우 순간적인 공황과 혼란을 야기할 수 있습니다. 실제로 사람들이 가짜 콘텐츠를 구별하지 못하고 영상이 딥페이크인지 아닌지에 대한 혼란을 야기하는 것만으로도 심각한 문제를 야기할 수 있습니다.

딥페이크의 부정적 영향을 고려하면 테러 집단이나 개인이 딥페이크 기술을 이용해 소셜 미디어에서 허위 정보 캠페인을 벌여 여론을 조작하거나 국가 기관에 대한 국민의 신뢰를 훼손하려 할 가능성이 있습니다.178

이러한 기술은 선전, 급진화 또는 행동 촉구를 위한 효과적인 도구로도 활용될 수 있습니다. 예를 들어, 특정 정치인이 특정 집단에 대해 모욕적인 발언을 하는 "딥페이크" 콘텐츠를 제작하여 해당 집단 내부의 분노를 고조시키고 동조자 수를 늘리는 것을 통해 이러한 목적을 달성할 수 있습니다.

AI는 오디오-비주얼 딥페이크를 제작하는 것 외에도 맞춤형 급진화 이야기를 생성하는 데 사용될 수도 있습니다.
OpenAI의 널리 알려진 GPT-3,179를 포함한 NLP의 새로운 고급 기술은 마이크로 프로파일링 및 마이크로 타겟팅, 모집 목적을 위한 자동 텍스트 생성 또는 ISIL의 주장과 같은 맞춤형 가짜 뉴스 및 테러 관련 음모론 확산에 이 기술을 사용할 가능성에 대한 우려를 불러일으켰습니다.

175 조르지오 파트리니. (2019년 10월 7일). 딥페이크 환경 매핑. DeepTrace.

176 마리-헬렌 마라스(Marie-Helen Maras)와 알렉스 알렉산드루(Alex Alexandrou). 인공지능 시대와 딥페이크 영상의 여파. 국제 증거 및 증명 저널, 23(3): 255–262.

177 다니엘 토마스. (2020년 1월 23일). 딥페이크: 민주주의에 대한 위협인가, 아니면 그저 재미일까? BBC. https://www.bbc.com/news/ 에서 확인 가능 사업-51204954

178 미카 웨스터룬드. (2019). 딥페이크 기술의 등장: 서평, 기술혁신경영리뷰, 제9/11권.

179 OpenAI. (2021년 3월 25일). GPT-3는 차세대 앱을 지원합니다. OpenAI. https://openai.com/blog/gpt-3-apps/ 에서 확인 가능

그리고 알카에다는 COVID-19 팬데믹이 "서방에 대한 신의 분노"라고 말했습니다.180 제목에 따라 기사를 공유하거나 문제의 웹사이트에 대한 실질적인 실사를 하지 않고 기사를 훑어보는 온라인 독자 추세가 증가하고 있는 상황에서 AI 기반 가짜 뉴스 미디어 사이트를 사용하면 해로울 수 있습니다.181 따라서 테러 조직이 언젠가 실제 뉴스 헤드라인을 자동으로 읽고 잘린 가짜 메시지를 만들어 소셜 미디어 및 기타 채널을 통해 퍼뜨려 자신들의 목적을 달성할 수 있는 AI 시스템을 보급할 가능성이 여전히 남아 있습니다.

Pixabay의 memyselfaneye가 찍은 사진

마지막으로, AI는 쉽게 모집되거나 급진화될 수 있는 남녀를 찾아내 테러 관련 콘텐츠나 메시지를 특정 대상에게 배포하는 데 활용될 수 있습니다. 이 경우, 테러리스트들은 AI를 "알고리즘 증폭기" 및 "추천 도구"로 활용하여 선전을 유포할 수 있습니다. 예를 들어, 온라인에서 폭력적인 콘텐츠를 반복적으로 검색하거나 소외되고 분노한 반영웅을 묘사한 영화를 스트리밍한 사람들에게 특정 메시지를 전달하는 것입니다.

v. 기타 작전 전술

에이. 감시

머신 러닝의 발전으로 인해 디지털 이미지나 비디오에서 정보를 수집, 처리, 분석 및 추출하는 방법인 컴퓨터 비전 분야의 중요한 발전은 AI 분야의 최근 주요 발전 중 하나일 수 있습니다. 딥 러닝은 이미지 및 비디오 처리, 특히 객체 인식에 혁명을 일으켜 기계가 얼굴을 감지하고 인식하며, 얼굴 인식을 수행할 수 있도록 했습니다.

180 UNICRI. (2020년 11월). 허위 정보 바이러스를 막아라. 유엔. 다음에서 접속 가능 http://www.unicri.it/sites/default/files/2020-11/SM%20misuse.pdf

181 케이틀린 듀이. (2016년 6월 16일). 새롭고 우울한 연구 에 따르면 , 여러분 중 10명 중 6명은 이 링크를 읽지 않고 공유할 것이라고 합니다. 워싱턴 포스트. https://www.washingtonpost.com/news/the-intersect/wp/2016/06/16/six-in-10-of-you-will-share-this-link-without-read-ing-it-according-to-a-new-and-depressing-study/ 에서 확인 가능

표현.182 얼굴 인식이라는 격렬하게 논쟁의 여지가 있는 영역을 넘어,183 컴퓨터 비전의 발전은 인체 감지, 개인 식별, 속성 인식, 인간 행동 인식 및 신체 움직임(보행) 인식에서도 개선을 가능하게 했습니다.184 딥 러닝은 또한 차량 식별 및 재식별, 번호판 인식 등을 포함하여 객체 감지, 인식 및 추적을 개선했습니다.185 동시에 이러한 발전으로 인해 오식별 오류율이 상당히 감소했습니다.186

폐쇄회로 텔레비전(CCTV), 바디캠, 순찰 드론 등 다양한 감시 기술을 오랫동안 활용해 온 법 집행 기관은 피해자, 가해자 또는 기타 용의자의 신원을 파악하는 데 딥러닝 기반 컴퓨터 비전의 잠재력을 빠르게 인식해 왔습니다. 최근 몇 년 동안 AI 기반 감시 기술에 대한 법 집행 기관의 관심이 크게 증가했습니다. 카네기 국제평화재단이 편찬한 AI 글로벌 감시 지수에 따르면 분석 대상 176개국 중 75개국이 스마트 시티/안전 도시 플랫폼, 얼굴 인식 시스템, 스마트 경찰 활동 등을 포함해 감시 목적으로 AI 기술을 적극적으로 사용하고 있어 전 세계적으로 AI 감시 도입이 빠르게 증가하고 있음을 보여줍니다.187, 188 COVID-19 팬데믹도 AI 기반 감시 기술에 대한 관심을 높이는 데 중요한 역할을 했습니다. 여러 국가 당국이 디지털 접촉 추적 노력을 지원하거나 격리 조치 시행을 용이하게 하는 데 이 기술이 지닌 논란의 여지가 있는 잠재력을 보여주었습니다.189

Unsplash의 Dmitry Ratushny 사진

182 Shan Li 및 Weihong Deng. (2018). 심층 얼굴 표정 인식: 개괄. IEEE 감성 컴퓨팅 저널. PP. 10.1109/TAFFC.2020.2981446.

183 유엔 인권 고등판무관 사무소. (2020년 6월 25일). 바첼레트, "새로운 기술은 평화적 시위의 권리를 방해하는 것이 아니라, 그 권리를 증진하는 데 기여해야 한다"고 각국에 촉구. 유엔. https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx- 에서 확인 가능 ?뉴스ID=25996&컬러ID=E

184 Honghua Xu, Li Li, Ming Fang 및 Fengrong Zhang. (2018). 머신러닝을 기반으로 한 움직임 인간 행동 인식. 국제적인 온라인 엔지니어링 저널(iJOE). 14. 193. 10.3991/ijoe.v14i04.8513.

185 Jianzong Wang, Xinhui Liu, Aozhi Liu, Jing Xiao. (2019). 자연 장면에서 차량 번호판 인식을 위한 딥러닝 기반 방법. APSIPA 신호 및 정보 처리 논문집, 8, E16. doi:10.1017/ATSIP.2019.8.

186 더글러스 헤븐. (2019년 10월 9일). 딥러닝 AI가 속기 쉬운 이유 인공 지능 연구자들은 딥러닝의 결함을 수정하려고 노력하고 있습니다. 신경망. Nature. https://www.nature.com/articles/d41586-019-03013-5 에서 확인 가능

187 스티븐 펠드스타인. (2019년 9월). AI 감시의 세계적 확장. 카네기. https://carnegieendowment.org/files/ 에서 열람 가능 WP-펠트슈타인-AISurveillance_final1.pdf

188 저자가 그의 평가에서 반복적으로 관찰한 것처럼 AI 기반 감시 기술의 사용은 반드시 이러한 시스템이 인권을 침해하는 방식으로 남용되거나 사용되고 있다는 것을 의미합니다.

189 Yann Sweeney. (2020). COVID-19 감시 도구에 대한 논쟁 추적. Nature Machine Intelligence 2, 301–304. https://doi.org/10.1038/s42256-020-0194-1

그러나 이러한 활용 사례를 뒤집으면 AI 기반 감시 기술의 악의적 사용 가능성이 제기됩니다. 특히 대규모 테러 공격의 경우, 계획 및 준비를 위해 장기간의 감시가 필요한 경우가 많습니다. 테러 단체는 목표물을 식별 및 탐지하고, 공격 적합성을 판단하며, 공격을 용이하게 할 수 있는 취약점을 파악하기 위해 장소와 사람을 모두 감시합니다. 전통적으로 이러한 감시는 도보, 주차된 차량, 또는 소셜 네트워크를 통해 온라인으로 수행되며, 몇 주, 몇 달, 심지어 몇 년에 걸쳐 진행될 수 있습니다. AI 기반 감시 기능의 발전은 이론적으로 감시의 시간 소모적인 측면을 상당 부분 줄일 수 있습니다. 기술의 도움을 받아 테러리스트는 장소를 감시하고 사람들의 움직임을 추적하며, 표적 개인과 자산을 식별하고, 목표 지점의 물리적 보안 조치를 자동 및 원격으로 평가할 수 있습니다.

b. 소셜 네트워킹 플랫폼에서의 가짜 온라인 신원 및 사칭

인터넷은 본질적으로 사용자에게 일정 수준의 익명성을 제공합니다. 이는 온라인 트롤링, 사이버 괴롭힘, 아동 그루밍 및 성착취 등 인터넷의 가장 악랄하고 유해한 사용을 가능하게 하는 요인 중 하나였습니다. 페이스북에는 미리 결정되고 훼손된 콘텐츠를 온라인에 적극적으로 유포하는 데 사용되는 가짜 계정이 7억 5천만 개가 넘는 것으로 추산됩니다. 페이스북은 이러한 문제의 심각성을 인지하고 2019년에만 29억 개의 가짜 계정을 삭제했다고 밝혔습니다.190

유엔 사무총장 안토니우 구테흐스는 2019년 9월 소셜 미디어와 다크 웹을 이용하여 공격을 조직하고, 선전을 유포하고, 새로운 추종자를 모집하는 것이 사이버 테러리즘의 새로운 전선이 되었다고 지적했습니다. 테러 조직의 주요 모집 대상은 17세에서 27세 사이의 젊은이들이기 때문에,191 이러한 연령대에서 소셜 미디어 플랫폼의 인기가 높아지면서 테러 조직에게 소셜 미디어의 효과적인 활용은 더욱 중요해지고 있습니다. 과거 ISIL의 소셜 미디어 플랫폼 사용은 분쟁 지역으로 이동하는 외국인 테러 전투원의 수를 전례 없이 증가시켰습니다. 더욱이 소셜 미디어 플랫폼의 악용을 통해 테러 조직은 공격을 수행하고, 잠재적인 모집자를 파악하고, 선전을 전파하고, 훈련 자료를 배포하고, 불법 거래에 참여하고, 자금을 조달할 수 있게 되었습니다.192 테러 조직이 이미 이처럼 효율적으로 소셜 미디어 플랫폼을 활용하고 있기 때문에, AI를 적용하면 성공률이 더욱 높아질 수밖에 없습니다.

