Research Paper Informatization Policy Vol. 33, No.2 (2026), pp.044-065
https://doi.org/10.22693/NIAIP.2026.33.2.044

인공지능의 국가안보화: 글로벌 동향과 정책적 함의*

장 영 균** · 최 민 석***

요 약 본 연구는 인공지능(AI)이 단순한 기술혁신을 넘어 국가안보의 핵심 전략 자산으로 부상하고 있는 현실을 반영하여 AI의 안보화(securitization) 과정을 중심으로 주요 글로벌 동향과 정책적 대응을 체계적으로 분석했다. AI는 군사, 사이버, 정보, CBRN 등 다양한 분야에서 국가의 생존과 직결되는 전략적 변화를 유발하고 있으며, 미국과 중국을 중심으로 한 기술 패권 경쟁은 국제질서의 구조적 재편을 가속화시키고 있다. 특히, 자율무기, 생성형 AI 기반 정보 조작, AI-CBRN, AI 기반 사이버전 등은 복합적이고 실존적인 안보 위협으로 부상하고 있으며, 이에 따른 전략적 대응의 필요성이 제기되고 있다. 이에 본 연구는 AI 국가안보 개념을 기존 AI 안전 및 보안의 개념과 구분하여 제시하고, 국가전략, 군사적 사용, CBRN, 정보 조작, 사이버 위협 등 다차원적 분석을 통해 AI 시대의 새로운 안보 지형을 조망했다. 더 나아가 AI 기반 위협에 대한 글로벌 대응과 규범 형성의 중요성을 강조하며, 한국이 능동적 규범 설계자로서 국제사회에서 전략적 입지를 확보해야 하는 방향성을 제시한다.

주제어 : 인공지능, 안보화, 이중용도 기술, CBRN, 정보 조작, AI 사이버전

The Securitization of Artificial Intelligence in National Security: Global Trends and Policy Implications*

Jang, Yeongkyun** · Choi, Minn Seok***

Abstract This study examines the securitization of artificial intelligence (AI) as it emerges as a core strategic asset in national security, transcending its role as a mere technological innovation. AI is reshaping the security landscape by driving structural transformations across the military, cyber, informational, and CBRN (chemical, biological, radiological, and nuclear) domains directly tied to state survival. The intensifying technological competition between the United States and China is accelerating the reconfiguration of the international order. In particular, the rise of autonomous weapons systems, generative AI-enabled information manipulation, AI-driven cyber attacks, and the convergence of AI with CBRN threats underscore the growing complexity and severity of national security risks. This study conceptually distinguishes AI national security from AI safety and AI security, and provides a multidimensional analysis covering military applications, strategic supply chains, information warfare, and technological sovereignty. It concludes by emphasizing the need for proactive global governance and the establishment of international norms, proposing that South Korea should actively position itself as a normative leader in shaping the future security architecture in the AI era.

Keywords : artificial intelligence, securitization, dual-use technology, CBRN, information manipulation, AI-driven cyber warfare

접수: 2025. 9. 29; 최종수정: 2025. 11. 22; 게재확정: 2025. 12. 9
* 이 논문은 한국전자통신연구원 '생성형 AI 안전성 평가기반 마련' 과제의 논문임.
** 주저자, 한국전자통신연구원 인공지능안전연구소 선임연구원(zamar43@etri.re.kr, https://orcid.org/0000-0003-2881-2345)
*** 교신저자, 한국전자통신연구원 인공지능안전연구소 AI안전정책 및 대외협력실장(cooldenny@etri.re.kr, https://orcid.org/0009-0005-5882-0866)

