FOCUS

인공지능과 국가안보

윤정현 국가안보전략연구원(INSS) 신안보연구실 부연구위원 | xavier94@naver.com

21세기 안보의 최우선 화두인 인공지능

‘안보(Security)’는 위협이나 공포로부터 자유롭고 걱정이 없는 상태의미하는 사회과학 용어다. 미국과 소련이 대치하던 ‘냉전(Cold War)’ 시기, 안보는 외부의 물리적인 위협으로부터 국가라는 정치단위를 보호하는 ‘국가안보(National Security)’ 개념과 동일시되었다. 탄도미사일과 핵잠수함, 장거리폭격기 등 군사무기들이 안보를 위협하는 대표적인 요소였다. 그러나 냉전의 종식과 정보·디지털 혁명이 심화되면서 수호해야할 안보의 대상과 위협을 바라보는 시각 역시 변화가 불가피해졌다. 디지털 기반 신기술 체계가 군사·산업·문화 전반에 광범위하게 적용됨으로써 새로운 불확실성과 도전이 제기되었기 때문이다. 이는 단순히 신무기 개발 경쟁이 초래하는 ‘안보딜레마(Security Dilemma)’[1] 문제뿐만 아니라, 지속 가능한 혁신과 번영을 위한 핵심 기술자원에 접근할 수 있는지를 판단하는 ‘기술주권(Technology Sovereignty)’의 문제도 결부시켰다. 또한, 이들의 통제와 글로벌 규범 정립의 주도권까지 고려해야 하는 매우 어려운 난제로 자리하게 되었다. 이러한 21세기의 복합적인 환경에서 볼 때, ‘인공지능(Artificial Intelligence, AI)’은 그 자체가 국가안보의 수호 대상이자 위협 수단으로서 양면성을 가진 대표적인 기술이다. 더 나아가, 핵무기와 같이 지구적 차원에서 적절한 통제 수단과 올바른 활용에 대한 논쟁이 끊이지 않는 품목이기도 하다.

1 안보딜레마는 자국의 안보를 강화하기 위해 군사력을 증강한 행위가 도리어 다른 국가의 맞대응적 군사력 증강을 낳는 결과를 초래함으로써 이전보다 안보 불안에 놓이게 된 역설적 상황을 의미한다.

치열한 글로벌 혁신경쟁을 초래하는 인공지능

AI는 사회경제 전반의 디지털 혁신을 촉진시키는 원동력이다. AI는 기존 산업 내 요소의 자동화와 지능화를 통하여 제조·서비스의 혁신을 창출한다. 나아가 다양한 산업과 융합되어 시너지를 창출하며 마치 에너지원과 같이 산업 곳곳에 스며들어 영향을 미치는 범용기술로 기능하고 있다. 실제로 2020년대에 들어서며 자율주행자동차, 지능형 로봇, 스마트 팩토리 등 제조업 융합분야와 의료용 AI, 지능형교육, 핀테크 등 서비스업 융합 분야에서 산업 간의 치열한 주도권 경쟁이 벌어지는 중이다.

이 같은 전방위적 파급력 때문에 국가 AI 경쟁력과 활용 역량은 안보적으로 중요한 사안이 되었다. 자연스럽게 이를 누가, 어떻게 지배할 것인가의 문제 역시 대두될 수밖에 없었다. 특히 비인간 행위자인 AI 알고리즘은 중립적으로 작동하는 것이 아니라 이를 설계한 기업과 집단의 의도와 사회경제적 가치를 반영할 수밖에 없다. 따라서 AI 권력은 이를 개발하는 기술과 지식을 확보하는 능력과 관련되며, 여기서 더 나아가 작동 과정에서 활용되는 데이터 규모와 자유로운 접근성도 영향을 미친다. 즉, 국제정치에서 AI는 새로운 혁신의 주도권을 둘러싼 권력게임의 핵심으로 자리 잡게 된 것이다. 실제로 AI로 대변되는 선도 부문 경쟁의 승패는 패권의 향배에 영향을 미친다. 미·중 전략경쟁은 사실상 기술패권 경쟁이며, 그 전면에 AI가 자리하고 있다. AI의 권력적 함의가 커지면서 이를 뒷받침하는 정책과 제도의 거버넌스를 정비하는 문제 역시 중요한 국가안보적 사안이 되었다. 이러한 측면에서 AI의 부상은 단순한 기술 트렌드에만 그치는 것이 아니라 경제·사회·문화의 혁신, 그리고 혁신경쟁의 규칙을 결정하는 거버넌스의 양상으로 진화하고 있다.