실제로 AI의 발전은 이러한 현상에 완전히 새로운 차원을 가져올 것으로 기대됩니다. 딥페이크의 기반 기술인 GAN은 매우 사실적인 가짜 얼굴 이미지를 합성하는 데 특히 효과적입니다. 예를 들어, "ThisPersonDoesNotExist.com" 웹사이트는 GAN을 사용하여 페이지에 접속하거나 새로 고침할 때마다 완전히 조작된 새로운 얼굴 이미지를 생성합니다.

이러한 기술의 악의적 사용 가능성은 이미 확인되었습니다. 2019년에는 케이티 존스라는 이름으로 링크드인 계정에 AI가 생성한 프로필 사진이 사용되었습니다. 30대 젊은 전문직 종사자인 케이티는 워싱턴에 본사를 둔 싱크탱크에서 근무하며 여러 미국 정부 관계자들과 관계를 맺고 있는 것으로 나타났습니다. 케이티의 프로필을 검토한 전문가들은 해당 프로필이 허위라고 판단하고, 이 가짜 프로필은 정보 수집 작전의 일환으로 요주의 인물들을 유인하여 정보를 수집하려는 시도였을 가능성이 높다고 결론지었습니다.193

또한 범죄 포럼에서는 소셜 미디어 플랫폼을 위한 AI 기반 가짜 계정 및 채팅봇 생성이 증가하고 있습니다.194 연구에 따르면 이러한 가짜 계정 및 봇은 점점 더 정교해지고 있으며

190 Karen Hao. (2020년 3월 4일). 페이스북이 머신러닝을 활용하여 가짜 계정을 탐지하는 방법. MIT Technology Review. https:// 에서 확인 가능 www.technologyreview.com/2020/03/04/905551/facebook이 머신러닝을 이용해 가짜 계정을 감지하는 방법/

191 안토니우 구테흐스. (2019년 9월 25일). 사무총장, 소셜 미디어와 다크웹을 이용한 사이버 테러를 안전보장이사회 장관급 토론의 "새로운 지평"으로 규정. 유엔. https://www.un.org/press/en/2019/sgsm19768.doc.htm 에서 열람 가능

192 유엔. (2016년 12월 1일). 개념 노트: 회원국 및 관련 기관과 함께하는 대테러위원회 특별 회의 국제 및 지역 기구, 시민사회, 그리고 민간 부문이 참여하는 "인권과 기본적 자유를 존중하면서 테러 목적의 정보 및 통신 기술 악용 방지"에 관한 공동 연구. https://www.un.org/sc/ctc/wp-content/uploads/2016/11/Concept-Note.pdf

193 라파엘 새터. (2019년 6월 13일). 전문가: 스파이, AI가 생성한 얼굴 인식을 이용해 표적과 연결. ABC 뉴스. https://abcnews. 에서 확인 가능. go.com/Technology/wireStory/experts-spy-ai-generated-face-connect-targets-63674174

194 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philip Amann, Aglika Klein, Odhran McCarthy 및 Maria Eira. (2020년 11월 19일). 인공지능의 악의적 사용 및 남용. Trend Micro, EUROPOL, UNICRI. http://unicri.it/sites/default/ 에서 확인 가능 파일/2020-11/AI%20MLC.pdf

해당 소셜 미디어 플랫폼의 평균 사용자는 이를 통해 사용자의 신고나 플랫폼에서의 감지 및 금지를 피할 수 있습니다.195 이러한 가짜 계정과 봇은 특정 페이지나 콘텐츠의 조회수, 팔로워 또는 "좋아요"를 늘리는 등 여러 가지 목적으로 사용됩니다.

테러 조직은 이러한 AI 기반 기술을 활용하여 소셜 미디어 플랫폼 사용의 효율성을 높일 수 있습니다. AI가 생성한 가짜 계정과 일반 사용자를 사칭하는 봇은 테러 조직이 소셜 미디어 플랫폼에서 더욱 쉽게 메시지를 전파하고, 실제 사용자가 차단될 위험을 줄이며, 원하는 정보를 얻거나 급진화를 지원하기 위한 사회 공학적 노력을 촉진하는 데 도움이 될 수 있습니다.

c. 변형된 여권

현대 테러 집단에게 여행 서류를 부적절하게 취득, 변경 또는 위조하는 것은 필수적입니다. 위조된 서류는 테러리스트들이 국제 여행을 용이하게 하기 위해 흔히 사용하는 수단입니다. 9/11 테러범들 중 몇몇의 사례가 그 예입니다.196 또한 다른 행정적 목적을 위한 신원 증명에도 자주 사용됩니다. 예를 들어, 2015년 파리 테러의 배후 세력은 공격에 앞서 대출을 받기 위해 위조 여권을 사용했습니다.197

위조 여권을 테러 목적으로 사용하는 것은 너무 널리 퍼져 있어서 조사관들은 알카에다와 같은 테러 조직이 다양한 국가에 자체 전문 구성원을 두고 있으며, 그들의 유일한 임무는 조직의 다른 구성원에게 요청 시 여권 및 기타 관련 문서를 제공하는 것이라고 믿었습니다.198 최근에는 ISIL이 위조 여권 생산을 "산업화"한 것으로 간주됩니다.199

Unsplash의 CardMapr.nl이 찍은 사진

위조 여권을 만드는 위험하고 새로운 AI 기반 방법이 곧 등장할 것으로 예상되는데, 이는 "변형된" 여권이라고 할 수 있습니다. MorGAN(생성적 적대 신경망을 통한 변형) 방법을 사용하면 범죄자들이 여러 사람과 일치하는 여권 사진을 만들 수 있다는 것이 밝혀졌습니다. 200

195 위와 같음.

196 미국 테러 공격에 관한 국가위원회(2004). 9/11 테러범들의 미국 입국: 직원 성명서 1호. 미국 정부. https://govinfo.library.unt.edu/911/staff_statements/staff_statement_1.pdf 에서 열람 가능.

197 유로폴. (차) 행정 문서 위조 및 불법 거래. 유로폴. https://www.europol.europa.eu/ 에서 확인 가능 범죄 지역 및 추세/범죄 지역/행정 문서 위조 및 인신매매

198 오리아나 질. 국경 횡단: 테러리스트들이 위조 여권, 비자 및 기타 신분증을 사용하는 방법. PBS. https://www. 에서 접속 가능. pbs.org/wgbh/pages/frontline/shows/trail/etc/fake.html

199 브라이언 로스, 미셸 맥피, 리 페란. (2016년 1월 25일). ISIS, 위조 여권 "산업" 장악. 관계자 발언. ABC. 다음에서 확인 가능 https://abcnews.go.com/International/isis-fake-passport-industry-official/story?id=36505984

200 나세르 다머 박사(차). 얼굴 변형: 새로운 위협인가? 프라운호퍼 IGD. https://www.igd.fraunhofer.de/en/press/annu- 에서 확인 가능 al-reports/2018/얼굴 변형-새로운 위협

즉, 변형 여권은 두 명 이상의 개인이 사용할 수 있는 단일 여권을 말합니다.201 이를 악용하여 악의적인 행위자는 인간 기반 인식 시스템과 기계 기반 인식 시스템을 모두 속여 전통적으로 강력한 국경 보안을 쉽게 무너뜨릴 수 있습니다.202 변형 여권을 둘러싼 일부 보안 문제를 해결하기 위해 일부 국가의 당국은 개인이 자신의 사진(아마도 가짜)을 제공하는 대신 여권 사무소에서 여권 사진을 찍도록 의무화하는 등 추가 조치를 채택하기 시작했습니다.203, 204

이러한 추가 조치가 충분한지는 아직 알 수 없지만, 확인되지 않은 변형된 여권으로 인해 테러리스트가 국경 통제와 공항의 보안 검사를 통과하거나 심지어 공공장소에서도 감지되지 않고 이동할 수 있는 능력이 크게 향상될 수 있다는 것은 분명한 사실입니다.

디. 온라인 소셜 엔지니어링

사회 공학은 인간의 상호작용을 통해 취약점을 악용하는, 흔히 조작을 수반하는 잘 확립된 공격 벡터입니다. 범죄자와 기타 악의적인 행위자들이 금전이나 기밀 정보를 획득하거나 피해자들이 평소에는 하지 않을 행동을 하도록 설득하기 위해 사기 행각을 벌이는 데 자주 사용됩니다 .205 이 공격 벡터는 주로 소셜 미디어를 통한 온라인이나 직접 대면하는 오프라인 모두에 적용될 수 있습니다.

챗봇은 현대 사회에서 AI의 가장 가시적이고 명확한 활용 사례 중 하나입니다. AI가 계속 발전함에 따라, 봇은 소셜 엔지니어링 기법을 포함한 온라인 사기에 점점 더 큰 역할을 할 것으로 예상됩니다.