Ⅰ. 서론

1. 연구 배경

21세기 들어 인공지능(Artificial Intelligence, AI)의 발전은 단순한 기술혁신의 차원을 넘어서 국제질서의 구조적 재편과 안보 지형의 근본적인 변화를 야기하고 있다. AI는 정보 처리, 의사결정 자동화, 자율무기 시스템, 사이버보안 등 다양한 분야에서 혁신적 가능성을 제공함으로써 국가 역량의 핵심 척도로 급부상하고 있으며, 이로 인하여 AI 기술은 기존의 군사력이나 경제력과 어깨를 나란히 하는 새로운 형태의 국가전략 수단으로 자리잡고 있다. 특히, 글로벌 패권 경쟁의 주요 무대가 점차 AI 기술 확보 및 활용 능력으로 옮겨가고 있다는 점에 집중할 필요가 있다. 과거 국가 간 힘의 균형(balance of power)은 군사력, 경제력, 인구, 자원, 영토 등 물리적 요소들에 의해 결정되었다. 하지만 오늘날에는 어느 국가가 얼마나 빠르게 그리고 효과적으로 AI 기술을 개발하고 전략적으로 활용하는지가 국제정치에서의 영향력과 지위를 결정짓는 핵심 요소가 되고 있다. 이러한 흐름 속에서 AI 기반의 기술적 우위를 차지하려는 경쟁은 단순한 기술 선점 차원이 아니라, 국제정치의 규칙과 질서를 좌우하려는 전략적 의도를 포함한다는 점에서 매우 중대한 의미를 지닌다(장기영, 2024; 조은일, 2024; Horowitz, 2018).

이와 같은 배경 속에서 미국과 중국 등 주요 강대국들은 AI 기술의 상대적 우위를 선점하기 위한 치열한 경쟁에 참여하고 있다(Jouvenet, 2025). 이 경쟁은 기술의 군사적 그리고 안보적 활용을 포함한 전방위적 패권 다툼으로 확장되고 있으며, 이는 마치 냉전 시기의 핵무기 경쟁을 연상케 하는 신기술 냉전(New Tech Cold War)의 양상을 띠고 있다(Asia Pacific Task Force, 2025). 대표적으로 범용 인공지능과 같은 고도 AI 기술의 조기 확보를 둘러싼 주도권 경쟁은 각국이 '퍼스트 무버 어드밴티지(first-mover advantage)'를 확보하려는 전략적 계산과 맞물려 국제사회 전반의 긴장감을 고조시키고 있다(Hannas & Chang, 2021). 무엇보다 최근 중국의 딥시크(DeepSeek)와 같은 초대형 AI 모델 개발은 중국이 AI 기술 생태계의 중심지로 급부상하는 결정적인 분기점이 되었고, 이는 미국과 중국 간의 기술 패권 경쟁을 더욱 격화시키는 계기로 작용하고 있다(Yang, 2025).

더 나아가 AI 기술의 활용 범위는 단순한 산업 응용을 넘어 군사, 사이버보안, 생화학 및 핵(CBRN), 정보전, 심지어 정치적 여론 조작에까지 확장되면서 국가의 핵심 이익을 침해할 수 있는 수준에 도달하고 있다. 자율 무기체계는 군사작전의 자동화 및 인간 개입의 최소화를 가능하도록 구현하고 있으며, 생성형 AI는 가짜뉴스 및 심리전의 핵심 도구로 악용될 수 있다는 점에서 AI는 기존의 비군사적 기술 영역을 넘어 직접적인 안보 위협으로 인식되고 있다. 이에 따라 주요 국가들은 AI 기술의 악용 가능성에 대한 우려를 공식적으로 제기하며, AI를 국가안보의 범주 안으로 본격적으로 포함시키는 '안보화(securitization)' 전략을 추진하고 있다.

실제 미국은 'AI 국가안보각서(National Security Memorandum on Artificial Intelligence)'를 발표하여 AI의 국가안보 관련 중요성을 제도적으로 명문화했으며(The White House, 2024), 동시에 'AI 액션 플랜'을 통해 안전하고 신뢰할 수 있는 AI 개발 및 활용, 국가안보 목적 달성, 글로벌 리더십 강화를 위한 구체적 이행 과제를 제시했다(The White House, 2025b). 영국 또한 AI 안전성을 다루던 기존 기관의 명칭을 'AI Security Institute'로 변경함으로써 AI 영역에서 안보에 대한 국가적 대응을 제도화하고 있다(Department for Science, Innovation and Technology & AI Security Institute, 2025). 이는 AI 기술이 더 이상 민간 기술로만 간주되지 않으며, 외교, 군사, 정보 정책 전반을 아우르는 범정부 차원의 전략 자산으로 전환되고 있음을 보여준다.