AI 군사안보의 쟁점: 전력화를 위한 책임과 윤리의 문제

4차 산업혁명 시대의 지능화와 디지털 전환의 흐름은 민간 부문뿐만 아니라 국방 분야에서의 전투개념 또한 질적으로 변환시키고 있다. 현대전은 전장요소의 정밀화, 자동화, 네트워크화의 양상으로 변한지 오래이며 전통적인 육·해·공의 전장 범위는 사이버와 우주 공간을 포함한 5차원의 영역으로 확대되는 중이다. 이 같은 변화는 국방 시스템의 복잡·정밀화를 필연적으로 수반하게 되며, 전통적인 인적 역량만으로는 첨단화된 국방체계를 통제·제어하기 어렵다는 사실을 주지시키고 있다. 이러한 점에서 인공지능 기술의 도입 여부는 미래 국방력의 향상과도 밀접히 연결된다고 볼 수 있다. 병력 수급 문제의 해소뿐만 아니라 효율적인 자원관리, 첨단 전투력의 강화 등을 위한 기술적 돌파구로 주목받고 있기 때문이다. 특히, 한반도의 지정학적 현실에서, 인공지능에 기반한 미래전의 흐름이 우리에게 갖는 의미는 더욱 클 수밖에 없다. 향후 지능화된 국방체계는 전장 및 전투지원 환경에서 핵심적인 역할을 수행할 것이며, 이를 위한 활용역량의 확보는 국방전략의 최우선적 고려사항이 될 가능성이 높다.

인공지능의 군사안보적 도입은 해당 국가에만 국한되지 않으며 주변국과 동맹국 간의 민감한 정치적 이슈의 쟁점이 된다. 특히 군사동맹의 경우, 첨단 기술 분야의 보안과 운용 전력의 효과성이 매우 중요하게

고려된다. 미국은 바이든 정부 출범 당시 한·미 동맹의 재구축과 현대화 필요성이 언급되었는데, 실천 방안으로 인공지능 분야에서의 공동 투자를 강화하고 긴밀한 기술협력을 위한 공동의 규범 준수의 필요성이 제기된 이유이기도 하다. 그러나 군사·보안시스템 부문은 고난도의 정확성과 안전성, 신속성이 필요한 분야이다. 특히 무기체계 부문은 전통 시스템을 대체하는 인공지능 기술의 도입·적용이 매우 조심스러울 수밖에 없는 현실이다. 실제로, 비정형 데이터가 대부분인 국방 분야에서는 설명 가능한 데이터 정제 기술과 수집·관리 알고리즘이 고도화되어야 한다. 예를 들어, 현재 영상 정보 및 음성 정보 인식기술과 같은 센서 단위의 기술은 발달이 되었지만, 방책을 추천하는 기술 등 사람의 결심의 판단까지 지원할 수 있는 고도화된 AI 기술은 부재한 상황이다. 이는 안전성의 문제와도 긴밀히 맞물려 있다. 국제사회의 인공지능 활용 확대가 야기하고 있는 또 하나의 쟁점은 책임성과 윤리적 문제이다. 이는 자율무기체계(AWS), 대표적으로 '킬러로봇'의 살상행위에 대한 권한 부여의 정당성과 책임의 문제로 귀결된다. 실험실과 같은 제한적 환경에서 사전에 입력된 프로그램에 따라 작동하는 '자동화(Automation)'와 달리 '자율화(Autonomy)'는 개방적이고 비구조화된 실제 환경에서 인공지능 알고리즘에 의해 수준 높은 의사결정을 수행하게 된다. 즉, 인간의 개입이 전무하고 인공지능이 표적을 임의로 선정하여 스스로 타격하는 무기이다. 이들은 전투원과 민간인을 구분해야 하는 법적 의무를 수행할 능력이 불완전할 수 있으며, 인간의 생사 여부를 인공지능이