말할 필요도 없이 테러 조직들은 온라인에서 사회 공학 전술을 악용하는데, 주로 새로운 구성원과 동조자를 파악하고 모집하는 데 활용합니다. 실제로 이러한 조직들은 이미 봇 계정 사용에 상당한 경험을 가지고 있습니다.206 2015년 파리 테러 이후, 핵티비스트 단체 어나니머스(Anonymous)는 ISIL을 상대로 온라인 캠페인을 시작하여 최대 25,000개의 ISIL 봇을 온라인에서 제거했다고 주장했습니다.207

현재 챗봇은 전자상거래 지원이나 고객 서비스와 같이 반복적인 요소가 있는 매우 좁은 맥락에서 탁월한 성과를 보이고 있습니다. 자연어 처리(NLP) 기술이 발전함에 따라 챗봇은 인간과의 상호작용을 기반으로 시간이 지남에 따라 학습하여 인간과 더욱 유사한 방식으로 응답할 수 있습니다. 챗봇과 인간을 구분하는 능력이 더욱 어려워짐에 따라, 소셜 엔지니어링 공격에 챗봇을 활용할 가능성이 커지고 있습니다.208

동시에, 사회 공학 전술의 성공은 그 타당성에 달려 있으며, 이와 관련하여 대상에 대한 상세하고 정확한 정보를 얻는 것이 필수적인 역할을 합니다. AI도 여기서 역할을 할 수 있습니다. 예를 들어, 얼굴 인식 알고리즘을 사용하는 새로운 AI 기반 계정 탐지 도구가 온라인 포럼에서 연구되고 있으며, 이를 통해 사용자는 프로필 사진이 동일하지 않더라도 서로 다른 소셜 미디어 플랫폼에서 동일한 사람의 여러 계정을 일치시킬 수 있습니다. 이 기술을 사용하면 악의적인 행위자가 대상의 여러 소셜 미디어 프로필을 빠르게 식별할 수 있습니다.209 이러한 프로필을 분석함으로써 악의적인 행위자는 문제의 개인에 대해 더욱 완벽하게 이해하고 문제의 개인을 더 잘 조종하는 데 필요한 정보를 수집할 수 있으며, 예를 들어 속임수나 강압을 통해 기밀 정보를 공유하도록 할 수 있습니다.

201 데이비드 J. 로버트슨, 앤드류 멍걸, 데릭 G. 왓슨, 킴벌리 A. 웨이드, 소피 J. 나이팅게일, 스티븐 버틀러. (2018). 변형된 여권 사진 감지: 훈련 및 개인차 접근법. 인지 연구: 원리와 함의, V. 3/1.

202 데이비드 J. 로버트슨, 앤드류 멍걸, 데릭 G. 왓슨, 킴벌리 A. 웨이드, 소피 J. 나이팅게일, 스티븐 버틀러. (2018). 변형된 여권 사진 감지: 훈련 및 개인차 접근법. 인지 연구: 원리와 함의.

로이터 통신 직원 203명. (2020년 6월 3일). 독일, 디지털 도플갱어 여권 사진 금지. 로이터. https://www.reuters.com/ 에서 확인 가능 기사/미국-독일-기술-모핑/독일-디지털-도플갱어-여권-사진-ID-USKBN23A1YM

204 루아나 파스쿠. (2020년 6월 17일). 독일, 국경 검문소 생체 정보 위조 방지 위해 여권 사진 변형 금지. 생체 정보 업데이트. https://www.biometricupdate.com/202006/germany-bans-passport-photo-morphing-to-prevent-biometric-spoofs-at-bor- der-checks 에서 접근 가능

205 인터폴. 사회 공학 사기. 인터폴. https://www.interpol.int/en/Crimes/Financial-crime/Social-engineer- 에서 확인 가능 ing-scams

206 스티븐 스탈린스키 & R. 소스노. (2020년 8월 5일). 암호화된 메시징 플랫폼 텔레그램에서의 지하디스트 봇 사용. Memri. 다음에서 접근 가능 https://www.memri.org/reports/jihadi-use-bots-encrypted-messaging-platform-telegram

207 리나 가필드. (2015년 12월 14일). ISIS는 수천 개의 정치 봇을 만들었고, 핵티비스트들은 당신이 그것들을 파괴하기를 원합니다. 비즈니스 인사이더- https://www.businessinsider.com/anonymous-battles-isis-political-bots-2015-12 에서 확인 가능합니다.

208 Simon Chandler. (2018년 12월 21일). 사악한 챗봇의 진화가 코앞으로 다가왔습니다. Daily Dot. https://www.dailydot에서 확인 가능. com/debug/evil-chatbot-hackers-ai/

209 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philip Amann, Aglika Klein, Odhran McCarthy 및 Maria Eira. (2020년 11월 19일). 인공지능의 악의적 사용 및 남용. Trend Micro, EUROPOL, UNICRI. http://unicri.it/sites/default/ 에서 확인 가능 파일/2020-11/AI%20MLC.pdf

VII. AI의 테러적 활용에 대한 분석

AI의 악의적 사용은 매우 새롭고 아직 완전히 이해되지 않은 영역입니다. 테러 목적으로 AI를 사용할 경우 발생할 수 있는 잠재적 영향 또한 아직 미지수입니다. 따라서 법 집행 기관, 보안군, 대테러 기관뿐 아니라 정책 입안자, 업계, 학계 모두 AI의 악의적 사용을 받아들이는 데 어려움을 겪을 수 있습니다. 결과적으로 위험에 대한 철저하고 정보에 기반한 평가를 수행하기가 어려울 수 있습니다.

이러한 어려움을 극복하기 위해 일련의 가상 시나리오가 개발되었습니다. 이 시나리오들은 독자들이 AI가 가상 테러 집단의 행동 방식에 어떻게 통합될 수 있는지 시각화하는 데 도움을 주기 위해 아래에 설명되어 있습니다. UNCCT 참가자들이 인식하는 이러한 시나리오의 가능성은 다음과 같습니다.

UNOCT-UNICRI 전문가 그룹 회의는 각 시나리오 이후에 추가로 제공됩니다.

시나리오 1:

팬데믹 문제 - 란티움 정부와 백신, 복잡한 사이버 공격으로 치명타 입어

란티움 공화국의 경기 침체 이후, 테러 단체가 두각을 나타냈습니다. "리디머스(The Redeemers)"로 알려진 이 단체는 정부와 기관을 약화시키고 전복하기 위해 여러 차례 사이버 및 물리적 공격을 감행했습니다. 이 단체는 또한 지역 전역의 여러 유사 단체들과 연계되어 있으며, 공동의 목표를 달성하기 위해 광범위한 사이버 역량을 제공하는 것으로 알려져 있습니다. "더 허브(The Hub)"라는 종단간 암호화 기능을 갖춘 소셜 미디어 애플리케이션을 개발하여, 단체 간의 정보 교환과 대중에 대한 선전 활동을 통해 리디머스의 호소력을 강화하고 새로운 구성원을 모집하는 데 활용하고 있습니다.

현재 전 세계적인 팬데믹의 한가운데에 있는 란티움은 발병에서 회복하고 시민들을 바이러스로부터 안전하게 지키기 위한 백신 접종 프로그램을 시작하는 데 어려움을 겪고 있습니다. 팬데믹 시작 이후 인구의 3% 이상이 감염되었으며, 사망률은 대부분 국가에서 기록된 것보다 3.4% 높습니다. 백신 접종 프로그램 초기 단계에 있는 란티움은 퓨처팜 백신을 주당 4만 회분 확보했습니다. 이 어려운 시기에 모든 시선이 정부의 백신 접종 프로그램 시행에 쏠려 있습니다. 이를 인지한 리디머즈는 이러한 상황을 악용하여 정부의 신뢰도에 치명적인 타격을 가할 새로운 공격을 계획하기 시작합니다.

광범위한 동조자 네트워크를 활용하여, 리디머스는 경찰청 생체정보팀이 운영하는 란티움의 CCTV 감시 시스템에 접근하는 데 성공했습니다. 시스템에 접근하여 란티움 국립병원 직원들의 움직임을 모니터링한 지 몇 시간 만에, 리디머스는 병원 행정동을 드나드는 직원들의 움직임을 바탕으로 병원의 주요 직원들을 식별했습니다. CCTV 카메라 녹화 영상의 얼굴 스크린샷을 이용하여 안면 인식 소프트웨어 프로그램을 사용하여 신원을 확인할 수 있었습니다.

확인된 개인의 이름을 전문가 네트워킹 사이트 "con-nected.com"의 사용자 프로필과 교차 참조하여 Redeemers는 병원에서 적절한 대상을 선택합니다. 그 대상은 병원의 백신 프로그램 출시를 담당하는 수석 프로그램 책임자 중 한 명인 Alison Apple 여사입니다.

다음 날 밤, 리디머스는 애플 씨를 공격의 진입점으로 삼아 광범위한 사이버 공격을 개시합니다. 그들은 다크 웹에서 얻은 도난된 비밀번호 데이터베이스를 활용하여 훈련된 신경망을 사용하여 애플 씨를 대상으로 AI 비밀번호 추측 공격을 시작합니다. 신경망이 작동하면, 고품질 비밀번호 추측을 생성하여 궁극적으로 병원 네트워크에서 애플 씨의 공식 계정에 접근합니다.

리디머들은 병원에 도착하자마자 백신 프로그램 관련 문서를 수색합니다. 곧 이 민감한 문서들이 무결성 보호를 위해 암호화되어 있음을 알게 됩니다. 이에 대응하여 리디머들은 AI 기반 복호화 도구를 사용하여 문서의 보안 기능을 해제하고, 병원 내 백신의 주요 저장 허브와 백신 보관 프로토콜 및 시스템 정보를 포함한 백신 프로그램과 관련된 필수 데이터와 정보를 공개합니다. 병원 네트워크에 접근한 리디머들은 백신의 냉장 시스템을 공격하여 냉동고 온도를 권장 온도인 영하 40도 이상으로 5시간 동안 밤새도록 높입니다. 백신의 무결성을 유지하는 데 필요한 온도 범위를 벗어나면 활성 성분이 중화되고, 백신의 효능은 외관 변화 없이 95%에서 5%로 감소합니다. 리디머들은 뛰어난 컴퓨터 과학 지식을 바탕으로 온도 센서와 병원 직원에게 온도 변화를 알리는 알림 시스템을 비활성화합니다. 이 작업을 마친 리디머들은 시스템을 종료하기 전에 기록을 삭제하여 어떠한 개입의 흔적도 남기지 않습니다. 다음 날 아침 병원 직원들이 업무에 복귀할 무렵, 모든 냉동고는 권장 보관 온도로 다시 작동하고 있었습니다. 구세주의 개입을 알지 못한 채, 병원 직원들은 백신을 계속 공급했습니다.