일면에서는 AI 기술이 지정학적 위상을 강화할 수 있는 전략적 기회로 작용할 수 있으나, 다른 한편으로는 일부 국가의 경우 기술 종속으로 인한 안보적 취약성 심화라는 이중적 리스크에 직면할 수 있다. 무엇보다 기술

표준과 생태계가 강대국 중심으로 재편될 경우 후발 국가들은 기술적 자율성을 상실하고 외부 기술 의존도를 심화시키게 될 우려가 있다. 이러한 상황은 국제 안보 질서의 다극화를 심화시키는 동시에 새로운 형태의 디지털 식민주의로도 전이될 수 있는 잠재성을 내포하고 있다(Pohle & Thiel, 2020; Kwet, 2019).

특히, 현재까지의 AI 안보전략에 관한 논의는 미중 전략 문서나 개별 분야별 위험 요인을 중심으로 단편적으로 전개되는 경향이 뚜렷하다. AI 군비경쟁 같은 특정 영역에서의 동향은 비교적 활발히 검토되고 있으나, 이러한 위험들이 국제정치의 구조적 맥락 속에서 어떠한 의미를 갖는지에 대해서는 충분한 논의가 이루어지지 못하고 있다. 즉, 지금까지의 접근은 AI가 불러올 수 있는 개별적 위험을 진단하는 데 집중되어왔으나, 그 결과가 국가 간 힘의 배분, 동맹구조의 재편, 규범 기반 국제질서의 변동과 같은 거시적 차원의 변화를 어떻게 설명할 수 있는지에 대해서는 설명력이 부족하다. 이는 AI가 단순히 새로운 위협 요소로 나열되는 데 그칠 위험을 내포하며, 안보 패러다임 자체에 가해지는 근본적 변화를 파악하는 데 있어 일정한 한계를 드러낸다.

더욱이 AI는 기술, 경제, 정치가 긴밀하게 교차하는 복합적 성격을 지니면서 전통적 군사안보의 범위를 넘어서는 전환을 촉발하고 있다. 이에 국가들은 안보 위협을 새로운 방식으로 인식하고 대응체계를 재구성하지 않을 수 없게 되었다. 이는 단순히 기존 안보 틀 속에 AI를 편입하는 차원을 넘어 국제질서의 규칙과 기준 자체를 재정의하는 구조적 변화로 이어지고 있다. 결국 AI는 개별 기술의 위험 요소가 아니라, 국제정치의 기본 구조와 국가전략의 패러다임을 근본적으로 전환시키는 요인으로 자리잡고 있다는 점에서 더욱 심층적이고 체계적인 분석이 요구된다.

이와 같은 내용을 고려했을 때 AI 기술의 발전이 국가안보에 미치는 함의를 면밀하게 분석하고, '안보화(securitization)'라는 이론적 틀을 통해 주요 국가들의 정책적 대응 양상을 비교 및 검토하는 것은 매우 시의적절한 연구 주제이다. 이에 본 연구는 이러한 문제의식에 기반하여 AI의 안보화 과정에서 나타나는 주요 글로벌 동향을 체계적으로 분석하고 AI 기술이 국제 안보 질서에 어떠한 구조적·정책적 변화를 초래하고 있는지를 종합적으로 제시하는 데 목적이 있다.

2. 연구 방법

본 연구는 이론적 그리고 개념적 논의를 출발점으로 하되, 이를 토대로 주요국 전략 문서, 국제기구 및 싱크탱크 보고서, 학술논문, 정책 자료 등을 체계적으로 분석하는 '질적 문헌분석(qualitative document analysis)'을 적용했다 (Morgan, 2022). 단순 기술적 수준의 동향 소개를 넘어 안보화 이론을 기반으로 AI 국가안보의 개념 틀을 먼저 정립한 후 이 틀을 기준으로 글로벌 사례와 정책 흐름을 재구성하는 이론 주도형(theory-driven) 연구 디자인을 취한 것이 특징이다. 이에 본 연구에서는 구체적으로 연구 절차는 세 단계로 설계했다. 첫째, 개념적·이론적 정교화 단계에서는 전통적 국가안보 개념과 탈냉전 이후 안보 개념의 확장 논의를 검토하고, AI 안전, AI 보안, AI 국가안보를 상호 비교함으로써 본 연구에서 사용하는 AI 국가안보의 정의와 범위를 도출했다.