[표 1] 군사용도로 활용되는 인공지능의 주요 기능

가짜 미디어 탐지/생성	가상 지휘관	자동화 통신
		
실황에 반응하거나 대상화된 개인 또는 집단과 실시간으로 소통하면서 가짜 미디어 보고서, 비디오, 오디오 및 소셜 미디어 게시물을 자동으로 탐지하거나 형성.	종합적인 운영 인식을 활용하여 인간과 같은 추론과 이전 작업을 기반으로 한 조언으로 운영 지휘관을 실시간으로 지원하고 조언.	개별 병사가 언제 어디서나 언어, 신체 언어 및 인간의 감정을 자동으로 즉시 식별하고 정확하게 번역할 수 있도록 함.
집단 행동 예측	**정밀 교전**	**자동 타겟팅**
		
배경 데이터(예: 소셜 미디어, 감시, 생체 인식 장치)로부터 인간 또는 집단의 행동을 정확하게 예측.	고도로 국소화된 효과(운동적 또는 에너지 기반)와 선택적인 치명성을 지닌 혼잡하고 어수선하거나 역동적인 환경에서 표적을 포착하고 교전.	원하는 운영/전략적 효과를 달성하기 위해 군사, 경제, 정보 및 외교 스펙트럼 전반에 걸쳐 정확한 맞춤 조언을 제공.

출처: NATO Science & Technology Organization(2020), p. 58.

결정하게 된다는 측면에서 인간 존엄성의 원칙에도 위배될 수 있다. 따라서 개발 단계에서부터 개발자 및 사용자가 전장에서의 살상을 목적으로 인공지능을 악용하는 것에 대한 윤리적 책임에 대한 논의가 제기될 수밖에 없다.

그럼에도 불구하고 인공지능은 침투성, 향상성, 혁신창출성이라는 고유한 속성으로 인해 군사분야의 전력 강화를 위한 매력적인 수단으로 간주되고 있다. 에 즉각적인 영향을 미치는 변수가 되었다. 민군양용기술, 무기체계·전력지원체계, 살상·비살상 분야에 적용되어 미래전의 게임체인저가 될 것이라는 평가를 받는 것이 이러한 이유이다. 이른바 신흥·파괴적(Emerging and Disruptive Technology) 기술로서 AI의 군사전략적 파급력을 평가했던 NATO는 향후 AI가 가짜 미디어 탐지 생성, 가상의 지휘관, 전장 병력 간 자동화 통신, 집단행동 예측, 정밀교전, 자동타겟팅 등에 활용될 것으로 전망한 바 있다.

국가안보의 논리로 정당화되는 AI 기술표준·통상 규제

최근의 미·중 갈등은 무역 및 통상 분쟁에서 시작해서 '핵심·신흥기술(Critical and Emerging Technology)' 분야로 확대되었다. 미래 혁신 경쟁력을 담보하는 첨단기술 이슈에서 중국에 대한 견제는 미국의 국가안보를 위한 최우선 정책으로 자리한지 오래다. 특히 AI 기술에 있어 미국의 접근전략은 크게 세 가지의 특징을 보인다. 첫째, 동맹이나 우호국, 동류 국가(Like-minded Countries) 간 네트워크 구축을 통해 글로벌 연대를 강화하고 있다. 예를 들어 2023년 5월 제4차 미·EU 무역기술협의회(TTC) 종료 후 양국은 대중국 기술견제를 목표로 무역·투자를 제한하는 '경제적 위협'에 공동으로 대응 발표했는데, 여기에는 AI, 양자컴퓨터, 바이오 기술 등이 중국 전역에 걸쳐 군사 및 시민감시 등에 사용되는 차단한다는 내용이 포함된 바 있다.

둘째, 글로벌 표준측면이다. AI의 경우, 2017년 출범한 국제표준화회의(ISO/IEC JTC1/SC42)가 해마다 2회의 국제회의를 개최, 운영되고 있다. 다만, 주목할 부분은 안보 및 통상이슈에 보다 민감한 미국과 규범, 개인정보를 중시하는 유럽과 이견이 존재한다는 점이다. 즉, 미국은 개방형 데이터 거버넌스를 지향하며 이를 표준으로 주도하는 상황이며, 유럽은 역내 안보, 중국은 데이터 자산화와 주권차원에서 대립각을 세우고 있어 양자 간 간극이 명확하다. 이 문제는 지난해 11월 '인공지능 안전 정상회의(AI Safety Summit)에서 AI 위험을 평가하는 국제 안전문제 연구소의 유치를 두고 미국과 영국 간의 신경전으로 표출된 바 있다. 가장 최근인 2024년 3월13일 유럽의회가 'AI 규제법(EU AI Act)'을 밀어붙이고 미국 빅테크 기업 다수가 우선 조사대상이 된 점도 이 같은 경쟁구도를 방증한다.