백신 접종이 몇 주 동안 이어진 후, 주요 언론들은 백신을 완전히 접종했음에도 불구하고 바이러스에 감염되고 심지어 사망하는 사례가 증가하고 있다는 보도를 쏟아내기 시작했습니다. 이러한 보도에 대한 광적인 반응과 백신에 대한 우려가 커지면서, 리디머스(The Redeemers)는 정부의 바이러스 대응을 확고히 지지하고 공개적으로 대변했던 크리스티나 카발레로 총리의 허브(The Hub)에 딥페이크 오디오 클립을 공개했습니다. 이 가짜 통화에서 카발레로 총리는 내각 팀원들에게 백신이 효과가 없으며, 이는 정부의 국민 통제를 강화하기 위해 시민들의 유전 물질을 수집하려는 교묘한 계략의 일부였다고 주장했습니다. 이 딥페이크는 리디머스가 스마트폰용 딥페이크 앱과 총리의 공식 발언 오디오 샘플을 사용하여 제작했습니다. 카발레로 총리는 즉시 해당 오디오 클립의 진위 여부를 부인했지만, 이 클립은 빠르게 퍼져나가며 대중의 분노를 촉발했고 결국 사임하게 되었습니다.

부총리가 그녀의 자리를 대신하지만, 정부에 대한 신뢰는 사상 최저 수준으로 떨어졌고, 카바예로 씨의 강력한 리더십 없이 정부는 내부 갈등으로 서서히 무너지기 시작하면서 국민들은 총선을 요구하게 되었습니다. 동시에 시장 분석가들은 오디오 클립 공개 이후 "분당 수백 명의 사용자"가 허브에 가입했다고 보고했습니다.

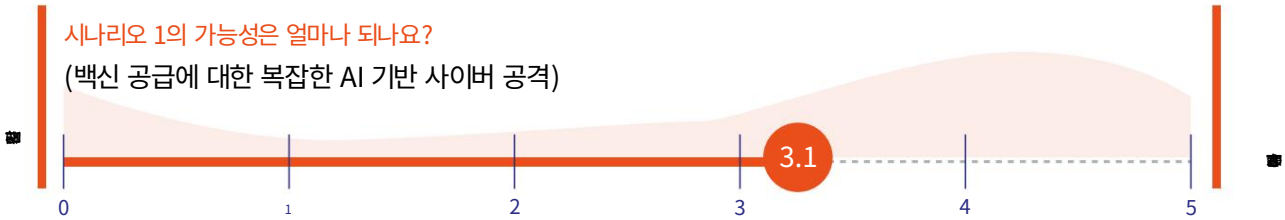


그림 4: 전문가 그룹 회의 참석자들은 시나리오 1의 가능성을 인식했습니다.

시나리오 2:

숨을 곳 없음 - 드론 떼가 유명 활동가를 죽이다

뉴 바인랜드의 화창한 오후, 다섯 대의 드론이 일제히 비행하며 인파로 북적이는 도심 광장을 휙 지나가자 사람들은 혼란에 빠졌습니다. 불과 몇 미터 떨어진 곳에서 이 기이한 광경을 보려는 군중의 관심은 작은 폭발이 갑자기 광장을 뒤흔들면서 순식간에 공황과 혼란으로 변했습니다. 현장으로 출동한 경찰은 폭발로 한 명이 사망하고, 폭발 현장 근처에 있던 다른 여러 명이 중상을 입은 것을 발견했습니다. 사망자는 나중에 벤저민 브라운 씨로 확인되었습니다.

브라운 씨는 지역 사회에서 거침없는 목소리를 내는 구성원이었으며, 지역 언론의 논평과 "선택받은 자(the Chosen)"로 알려진 테러 집단의 이념을 깎아내리는 강렬하고 통렬한 발언을 쏟아내 언론의 상당한 주목을 받기 시작했습니다. 시간이 흐르면서 그의 발언에 담긴 열정과 힘은 그를 국제적으로 인정받는 인물로 만들었고, 최근에는 TED 강연에서 선택받은 자들의 수법과 동기를 다시 한번 강력하게 비판하기도 했습니다. 공격 당일 아침, 브라운 씨는 센트럴 스퀘어에 있는 한 지역 서점에서 그의 최신 저서 "거짓말쟁이와 사소한 범죄자로만 남도록 선택받은 자(Chosen to be nothing more than liears and petty criminals)" 출간 관련 행사에 참석하여 강연을 하고 있었습니다.

'초즌'은 해당 단체와 연계된 소셜 미디어 채널에 공유된 영상 메시지를 통해 이번 공격의 배후를 재빨리 자처했습니다. 영상에서 해당 단체는 브라운 씨의 죽음을 애도하며 "복수는 그들이 숨는 곳마다 찾아올 것"이라고 과시했습니다.

공격에 대한 조사가 진행됨에 따라, 당국은 공격 직전 중앙 광장에서 목격된 드론들이 핵심적인 역할을 했으며, 초즌(Chosen)이 급조폭발물(IED)을 드론 무리에 무기화했다는 사실을 밝혀냈습니다. 각 드론에는 카메라와 안면 인식 기술이 탑재되어 있어 목표물을 다양한 각도에서 스캔했기 때문에, 각기 다른 시점에서 얼굴을 감지했습니다. 오탐(false positive)을 방지하기 위해, 소프트웨어는 각 드론에 설치된 안면 인식 애플리케이션의 결과를 자동으로 비교하여 최소 두 대의 드론이 목표물을 감지했을 때만 폭발물을 투하했습니다. 이 방법을 통해 드론의 목표물 감지 정확도가 90%까지 향상되어, 초즌이 이 골칫거리를 제거할 기회를 낭비하지 않도록 했습니다. 드론 함대는 서점에서 나오는 목표물을 스캔했고, 브라운 씨의 신원이 확인되자마자 드론들은 급강하하며 폭발물을 투하하여 목표물을 사살했습니다.

몇 주 후, 경찰은 광범위한 수사를 마무리하고 공격에 연루된 '초즌(Chosen)' 소속 3명을 체포했습니다. 경찰은 보도자료를 통해 드론 애호가들을 위해 여러 업체를 통해 온라인에서 드론을 구매했다는 증거가 있다고 밝혔습니다. 또한, 초즌이 드론 함대를 조종하고 드론 비행을 완벽하게 조정할 수 있도록 해주는 멀티 플라이어 컨트롤러 스마트폰 앱도 같은 웹사이트를 통해 구입했다고 전했습니다. 안면 인식 소프트웨어와 관련하여, 경찰은 초즌이 컴퓨터 과학 분야에 대한 전문 지식이 부족한 것으로 알려졌지만, 온라인 리뷰와 설명을 바탕으로 가장 적합한 옵션을 선택하여 시중에서 판매되는 안면 인식 소프트웨어도 온라인에서 구입했을 것으로 추정된다고 밝혔습니다. 소프트웨어 자체에는 설치 과정이 자세히 설명된 설명서가 포함되어 있었습니다. 또한, 프로그램 사양에는 작동을 위해 여러 장의 표적 사진만 필요하다는 내용이 명시되어 있었는데, 이는 브라운 씨와 같은 유명 인사라면 쉽게 구할 수 있는 정보였습니다.

그림 5: 전문가 그룹 회의 참석자들은 시나리오 2의 가능성을 인식했습니다.

시나리오 3:

국경은 더 이상 없다 - 가짜 여권으로 수도 폭탄 테러 발생

올해 초, 인기 있는 지하 해킹 커뮤니티 포럼인 "Broke.it"은 해커 그룹 "Hacking For You"의 "AI as a service" 광고를 게재하기 시작했습니다. 포럼 채널의 광고에는 변형 여권 생성을 포함한 다양한 서비스가 포함되어 있었습니다. 이 광고는 관심 있는 고객이 대상 사용자의 얼굴 사진과 현지 여권 사진 한 장만 보내면 수수료를 지불하고 해커 그룹이 위조 여권을 제작하여 발급해 준다고 안내했습니다.

이 서비스는 변형된 서비스 제작이 비교적 쉽고 해커 집단이 서비스 이용료를 낮게 책정했기에, 곧 『해킹 포 유』(Hacking For You)의 베스트셀러가 되었습니다. 해커 집단의 활동과 명성은 범죄 조직 전체에서 빠르게 성장했으며, 초국적 범죄 집단은 이러한 위조 여권을 이용하여 국경 검문소를 무단으로 통과하고 국제적으로 더욱 쉽게 활동하기 시작했습니다.

이러한 변형 여권의 인기가 계속 높아짐에 따라, 테러리스트와 폭력적인 극단주의 단체를 포함한 다른 악의적인 행위자들에게도 소문이 퍼지기 시작했습니다. 7월, 해킹 포 유(Hacking For You)는 폭력적인 극단주의 단체 브림스톤(Brimstone)과 관련된 개인들로부터 변형 여권 제작 요청을 받았습니다. 테레 노스(Terre North)에 위치하며 그곳에서 활동하는 이 단체는 인근 사우스애니(Southany), 웨스트랜디아(Westlandia), 이스토폴리스(Eastopolis)를 포함한 이 지역 전역의 소프트 타깃에 대한 폭력적이고 악명 높은 공격으로 악명 높습니다. 해킹 포 유는 수수료를 선불로 지불한 후, 브림스톤의 요청을 수락하고 여러 개의 변형 여권 제작을 시작합니다.

브림스톤은 몇 주 만에 변형 여권을 수령했습니다. 이 여권은 단체 구성원의 얼굴과 테레 노스, 사우스애니, 웨스트랜디아, 이스토프-올리스 출신 국민들의 얼굴을 불법적으로 획득하여 합성한 것입니다. 국경 검문소 직원들이 여행자의 이름, 신분증 번호, 프로필 사진을 빠르게 확인하는 경향이 있기 때문에, 이 여권 덕분에 단체 구성원들은 발각되지 않고 국경을 넘나들 수 있습니다. 브림스톤은 이 기능을 활용하여 납치 및 몸값 요구, 불법 채굴, 갈취, 불법 마약 생산 및 유통 등 단체에 자금을 지원하는 불법 활동을 조장합니다.

하지만 그들은 곧 이 여권들이 훨씬 더 큰 잠재력을 가지고 있음을 깨닫습니다. 몇 주간의 논의 끝에 브림스톤은 다음 대규모 공격을 결정합니다. 웨스트랜디아의 수도에 있는 지하철역, 광장, 스포츠 경기장, 병원, 그리고 고등법원을 포함한 주요 목표물에 대한 연쇄 폭격입니다. 브림스톤은 연쇄 공격으로 인한 혼란 속에서 공격자들이 수도에서 동쪽으로 불과 100km 떨어진 국경을 빠르고 조용히 넘어 새롭게 변형된 여권을 이용해 테레 노스로 돌아갈 수 있을 것이라고 생각합니다. 체포를 성공적으로 피할 수 있을 것이라고 확신한 브림스톤은 계획을 실행하기로 결정하고, 공격 시기를 10월 중순으로 정합니다.