글로벌 동향에 대한 구조화된 문헌 및 자료 분석 단계에서는 미국, 중국, EU, 영국 등 주요 행위자의 국가전략, 국방전략, 사이버보안 전략, AI 안전 및 안보 관련 공식 문서, 국제기구 및 주요 싱크탱크의 리포트, 언론 및 전문 매체 기사 등을 대상으로 질적 내용 분석을 수행하였다. 자료 선정 기준은 군사, CBRN, 정보 조작 및 인지전, 사이버전 등 본 연구가 설정한 네 가지 안보 영역 가운데 하나 이상과 직접 관련될 것, 정책 및 전략적 함의를 도출할 수 있을 정도로 구체적인 내용을 포함할 것 등으로 설정했다. 이 과정에서 본 연구는 Ⅱ장에서 제시할 AI 국가안보의 네 가지 범위를 분석의 기본 축으로 삼았다. 이를 기반으로 네 가지 글로벌 이슈와 체계적으로 매핑(mapping)함으로써 이론적 개념틀과 경험적 동향 분석을 연결하는 분석 매트릭스를 구성했다.

이와 같은 설계는 본 연구가 명확한 이론적 틀과 분

석 절차를 갖춘 학술 연구임을 보여준다. 첫째, 본 연구는 단순히 AI 관련 사례를 모은 것이 아니라, AI 국가안보 개념과 안보화 이론을 전제로 사전에 설정된 범주에 따라 사례를 선정 및 배치했다. 둘째, 각 사례는 군사, CBRN, 정보 조작, 사이버전이라는 공통된 분석 범주와 AI 패권 경쟁이라는 상위 축을 기준으로 비교 및 검토되며, 이러한 구조화된 비교는 동일한 틀을 사용하는 연구자라면 유사한 분류와 해석에 도달할 수 있는 학술적 재현성을 뒷받침한다. 셋째, 본 연구는 하나의 자료 유형에만 의존하지 않고, 학술논문, 정부 및 국제기구 문서, 싱크탱크 보고서, 언론 기사 등을 상호 검증하는 방식을 취함으로써 분석의 신뢰도를 높였다. 이러한 점에서 본 연구의 동향 분석은 단순한 서술이나 정책 브리핑이 아니라, 명시적 이론 틀에 기반한 학술적 분석으로 평가될 수 있다.

이와 같은 동향 연구는 다양한 분야에서 진행되며 하나의 연구 모델로 정립되었다. 모빌리티 신산업(황성수·신용호, 2019), 국가 AI 전략(김병운, 2016), 생성형 AI 규제 논의(김법연, 2024) 등에서 보여지는 것처럼 다양한 분야에서 동향 연구 자체가 독립된 학술 연구 유형으로 자리 잡아 왔다. 복잡한 기술 및 산업 변화가 빠르게 진행되는 영역에서는 다층적 자료를 기반으로 한 구조화된 동향 분석이 학술적 기여를 수행한다는 점이 점차 공감대를 얻고 있다. 본 연구의 접근 역시 이러한 학술적 흐름과 동일한 인식을 공유하며 다양한 유형의 1차 및 2차 자료를 연계하여 분석의 틀에 따라 체계적으로 재구성함으로써 실제 특정 영역에서의 맥락을 이론적으로 설명하는 학술 연구로서의 가치를 확보한다. 즉, 본 연구의 동향 분석은 단순한 현황 정리가 아니라, 검증 가능한 범주화와 비교를 통해 이론-현실-정책 간 연결성을 강화하는 방식으로 학술적으로 승인받는 분석 모델인 것이다.