셋째, 미국은 AI를 구현하는 하드웨어 등을 통상 규제 품목에 포함시킴으로써 중국을 효과적으로 압박하고 있다. 실제로 2023년 엔비디아와 AMD사의 레거시 반도체나 범용 AI 반도체까지 수출통제 대상에 포함하였으며, 미국 클라우드 기업이 중국 AI 기업을 상대로 제공하는 서비스까지 규제 대상에 포함한다는 방침을 세웠다. 또한 백악관은 2023년 3월 AI와 반도체 분야의 민간 투자제한을 발표함과 동시에

외국인투자심의위원회(CFIUS)를 통해 기존 5세대 분야에서 시행되던 제재 강도의 범위를 확대하기도 하였다. 당시 4세대 분야에서 엔비디아, 인텔, 퀄컴 등 미국의 주요 반도체 기업들이 경계 리스트에 놓았던 화웨이에 수출하지 못하도록 기존에 확보한 수출 허가조차 취소하는 방안을 고려하기도 하였다.

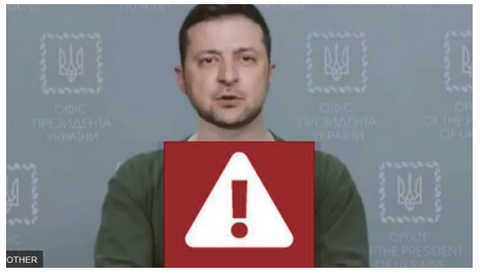
생성형 AI 시대의 새로운 국가안보 리스크: 가짜뉴스와 영향력 공작

무기체계에 도입되어 전략·전술자산의 성능을 향상시키는 AI가 전시의 안보 위협을 증대시키고 있다면, 최근의 '생성형 인공지능(Generative AI)'[2]은 AI 자체의 진화에 따른 결과로써 평시의 안보위협을 초래하는 은밀한 위협요소로 작용 중이다. 생성형 AI 시스템은 방대한 양의 데이터에서 패턴과 관계를 학습하여 기본 학습 데이터와 유사하지만 동일하지는 않은 새로운 콘텐츠를 생성한다. 문제는 생성형 AI가 첨단 생물학·생화학 무기의 설계, 사이버 공격 수단, 대규모 '영향력 공작(Influence Operation)' 감행 위협을 야기할 수 있다는 점이다. 특히, 생성형 AI 기술은 민간 영역에서 발전되고 있어 국가안보 영역 내에서 통제와 제재를 실행하는데 한계를 가지고 있다. 기능적 진화와 인터페이스의 진화가 두드러진 생성형 AI를 통해 누구나 쉽게 접근할 수 있게 됨에 따라, 악의적 공격과 예상치 못한 안보화 리스크에 직면할 가능성 역시 증대되었기 때문이다. 뿐만 아니라 생성형 AI 시스템은 그 알고리즘의 입력과 작동이 항상 가시적인 것은 아니기 때문에, 해당 모델 개발에 사용된 데이터의 신뢰성 평가 등 알고리즘 애플리케이션에 대한 감독을 회피할 수 있다. 여기에 학습 데이터에 의존하는 AI의 속성을 악용하여 누군가 의도적으로 확증 편향된 정보와 데이터를 주입한다면 AI가 사회에 심각한 가짜뉴스나 허위조작정보를 유포하는 부작용을 일으킬 수도 있다. 러·우 전쟁 초기 조작되어 유출된 거짓 항복 선언 영상 및 2022년 5월 '펜타곤 폭발 사건' 가짜뉴스로 인해 S&P 지수가 30포인트 폭락한 사례들이 대표적이다.