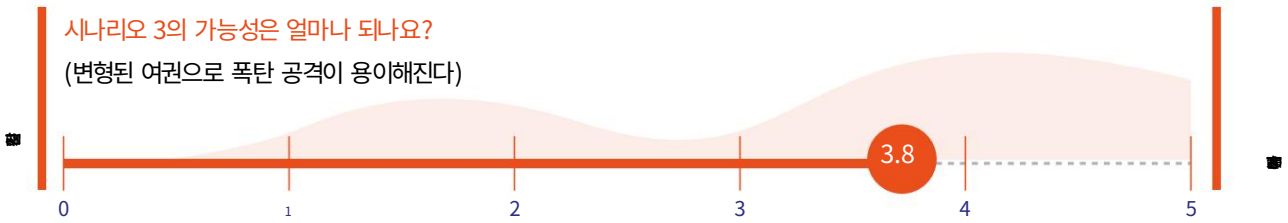


그림 6: 전문가 그룹 회의 참석자들은 시나리오 3의 가능성을 인식했습니다.

VIII. 위협 평가

AI를 테러리스트가 사용하는 현 상황을 살펴보고 테러리스트가 이 기술을 어떻게 활용할 수 있는지 보여주는 가정적이지만 상상 가능한 몇 가지 사례를 파악한 후, 중요한 질문 하나가 남습니다. 이전 장에서 묘사한 가상 시나리오에서 설명한 것과 유사한 방식으로 테러리스트 집단이나 개인이 AI를 직접 사용하는 것에 대해 우려할 만한 이유가 있을까요?

이 문제를 다루기 전에, 반 데르 비어(Van der Veer)가 지적했듯이 테러리스트의 기술 사용에 대한 논의에는 "이해관계와 의제"가 존재한다는 점을 유의하는 것이 현명합니다.210 그녀는 컨설턴트, 자문가 및 기타 민간 단체들이 자신들이 서비스를 판매하는 문제에 대해 선동적인 서사를 유지하는 데 관심이 있을 수 있다고 지적합니다. 따라서 토론에 참여하는 비전문가 청중은 편향되거나 영향을 받은 서사와 지속적이거나 중립적인 주장을 구분하는 데 어려움을 겪을 수 있으며, 특히 이러한 토론이 고도로 기술적인 문제에 초점을 맞출 때 더욱 그렇습니다.

이러한 점을 고려하여, 본 장에서는 테러리스트의 AI 사용 위협 수준을 분석함으로써 앞서 언급한 문제를 객관적으로 성찰하고 결론을 도출하고자 한다. "위협"이라는 용어는 일반적으로 의도와 능력의 결합으로 이해된다. 두 용어 모두 다음 절에서 다룰 것이다.

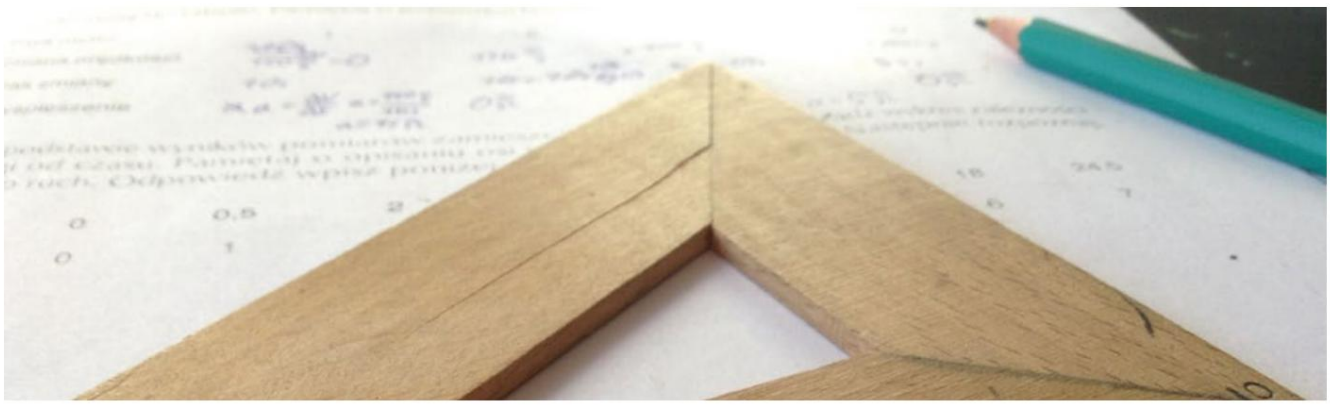


사진: Dawid Malecki, Unsplash

210 렌스케 반 데르 비어. (2019). 기술 시대의 테러리즘, 전략 모니터 2019-2020. 헤이그 전략연구센터와 클링엔달 연구소. https://www.clingendael.org/pub/2019/strategic-monitor-2019-2020/terror-ism-in-the-age-of-technology/ 에서 확인 가능

에이. 의도

테러리즘의 본질을 고려할 때, 테러 집단이나 개인의 의도를 파악하기는 어렵습니다. 그럼에도 불구하고 테러리스트의 AI 사용 의도를 평가하기 위해서는 이 기술이 테러리즘에 적합한지 살펴보는 것이 중요합니다. 크로닌은 개방적인 기술 혁신이 미래의 테러리스트들을 어떻게 무장시키고 있는지 분석하면서, 비국가 행위자들이 혁신적인 무기에서 찾는 특징들을 나열합니다.211 크로닌은 혁신적인 무기는 접근성, 가격, 사용 편의성, 휴대성, 은폐성, 그리고 효율성을 갖춰야 한다고 생각합니다. 이러한 특징들은 AI가 반드시 갖추고 있는 것은 아닙니다. 현실적으로 AI는 완벽하지 않습니다. 대중문화와 미디어에서 자주 언급되는 것과는 달리, AI는 만능 해결책이 아닙니다. 실패할 수도 있고, 실제로 실패하는 경우가 매우 많습니다.212 VentureBeat에 따르면, 데이터 과학 프로젝트의 약 87%가 실제 운영에 들어가지 못하는 것으로 추산 됩니다.213 TechRepublic의 보고서에 따르면 전 세계 CEO의 56%는 3~5년 안에 투자 수익을 기대하지 않습니다.214 AI를 효과적이고 신뢰할 수 있는 방식으로 개발하고 구현하려면 상당한 시간, 비용, 그리고 노력이 필요합니다. 호프만이 암시하듯이, 테러 집단이 한 세기 이상 총기와 폭발물이라는 두 가지 주요 무기 체계를 고수해 온 데에는 그럴 만한 이유가 있습니다. 바로 효과적이고 신뢰할 수 있기 때문입니다.215

한편, 크로닌은 혁신적인 무기가 테러리스트들에게 매력적이려면 광범위한 맥락에서 유용해야 한다고 지적합니다.216 이러한 무기는 효과를 증폭시키고, 상징적으로 공감을 불러일으키며, 예상치 못한 용도로 사용될 수 있는 기술 집합의 일부여야 합니다. 앞서 언급한 특징들과는 달리, AI는 여러 면에서 이러한 설명에 부합하며 테러리스트들이 이 기술에 관심을 가질 가능성을 뒷받침한다고 주장할 수 있습니다.

이러한 분석에도 불구하고, 이미 언급했듯이 테러 단체나 개인들이 AI 및 관련 기술에 관심을 보이는 초기 징후들이 있었습니다. 예를 들어, 드론은 ISIL과 같은 단체의 작전 방식에 점점 더 많이 통합되고 있습니다. 더욱이, 이러한 단체들이 과거에 어떻게 혁신하고 새로운 기술을 수용해 왔는지에 대한 방대한 역사를 다시 한번 생각해 볼 때, 테러 단체들이 AI가 악의적인 목적으로 어떻게 활용될 수 있는지를 어느 정도 탐구하거나 이해하려는 의도를 가지고 있을 가능성을 최소한 고려해 보는 것이 현명할 것입니다.

비. 능력

이 분석에 따르면 테러 집단이나 개인이 AI를 개발하거나 배포할 수 있는 역량은 성공과 실패를 가르는 중요한 요소가 될 수 있습니다.

AI 역량이 전 세계적으로 빠르게 성장하고 있다는 것은 부인할 수 없는 사실이며, 일반적으로 AI 기술과 이러한 기술을 개발하고 배포하는 수단은 상업적으로 확보할 수 있으며, 일부는 오픈 소스로 공개되어 있습니다. 예를 들어, 대규모 머신 러닝 및 수치 계산을 위한 오픈 소스 라이브러리인 TensorFlow를 사용하면 정교한 기술이나 컴퓨터 없이도 간단한 객체 감지 또는 얼굴 인식 모델을 갖춘 신경망을 쉽게 구축할 수 있습니다.217 Github은 사용 및 접근의 문턱을 낮춰 테러리스트와 같은 악의적인 행위자가 AI 를 악용할 가능성을 높일 수 있는 또 다른 오픈 소스 플랫폼입니다.218 그러나 기술을 활용하는 데 필요한 역량이 없다면 기술 접근성만으로는 충분하지 않습니다.

기술적 역량 수준을 검토한 결과, 전문가들은 ISIL과 같은 테러 집단이 필요한 역량이 부족하기 때문에 효과적이고 정교한 사이버 및 기술 기반 공격을 수행하지 못했다고 제안하는 경향이 있습니다.

---

211 오드리 커스 크로닌. (2020년 1월). 국민에게 권력을: 개방형 기술 혁신이 어떻게 미래의 테러리스트들을 무장시키고 있는가.

212 테렌스 쎄, 마크 에스포지토, 미즈노 타카아키, 대니 고. (2020년 6월 8일) AI 프로젝트가 실패하는 어리석은 이유. 하버드 비즈니스 리뷰. https://hbr.org/2020/06/the-dumb-reason-your-ai-project-will-fail 에서 확인 가능

Venture Beat 직원 213명 (2019년 7월 19일) 데이터 과학 프로젝트의 87%가 실제 운영에 들어가지 못하는 이유는 무엇일까요? https://venture- 에서 확인하실 수 있습니다. beat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/

214 앨리슨 드니스코 레이몸. (2018년 7월 9일). 제조업 CEO의 56%는 AI 투자에 대한 ROI가 최대 5년이 걸릴 수 있다고 답했습니다. https://www.techrepublic.com/article/roi-on-ai-investments-could-take-up-to-5-years-56-of-manufacturing-ceos-say/ 에서 확인 가능

215 유엔 마약범죄사무소. 테러리즘과 재래식 무기. https://www.unodc.org/images/odccp/ter- 에서 확인 가능 로리즘_무기_전통적.html

216 오드리 커스 크로닌. (2020년 1월). 국민에게 권력을: 개방형 기술 혁신이 어떻게 미래의 테러리스트들을 무장시키고 있는가.