3. 연구의 구성

본 연구의 구성은 다음과 같다. 이후 'Ⅱ장'에서는 AI 국가안보의 개념을 제시하고 범위를 설정한다. AI 안전(AI safety) 및 AI 보안(AI security) 등 유사 개념과의 비교를 통해 AI 국가안보의 독자성과 정책적 중요성을 도출하고, 본 연구가 분석하고자 하는 AI 국가안보 영역을 개념적으로 제시한다. 'Ⅲ장'은 AI의 안보화에 따라 나타나는 주요 글로벌 동향과 이슈를 분석하는 핵심 부분이다. 본 장에서는 AI 기술을 둘러싼 패권 경쟁, AI의 군사적 응용 확대, CBRN 영역에서의 위험 및 위협 증가, 정보 조작을 통한 사회적 혼란 유발 등 AI가 안보 환경에 미치는 실질적 영향, 다양한 사례, 그리고 주요국의 AI 국가안보 정책 함께 검토한다.

'Ⅳ장'은 분석 결과를 바탕으로 정책적 함의를 도출하여 글로벌 시각에서 AI 국가안보 정책을 분석할 필요성과 국제 사회 수준에서의 대응 방향을 제시한다. 마지막 'Ⅴ장'은 결론으로서 본 연구의 주요 분석 내용을 요약하고, AI 국가안보와 관련하여 향후 연구 및 정책 분야에서 다루어야 할 주제와 과제를 제안한다.

Ⅱ. AI 국가안보의 개념과 범위

1. 유사 개념과의 차이

안보(Security)란 본질적으로 인간 개인과 집단이 외부의 위협으로부터 생존과 안전을 보장받는 상태를 의미한다(전웅, 2004). 전통적으로 안보는 물리적 공격, 전쟁, 폭력과 같은 명시적이고 가시적인 위험에 대응하는 것으로 정의되어 왔으며, 그 대상은 개인, 사회, 국가 등 다양한 수준에서 설정될 수 있다. 특히, 국제정치학에서 안보는 특정 행위자의 생존과 자율성을 위협하는 외부의 강압 또는 공격적 행동에 대한 대응 역량을 중심으로 논의되어 왔다(민병원, 2012; 이근욱, 2009; 전웅, 2004). 이처럼 안보의 개념은 시대적 조건과 기술의 발전에 따라 그 범위와 성격이 확장되어 왔으며, 현대 사회에서는 사이버, 생명과학, 환경 등 비전통적 위협도 안보의 영역으로 포함하고 있다(정민섭·남궁승필 등, 2020).

그 가운데에서도 국가안보는 안보 개념의 중심적 위치를 차지해왔다. 전통적 국가안보 개념은 주로 군사력과 외교력에 기반한 국가 간 갈등 상황에서의 생존 문제를 다루어 왔으나, 냉전 이후 정보기술의 발달과 비국가 행위자의 부상은 안보의 범위를 비군사적 위협으로까지 확장시켰다. 오늘날 국가안보는 군사적 충돌뿐 아니라, 사이버 공격, 테러, 전염병, 에너지 위기 등 다양한 형태의 복합 위협에 포괄적으로 대응하는 개념으로 전환되고 있다(이재은, 2013).

이러한 변화의 연장선에서 최근 새롭게 대두되고 있는 것이 바로 'AI 국가안보'(AI national security)이다. 인공지능 기술은 자율적 판단과 학습이 가능하다는 점에서 기존의 물리적 무기나 정보체계와는 전혀 다른 수준의 전략적 가능성과 위험 및 위협을 동시에 제기하고 있다(안진우 등, 2020). AI는 단순한 기술혁신을 넘어 국가의 정책 결정, 군사 전략, 사회 여론, 경제 질서 등 광범위한 영역에 영향력을 행사할 수 있는 기술로 자리 잡고 있다. 특히, 자율무기 시스템, AI 기반 사이버 공격, 사회적 여론 조작, 가짜뉴스 유포, 생물학 및 화학무기의 개발 보조 등은 국가의 주권과 질서를 근본적으로 위협할 수 있다. 이러한 맥락에서 AI 국가안보란, 인공지능 기술이 군사, 정치, 사회, 경제 등 국가의 주요 기능과 질서에 구조적·전략적 위협을 가하여 국가의 핵심 이익(critical interests)을 침해하거나 불안정을 초래할 수 있는 가능성으로부터 국가가 자유로운 상태를 의미한다. 단순히 AI 시스템을 보호하거나, 오류를 수정하는 수준을 넘어서 AI 기술 자체가 국가에 위험과 위협이 될 수 있다는 점에서 그 위상은 단독의 안보 영역으로서 정립될 필요가 있다.