최근 생성형 AI의 문제점이 주목받는 이유는 민감한 정치적 국면이나 비상 상황에서 엄청난 정치적 결과를 미칠 수 있기 때문이다. 예를 들어, 선거를 앞둔 거짓 정보와 스캔들의 유통은 유권자의 판단에 개입함으로써 민주적 정당성을 침해하는 결과를 낳게 된다. 생성형 AI는 유해하고 악의적인 지시에도 이를 충실히 반영한 결과물을 제공하기 때문에 선동의 도구로 활용되기 쉽고, 여론의 압박을 조장함으로써 비합리적 의사결정이나 사회적 혼란을 야기한다. 2024년이 '슈퍼선거의 해'이자 생성형 AI 영향공작이 가장 기승을 부릴 시기로 전망되는 이유이다. 최근 국내에서는 가짜뉴스를 만들어 제공하는 서비스까지 등장했으며, 언론사 뉴스 제목을 조작하여 인터넷 뉴스의 속보 형식으로 허위정보를 유포하는 사이트가 논란을 일으킨 바 있다.

2 '생성가능한(generative)', '사전 학습된(pre-trained)' 거대, 혹은 대형 언어 모델(Large Language Model)'을 의미하는 생성형 AI는 Open AI 社의 챗GPT로 대중에 널리 사용되고 있다.

[표 2] 사회혼란과 경제적 충격을 낳는 AI 기반 영향력 공작의 예

조작된 젤렌스키 대통령의 항복 영상	조작된 펜타곤 폭발사건 가짜뉴스
	

출처: https://www.cctvnews.co.kr/news/articleView.html?idxno=236057

나가며

살펴본 바와 같이 AI는 디지털 전환 시대의 효율성과 생산성을 제고하고, 군사·경제·사회 전반의 전략적 우위를 뒷받침해주는 국가 안보의 핵심 축이 되었다. AI의 활용 문제는 단순한 AI 기술개발 경쟁을 넘어 내외부적 영향변수로서 지정학-제도-규범의 측면의 다차원적 경쟁 구도를 살펴봐야 하는 복합적 사안이며, 한국적 맥락에서의 제약과 기회요소를 재해석해야 하는 문제이다. 이미 미·중을 비롯한 주요국들은 AI 기술혁신과 안보적 활용을 위해 국내 제도의 개선뿐만 아니라 동맹세력과 국제기구, 비정부 행위자, 기업 및 이해관계자들과의 파트너십을 새롭게 정립하고 있다. 이 같은 상호 견제와 협력의 지정학적 구도 속에서 우리는 향후 AI 경쟁이 초래할 국가안보의 파급력과 우리의 전략적 위치를 다차원적으로 진단하는 것이 필요하다. 즉, 기술경쟁과 규제경쟁이 동시적으로 전개되고 있는 글로벌 환경 속에서 우리의 'AI 주권'을 증진시킬 수 있는 방안에 대한 고민이 요구되는 것이다. 최근 '군사 영역에서 인공지능의 책무성 강화를 위한 고위급 회의(REAIM)'와 'AI 안전 정상회의', 'EU AI 규제법 통과' 등 굵직한 국제적인 AI 이슈가 발표되고 있지만, 실제로 다양한 이해관계자들을 포괄하고 수용성을 확보하는 수준까지는 나아가지 못하고 있다. 이는 AI 기술성숙도와 발전단계를 고려한 규범의 적용과 이에 대한 정부와 비정부, 민간(빅테크 플랫폼 기업 등)과의 구체화된 논의가 더욱 필요함을 시사한다. 2024년 9월 제2회 REAIM 회의 개최국인 우리 정부로서는 국가안보를 넘어 국제안보의 관점에서, AI가 초래할 지구적 리스크를 의제로 제시해야할 의무가 있다. 비록 오늘의 AI는 국가 간 경쟁의 핵심 쟁점이나, 미래 인류의 존엄성을 위협할 수 있는 보편적인 문제로 발전할 수도 있기 때문이다. 따라서 이해당사국 상당수가 회피하는 공허한 선언에 머물지 않도록 활용 분야별 민감성을 고려하여 세부적인 논의를 전개해야 한다. 이를 통해 광범위한 수용성을 확보하고 보편적 원칙을 구현해가는 AI의 활용 관행을 만들어 나가야 할 것이다.