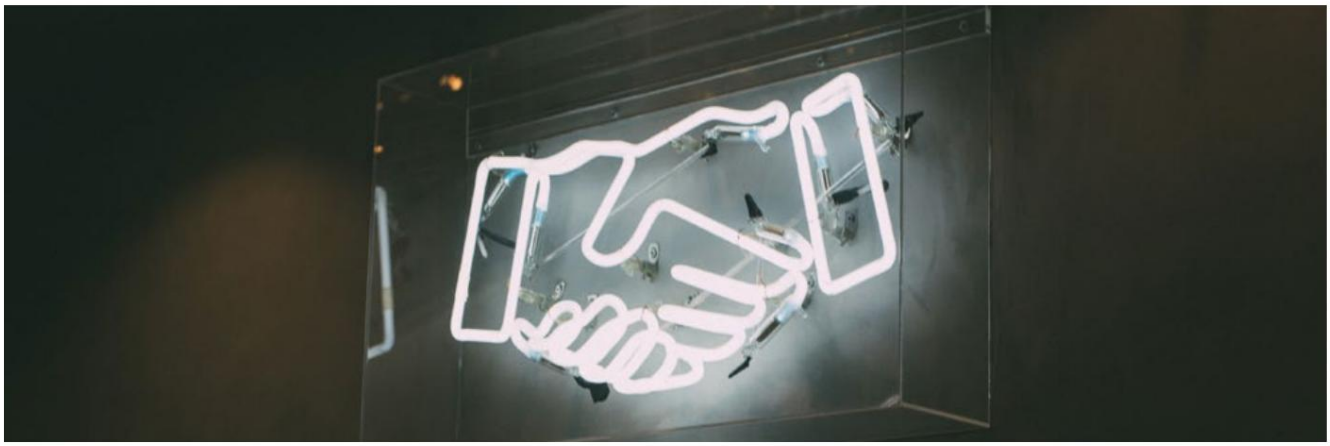
217 TensorFlow. (nd). TensorFlow 소개. TensorFlow. https://www.tensorflow.org/learn 에서 확인 가능

218 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philip Amann, Aglika Klein, Odhran McCarthy 및 Maria Eira. (2020년 11월 19일). 인공지능의 악의적 사용 및 남용. Trend Micro, EUROPOL, UNICRI. http://unicri.it/sites/default/ 에서 확인 가능 파일/2020-11/AI%20MLC.pdf

관계나 자금이 부족하거나, 단순히 그렇게 할 만큼 조직력이 부족합니다 .219 ISIL이 통합 사이버 칼리파국(United Cyber Caliphate)을 창설한 것이 우려스러운 상황이라 하더라도, 이 단체는 아직 초기 단계에 있는 것으로 여겨집니다.220 실제로 이 단체는 소위 "스크립트 키디(script kiddies)"라고 불리는 저숙련 해커들이 사용하는 기법을 사용합니다. 이들은 다른 해커들이 개발한 스크립트나 프로그램을 사용하여 작동 방식을 제대로 이해하지 못한 채 공격을 수행합니다. 그럼에도 불구하고 플래시포인트 연구원들은 "더욱 효과적이고 더 많은 지원을 받기 위해 적응하고 진화하려는 의지는 이러한 행위자들이 아직 미숙하지만, 학습하고, 방향을 전환하고, 재편하는 능력이 점차 위협적인 존재임을 시사한다"고 지적했습니다.221 통합 사이버 칼리파국과 같은 단체들이 현재로서는 필요한 기술적 역량을 갖추지 못했을지라도, 이러한 단체들의 행동은 다른 해커들의 상상력을 사로잡을 가능성이 높으며, 시간이 지나면서 더욱 유능한 차세대 사이버테러리스트들을 자극할 수도 있습니다.

테러 집단과 개인은 더욱 정교한 사이버 공격을 수행하고 이론적으로 AI 지원 공격을 시작하기 위한 자체 전문가 팀이나 사내 기술 역량을 보유하는 대신 기회를 아웃소싱하거나 분리하는 방법을 모색할 수 있습니다.222 예를 들어, 일부 알카에다/

ISIL에 영향을 받은 조직들은 자신들의 의도와 무관한 집단과도 손을 잡았습니다. 조직범죄와 테러리즘 사이의 진화하고 다면적인 연관성과 사이버 범죄 도구를 금전적 가치로 판매하는 "범죄 서비스(crime-as-a-service)" 모델의 등장을 고려할 때, 이러한 저숙련 테러 집단은 기성품 또는 맞춤형 알고리즘을 구매하여 바로 사용하거나, 그러한 시스템을 서비스로 이용할 수 있습니다. 이 "범죄 서비스" 모델의 잠재력에 대한 예로, 2020년 11월 한 디지털 권리 활동가가 텔레그램에 게시된 광고를 통해 모스크바의 얼굴 인식 시스템에 대한 액세스 권한을 16,000루블(약 200달러)에 구매한 것으로 알려졌습니다.223 이 수수료를 지불한 직후, 이 활동가는 지난 몇 달 동안 얼굴 인식 시스템에 기록된 자신의 움직임에 대한 자세한 보고서를 얻을 수 있었습니다. 이러한 의미에서 테러 조직은 반드시 그러한 기술을 직접 개발할 필요는 없지만, 블랙 마켓을 통해 "임대 해커"나 다른 사악한 범죄 집단에 공격을 아웃소싱할 수 있으므로 범죄의 접근성과 테러리스트 역량 간의 직접적인 상관관계가 더 큽니다.



Unsplash의 Charles Deluvio가 찍은 사진

게다가 이미 개발된 정교한 기술이 잘못된 사람의 손에 넘어갈 위험도 항상 존재합니다.
드론과 갈등 지역에서의 드론 사용에 대한 관심이 높아지고 전투 환경에서 점점 더 자율적인 무기 시스템이 개발되고 배치됨에 따라 이러한 무기가

219 그랜트 그로스. (2016년 4월 28일). ISIS의 사이버 공격 역량은 현재로서는 체계적이지 않고 자금도 부족합니다. IDG 뉴스 서비스의 PC 월드.
https://www.pcworld.com/article/3062980/isis-cyberattack-capabilities-are-unorganized-underfunded-for-now.html 에서 확인 가능

220 CS Liang. (2018). 『연합 사이버 칼리프국 출간과 전자 테러리스트의 탄생』, 조지타운 대학교 출판부.

221 L. 알쿠리, A. 카시러, A. 닉슨. (2016년 4월). ISIS를 위한 해킹: 새롭게 부상하는 사이버 위협 환경. 플래시포인트.
L. Alkhouri, A. Kassirer 및 A. Nixon, 2016. “ISIS를 위한 해킹: 새롭게 부상하는 사이버 위협 환경”, Flashpoint, 2016년 4월
L. Alkhouri, A. Kassirer 및 A. Nixon, 2016. “ISIS를 위한 해킹: 새롭게 부상하는 사이버 위협 환경”, Flashpoint, 2016년 4월

222 이는 21세기에 국가 및 비국가 행위자들이 인간 또는 기술적 대리자 중 하나 또는 둘 다를 사용하여 전쟁을 수행하는 경향이 점차 커지고 있다는 더 광범위한 경향의 일부로 이해될 수 있다.
Krieg, Andreas 및 Rickli, Jean-Marc (2019). Surrogate Warfare: The Transformation of War in the Twenty-first Century. Georgetown: Georgetown University Press. http://press.georgetown에서 접속 가능.
edu/book/georgetown/surrogate-warfare.

223 러셀 브랜덤. (2020년 11월 11일). 모스크바의 안면 인식 시스템을 단 200달러면 해킹할 수 있다는 보도가 나왔습니다. The Verge. 접근성
https://www.theverge.com/2020/11/11/21561018/moscows-facial-recognition-system-crime-bribe-stalking 에서

비국가 행위자(테러 집단 등)가 압수하거나 불법적으로 구매 또는 취득할 수 있습니다.224 이러한 가능성은 실제로 자율 무기 시스템 개발 금지를 요구하는 전문가들이 자주 사용하는 주장 중 하나입니다.225

궁극적으로 ISIL과 같은 집단이 AI를 직접 설계, 개발 및 구현할 역량이 부족한 것처럼 보일 수 있지만, 그러한 집단이나 개인이 이러한 기술을 배포할 역량을 확보할 가능성을 배제할 수는 없습니다. 역사적으로 역량 측면에서 상당한 발전이 비교적 짧은 기간 내에 이루어졌다는 점에 유의해야 합니다. ISIL이 드론을 자신들의 작전 레퍼토리의 일부로 사용하려는 초기 관심을 보인 후 1년도 채 걸리지 않아 성공적으로 작전에 활용했다는 것은 이를 보여주는 증거입니다.226 현재 가장 우려되는 기술은 진입 장벽이 낮은 기술이지만, 악의적인 행위자들은 시간이 지남에 따라 더욱 진보된 공격을 위한 기술을 축적할 가능성이 높습니다.

c. 우려할만한 사항?

2004년, 미국 9·11 위원회는 2001년 9·11 테러로 이어진 사건들에 대한 보고서를 발표했습니다. 이 보고서에서 위원회는 상상력의 실패가 초래할 수 있는 위험성을 강조했으며, 향후 테러 위협 평가에 상상력을 제도화할 것을 장려하기까지 했습니다.227 이러한 뼈아픈 교훈을 되새겨 볼 때, 테러리스트들이 AI를 사용할 가능성을 고려해 보는 것이 현명할 것입니다. 현재로서는 테러리스트의 의도나 역량에 대한 평가가 완전히 확정적이지는 않지만, 대비하지 못한 상황에 처하지 않기 위해서는 앞서 언급한 의도와 역량 평가에 있어 신중을 기하는 것이 바람직합니다.

여러 차례 언급했고 현재 보고서에서도 언급했듯이 기술은 테러 형태를 형성하는 데 중요한 역할을 합니다.228 AI가 일상생활에 빠르게 통합되고 있다는 점을 고려하면 테러 집단과 개인이 머지않은 미래에 AI 기반 기술을 사용하지 않을 가능성이 더 이상 낮다고 할 수 없습니다.229 이 보고서에 설명된 하나 이상의 악의적 사용이든 아니면 아직 상상할 수 없는 다른 방식이든 말입니다.230 이런 점에서 AI의 발전과 진보, 그리고 테러 집단과 개인이 이러한 기술과 관련 기술에 점점 더 관심을 갖는 것은 간과해서는 안 됩니다.

그럼에도 불구하고, 고려해야 할 또 다른 측면이 있습니다. F-Secure의 연구원인 앤디 파텔은 오늘날의 AI 시스템이 인간을 두려워해야 할 이유가 인간이 AI 를 두려워해야 할 이유보다 더 크다고 주장했습니다 .231 이와 관련하여 그는 테러 집단과 개인이 AI 시스템을 공격의 일부로 사용하기보다는 악용할 가능성이 더 높다고 지적합니다.

따라서 AI와 테러리즘의 교차점에 대한 이 고찰을 마무리하기 전에 마지막으로 한 가지 차이점을 짚어볼 필요가 있습니다. 바로 AI의 사용과 남용의 구분입니다. AI의 악의적 사용은 악의적인 행위자가 공격의 효율성을 높이기 위해 AI를 사용하는 것과 관련이 있는 반면, AI의 남용은 공격 대상을 겨냥한 공격과 관련이 있습니다.

224 콜린 P. 클라크. (2018년 8월 20일). 드론 테러는 이제 현실이며, 우리는 이 위협에 대응할 계획이 필요합니다. 세계경제포럼. 출판-
랜드 연구소(RAND Corporation)와 협력하여 작성되었습니다. https://www.weforum.org/agenda/2018/08/drone-terrorism-is-now-a- reality-and-we-need-a-plan-to-counter-the-threat/ 에서 확인 가능합니다.

225 P. Chertoff. (2018년 10월) 치명적 자율무기체계 확산의 위험: 비국가적 획득 방지, 전략적 안보
분석, 제네바 안보 정책 센터.

226 TH Tønnessen. (2017). 이슬람 국가와 기술 - 문헌 검토, 테러리즘에 대한 관점, V. 11/6.

227 9/11 위원회. (2004년 7월 22일). 9/11 위원회 보고서. Thomas H. Kean. https://www.9-11commission.gov/ 에서 열람 가능
보고서/911보고서.pdf

228 HK Tillema. (2002). 테러리즘과 기술에 대한 간략한 이론. TK Ghosh (편), 『테러리즘과 대응의 과학과 기술』(Science and Technology of Terrorism and Counter-
테러.

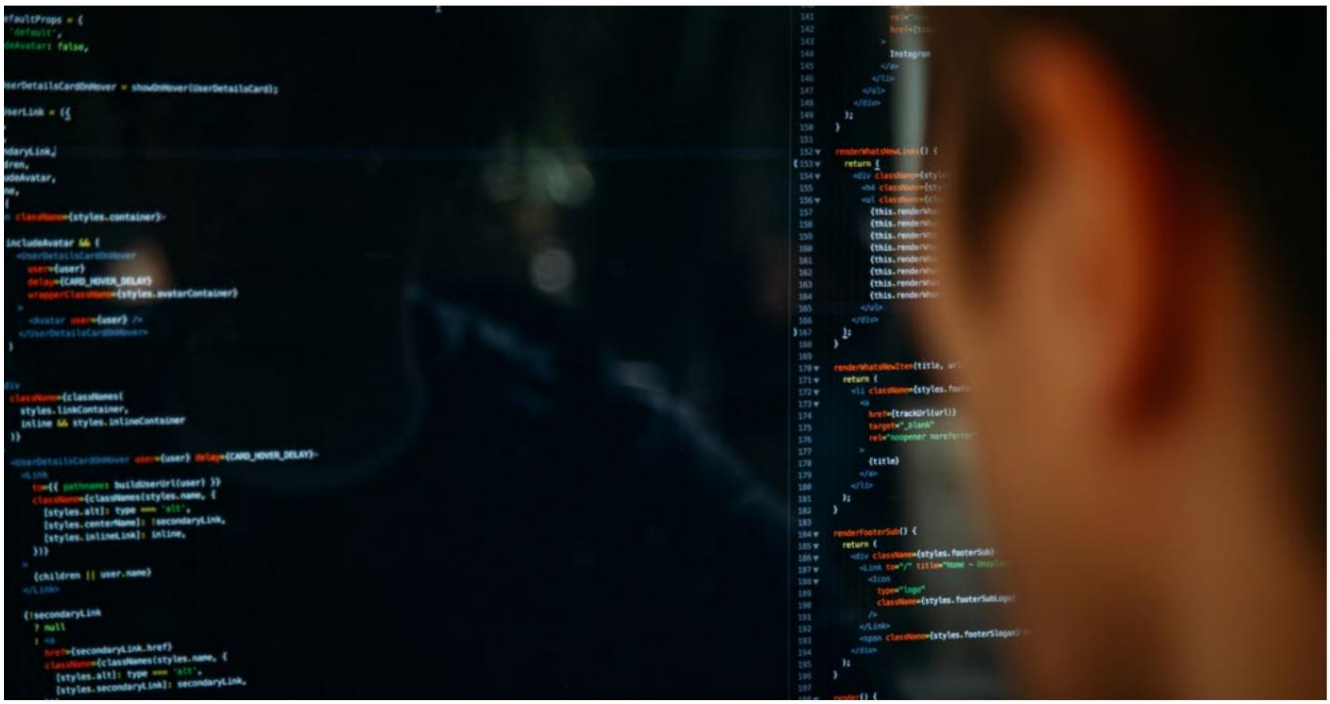
229 K. Steinmüller, (2017) 2040년의 세계. 잠재적 테러리스트 식별 및 새로운 테러 유형을 위한 기본 조건
적대적 계획: 새로운 기술과 새로운 테러 대응 전략, TJ Gordon 외 (편집자).

230 Adam Pilkey. (2019년 7월 11일). 인공지능 공격. F-Secure. https://blog.f-secure.com/artificial-intelligence-at- 에서 확인 가능
압정/

231 헬프넷 시큐리티(Help Net Security). (2019년 7월 16일). 공격자는 인공지능을 어떻게 악용할 수 있을까요? 헬프넷 시큐리티. https://www.helpnet- 에서 접속 가능
security.com/2019/07/16/인공지능-남용/

AI의 작동 또는 기능을 물리적 및/또는 사이버 기능을 이용하여 조작함으로써 침해할 수 있습니다.232 이러한 사용/남용 역학은 가상 자산을 포함한 다른 기술에서도 찾아볼 수 있습니다. 이러한 자산은 사이버 범죄를 조장하는 데 확실히 사용되지만, 암호 자산과 기업은 해커와 사기꾼의 표적이 되는 경우가 점점 더 늘고 있습니다.233 AI 남용은 테러 집단이 AI 시스템에 대한 공격을 선동하거나 그러한 시스템을 방해하려는 상황을 본질적으로 수반합니다. AI가 공공 및 민간 부문 모두의 시스템에 점점 더 통합됨에 따라, 특히 이러한 시스템이 중요 인프라에 내장된 경우 새로운 취약점이 발생합니다.

최근 플로리다의 한 정수 처리 사설을 독살하려는 해커의 시도는 기존 사이버 공격이 중요 인프라에 미칠 수 있는 엄청난 피해를 도시 전체 주민에게 보여주었습니다.234



Unsplash의 Charles Deluvio가 찍은 사진

AI 남용 가능성은 매우 광범위하지만, 몇 가지 시나리오를 생각해 볼 수 있습니다. 그러한 가능성 중 하나는 국가 기관이나 조직이 사용하는 AI 기술을 해킹하여 정보를 유출하고 이를 테러 목적으로 사용하는 것입니다. 사회가 점점 더 데이터 중심으로 변하고 정부와 민간 부문 모두에서 민감한 데이터를 수집함에 따라, 데이터 침해 및 손상 가능성이 비례적으로 증가합니다.

또 다른 매우 두드러진 가능성은 자율주행차 해킹입니다. 2000년대 초반부터 자동차 해킹 가능성은 이미 명백했습니다. 연구원들은 2013년 포드 이스케이프를 해킹하여 브레이크를 무력화하고, 2015년에는 미국 고속도로 한복판에서 지프 체로키를 정지시키는 능력을 시연했습니다.235 AI가 자동차 산업에 지속적으로 진출하고 자동차의 자율성과 연결성이 더욱 강화됨에 따라, 악의적인 공격자가 이러한 차량에 내장된 수억 줄의 코드 중 하나 이상을 해킹할 수 있는 가능성이 크게 증가하고 있습니다.

232 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philip Amann, Aglika Klein, Odhran McCarthy 및 Maria Eira. (2020년 11월 19일). 인공지능의 악의적 사용 및 남용. Trend Micro, EUROPOL, UNICRI. http://unicri.it/sites/default/ 에서 확인 가능 파일/2020-11/AI%20MLC.pdf.

233 유로폴. (2019년 10월 9일). 인터넷 조직범죄 위협 평가(IOCTA). 유로폴. https://www.europol.europa.eu/ 에서 확인 가능 활동-서비스/주요-보고서/인터넷-조직-범죄-위협-평가-iocta-2019

234 알렉스 마르쿼트, 에릭 레븐슨, 아미르 탈. (2021년 2월 10일). 플로리다 정수 시설 해킹은 휴면 상태의 원격 접속 소프트웨어를 사용했다고 보안관이 밝혔습니다. CNN. https://edition.cnn.com/2021/02/10/us/florida-water-poison-cyber/index.html 에서 확인 가능

235 앤디 그린버그(2015년 7월 21일). 해커들이 고속도로에서 지프차를 원격으로 파괴하다—나도 그 안에. Wired. https://www.wired에서 확인 가능. com/2015/07/해커-원격-킬-지프-하이웨이/

자율주행차 분야에서도 AI의 추가적인 오용에는 이미지 인식 시스템의 오용이 포함될 수 있습니다.
앞서 언급했듯이 자율주행차의 머신러닝 모델은 수신하는 정보의 정확성에 따라 달라집니다. 이 정보가 손상되면 차량 자체도 손상됩니다. 2020년 초, 해커들은 속도 제한 표지판에 테이프를 붙여 두 대의 테슬라 모델에서 오토파일럿 시스템을 속여 최대 시속 35마일(약 56km/h)이 아닌 최대 시속 85마일(약 136km/h)까지 가속하도록 했습니다.236 마찬가지로, 연구원들은 도로에 작은 스티커를 붙여 다른 테슬라 차량을 마주 오는 차선으로 들이받는 데 성공했습니다.237 테러 집단이나 개인이 이러한 악용을 통해 혼란을 야기할 수 있는 악의적인 잠재력은 자명합니다.

유사한 기법이 공격과 관련하여 활용되어 사람들을 목표 지역으로 유도하거나 공격 후 보안군 및/또는 응급 서비스의 도착을 지연시켜 공격의 효과를 증폭시킬 수 있습니다. 최근 한 연구에 따르면 자율주행차에 대한 비교적 소규모 해킹조차도 충돌과 교통 체증을 유발하기에 충분할 수 있습니다.238 이 연구는 맨해튼의 출퇴근 시간대 차량의 10~20%를 해킹하면 도시의 절반을 사실상 마비시킬 수 있다는 것을 발견했습니다.

Unsplash의 Sajjad Ahmadi가 찍은 사진

언급할 가치가 있는 또 다른 악용 사례는 악의적인 행위자가 AI를 사용하는 서비스나 애플리케이션을 방해할 가능성입니다. 즉, 시스템이 사용하는 매개변수를 수정하거나 시스템 학습에 사용되는 데이터 세트에 잘못된 데이터를 입력하여 AI를 오염시킬 수 있습니다. 이를 통해 악의적인 행위자는 AI 시스템을 원하는 방향으로 조종하거나, 예를 들어 잘못되거나 편향된 출력을 생성할 수 있습니다. 예를 들어, 2016년 마이크로소프트의 새로운 머신러닝 챗봇 "테이(Tay)"는 출시 직후 선동적이고 불쾌한 트윗을 올리기 시작하여 조기에 종료되었습니다.239 챗봇의 머신러닝 기능은 악의적인 공격을 받았으며, 챗봇이 트위터에서 공개적으로 자신을 표현하는 방식에 영향을 미치기 위해 의도적으로 인종차별적, 여성혐오적, 반유대주의적 언어를 조직적으로 주입했습니다. 이는 AI를 방해할 가능성을 보여주는 간단하면서도 효과적인 또 다른 사례입니다.

236 이소벨 애셔 해밀턴. (2020년 2월 19일). 해커들이 시속 35마일(약 56km) 속도 표지판에 5cm(2인치) 길이의 테이프를 붙여 테슬라 모델 S 두 대를 성공적으로 속였습니다. 시속 85마일(약 136km/h)까지 가속합니다. Business Insider. https://www.businessinsider.nl/hackers-trick-tesla-accelerating-85mph-us-ing-tape-2020-2?international=true&r=US 에서 확인 가능

237 카렌 하오. (2019년 4월 1일). 해커들이 테슬라를 속여 역차선으로 진입하게 했다. MIT 테크놀로지 리뷰. https://www.technologyreview.com/2019/04/01/65915/해커-트릭-테슬라-오토파일럿-향후-교통-방향-변화/

238 제이미 카터. (2019년 3월 5일). 연구원들은 해킹된 자율주행차가 도시 충돌 및 교통 체증을 유발할 수 있다고 경고했습니다. 포브스. https://www.forbes.com/sites/jamiecartereurope/2019/03/05/hacked-driverless-cars-could-cause-collisions-and-gridlock-in-cities-say-researchers/?sh=5fe14fbd2a09 에서 확인 가능

239 오스카 슈워츠. (2019년 11월 25일). 2016년, 마이크로소프트의 인종차별적 챗봇이 온라인 대화의 위험성을 드러냈다. IEEE 스펙트럼. https://spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/in-2016-microsofts-racist-chatbot-re-vealed-the-dangers-of-online-conversation 에서 접근 가능

AI 기반 시스템의 정상적인 작동은 2020년 2월, 한 독일 예술가가 구글 맵의 머신러닝 모델에 잘못된 데이터를 입력하여 베를린 거리의 교통량이 실제보다 많은 것처럼 구글 맵을 속인 사건에서 비롯되었습니다.240 이 예술가는 거리를 걸으며 99대의 휴대전화를 휴대했는데, 구글 맵은 이를 사람들이 차를 타고 있는 것으로 오인하여 시스템에 오류를 일으켰습니다. 여기서도 악의적인 잠재력이 드러납니다. 이러한 속임수나 표적 공격은 악의적으로 악용될 경우 혼란과 혼란을 야기하는 강력한 도구가 될 수 있습니다. 예를 들어, 사람들이 경로를 생성하기 위해 이러한 지도 애플리케이션에 점점 더 의존하는 도시 지역에서는 가짜 교통 체증 데이터를 생성하는 테러리스트가 공격을 실행하기 전에 군중을 지정된 구역으로 유인하거나, 보안군 및/또는 응급 구조대가 공격 장소에 조기 도착하는 것을 다시 한번 막을 수 있습니다.241

따라서 테러 목적으로 AI를 악의적으로 사용하는 것을 고려할 때, 그 반대 측면, 즉 AI의 악의적 남용 또한 고려해야 합니다. 테러리스트의 AI 악용과는 달리, 남용 위협은 테러 집단과 개인의 기존 역량 범위 내에 있을 수 있으며, 기존 공격 방식을 강화하고 "테러리즘의 현장"을 조성하는 데 매우 효과적일 수 있습니다.

IX. 평가에서 실행으로

AI를 테러 목적으로 사용하는 것은 아직 심각한 위협은 아니지만, 테러리즘은 결코 정체된 것이 아닙니다. 현재 테러 집단과 테러리스트들이 AI와 같은 기술을 활용할 수 있는 기술적 역량은 낮은 것으로 여겨질 수 있지만, 최신 기술 동향을 활용하려는 그들의 의지와 그 역량이 점차 향상되고 있다는 점을 과소평가해서는 안 됩니다. AI와 관련 기술이 대중에게 점점 더 쉽게 접근 가능해짐에 따라, 테러 대응 담당자들은 이러한 변화에 발맞춰 나가야 할 의무가 있습니다. 동시에, AI 기반 테러리즘이 임박한 위협은 아니더라도 테러 집단과 테러리스트들이 AI 시스템을 악용할 가능성에 대해 지속적으로 경계해야 합니다. AI가 공공 및 민간 부문, 특히 중요 인프라의 프로세스에 빠르게 통합되고 있는 것을 고려할 때, 이는 점점 더 우려되는 측면입니다.

이러한 점을 고려하여, 대테러 기관 및 법 집행 기관, 정책 입안자, 업계, 학계가 미래를 위해 고려해야 할 사항과 AI 기반 테러리즘의 잠재적 미래에 대비하기 위한 역량 강화를 위한 후속 조치의 방향을 제시하는 다음과 같은 권고안을 제시합니다. 이 권고안들은 UNCCT/UNOCT-UNICRI 전문가 그룹 회의 참가자들로부터 수집된 피드백을 바탕으로 작성 및 분류되었습니다. 권고안의 순서가 특정 우선순위를 나타내는 것으로 해석되어서는 안 됩니다.

Unsplash의 Andrew Neel이 찍은 사진

240 브라이언 배럿. (2020년 3월 2일). 한 예술가가 99대의 휴대폰을 사용하여 구글 지도 교통 체증을 조작했다. Wired. https://www.wired.com/ 에서 확인 가능 스토리/99-폰-가짜-구글-지도-교통-체증/

241 시모네 라포니, 사비오 시앙칼레포르, 가브리엘레 올리제리, 로베르토 디 피에트로. (2020). 고속도로의 냉장고: 도로 교통으로 인한 자동차 중독 게이션 앱, arXiv 사전 인쇄본 arXiv:2002.05051.

추가 연구

테러 집단과 개인이 AI를 도입하는 방식을 모니터링해야 합니다.

이 보고서를 바탕으로 연구 커뮤니티 내에서 추가적인 협의를 실시하여 추가적인 증거와 피드백을 얻어 향후 연구의 우선순위를 정해야 합니다.

특히 중요 인프라의 맥락에서 테러 집단이나 개인이 AI 시스템이나 AI 시스템에서 사용되는 데이터의 무결성을 사이버 공격할 수 있는 잠재적 위협은 추가로 평가되어야 합니다.

AI의 악의적 사용이나 남용을 둘러싼 법적 측면을 검토하고 분석해야 합니다.

생명공학, 뇌-컴퓨터 인터페이스, 데이터 추출 및 조작을 포함한 다른 기술 발전과 AI의 융합을 더욱 탐구해야 합니다.

다양한 이해관계자 간 협력

테러 목적으로 AI를 사용하는 것과 관련된 논의에 참여하는 이해 관계자의 범위는 모든 계층과 모든 지역으로 확대되어야 합니다.

기술 전문가와 비기술 전문가 간의 대화와 협력을 촉진하고 지속해야 합니다.

국가 당국과 정책 입안자는 보호 조치 및 정책의 초안 작성 및 이행에 대한 정보 제공, 예측 활동 개선 등을 포함하여 AI 연구 커뮤니티와 긴밀히 협력해야 합니다.

AI가 제기하는 중요한 과제와 관련하여 AI 실무자들 간의 논의는 기술 산업에만 국한되지 않고 시민 사회 단체, 인권 전문가, 젠더 자문가를 포함한 모든 이해 관계자를 참여시켜야 합니다.

인식 제고 및 지식 구축

AI 기반 도구가 테러 목적으로 악의적으로 사용되고 남용될 수 있다는 점에 대해 정부와 업계 파트너에게 인식을 제고하는 노력이 강화되어야 합니다. 이러한 인식은 잠재적 위협에 적시에 대응하는 데 중요하기 때문입니다.

개발 중인 기술의 악의적 사용 가능성에 대한 인식을 높이기 위해 AI 연구 커뮤니티와 긴밀히 협력해야 합니다.

연구 커뮤니티 내의 지식과 인식은 기술 개발의 초기 단계부터 구축되어야 하며, 예를 들어 학생 커뮤니티와 보조금을 신청하는 연구자를 타겟으로 삼아야 합니다.

AI 기술에 대한 정책 입안자들의 이해력, 특히 악의적 사용 및 남용 가능성에 대한 인식을 높여야 합니다.

위협 수준과 위협 시나리오의 특성을 과장하지 않는 한편, 주의를 환기하는 데 주의해야 합니다.

역량 강화

모든 이해관계자가 테러 목적으로 AI를 악의적으로 사용하고 남용하는 위협을 식별하고 대응하는 역량을 향상시켜야 합니다.

이해관계자 간의 협력과 조정, 경험 공유를 원활하게 하기 위한 활동을 조직해야 합니다.

정책 및 지침

AI 기반 공격에 대응하는 방법에 대한 명확한 정책과 실질적 지침은 국가와 조직에서 고려하고 개발해야 하며, 이는 유엔 헌장, 세계인권선언, 국제법의 규범과 기준에 명시된 가치에 부합하는, 이러한 공격에 대한 적절하고 충분한 대응 조치를 보장하기 위한 것입니다.

AI 시스템이 적대적 사용으로부터 보호되고 오용될 경우 책임을 질 수 있도록 하는 규제 및 인증 절차를 모색해야 합니다.

테러 대응에 AI 활용

AI와 관련 신기술을 활용해 AI 기반 테러 위협에 대응하는 방안을 모색해야 하며, 특히 테러리스트의 급진화를 막고 긍정적인 이야기를 퍼뜨리는 데 활용해야 한다.

AI와 테러 대응의 교차점에 대한 포괄적이고 심층적인 매핑이 이루어져야 합니다.

테러리즘에 대응하기 위해 AI를 사용하는 모든 경우, 인권이 중심에 있어야 하며, 테러 목적으로 AI를 악의적으로 사용하거나 남용하는 것을 방지해야 합니다.