이와 같은 AI 국가안보 개념은 기존의 AI 안전(AI safety) 및 AI 보안(AI security)과 구분된다. AI 안전은 AI 기술을 통해 발생할 수 있는 위험 및 위협 전반에 대하여 관리 및 대응하는 개념이다. 예컨대 인공지능 시스템이 예기치 않은 방식으로 행동하거나, 인간 통제를 벗어나는 경우, 또는 편향된 판단으로 사회적 차별을 유발할 위험 등 AI로 발생할 수 있는 모든 안전 문제를 다룬다. 한편, AI 보안은 AI 시스템이 외부로부터 공격을 받아 손상되거나, 탈취되거나, 왜곡되는 상황에 대응하는 기술적인 측면에 초점을 맞춘다(Brundage et al., 2018). 물론, 개념적으로 모호한 부분이 있지만, AI 안전은 AI 보안과 AI 국가안보의 개념을 모두 포괄하고 있으며, 안전 및 보안의 영역에서 국가의 핵심 이익을 침해할 수준으로 심각해질 경우 혹은 가능성이 있을 경우 '안보화(Securitization)'되었다고 규정하는 것이 타당하다.

특히, 코펜하겐 학파의 안보화 이론은 어떤 문제가 단순한 위험 관리의 대상에서 벗어나 실존적 위협으로 인식되는 과정에 주목한다는 점에서 AI 안전, AI 보안, AI 국가안보의 단계적 구분을 설명하는 데 매우 유용하다. 여기서 안보화 행위자는 정부, 규제기관, 군 조직뿐 아니라, 기술기업과 전문가 집단까지 폭넓게 포함되며, 이들이 AI 기술을 어떤 담론으로 형성하느냐에 따라 위협의 성격이 달라진다. 초기의 시스템 편향이나 오작동과 같은 AI 안전 문제는 주로 기술적 그리고 윤리적 보완의 대상으로 여겨지고, 이 단계에서는 연구자나 기술 전문가가 중심적인 역할을 한다. 하지만 AI 위험이 사회적 및 경제적 파급력을 동반한다고 판단되는 순간에 정부나 정치 지도자와 같은 행위자들이 전면에 등장하여 위협을 재정의한다. 이들이 내놓는 정책은 AI 위험을 단순한 기술 문제에서 공공의 안전 이슈로, 더 나아가 안보적 사안으로 끌어올리는 역할을 하며, 이를 사회가 어떻게 받아들이는지가 안보화 과정의 핵심 변수가 된다 (민병원, 2006).

이러한 관점에서 보면 AI 안전, AI 보안, 그리고 AI 국가안보로의 이행은 단절이 아니라, 안보화가 심화되는 단계적 과정으로 이해할 수 있다. 예를 들어, 모델 탈취와 같은 문제는 기술적 보안 취약성으로만 인식될 때는 관리 가능한 위험으로 남는다. 하지만 동일한 공격이 선거 조작, 전력 및 금융망 교란, 자율무기 오작동 등과 직접 연결되는 순간에 위협은 실존적 차원으로 격상된다. 이때 안보화 행위자는 기술 부처보다 군 및 정보기관과 국가 지도자에 가깝게 이동하고, 청중 또한 전

문가 공동체에서 사회 전체로 확장된다. 사회적 공감대가 형성되면 국가는 기존 규범을 넘어선 심화된 기술 통제, AI 무기체계 규율, 국제규범 형성 참여를 정당화하게 된다. 결국 AI 국가안보는 기술적 위험의 단순한 확대가 아니라, 위협 인식의 질적 변화와 사회적 승인 과정을 거쳐 도달하는 안보화의 최종 단계이며, 이 전환 조건을 명확히 하는 것은 향후 AI 거버넌스와 국가전략 설계의 중요한 출발점이 된다.

이를 더욱 명확히 하기 위해 몇 가지 사례를 제시할 수 있다. 예를 들어, AI를 적용한 채용 과정에서 특정 부분에 편향된 판단을 내리는 경우는 AI 안전의 영역에 속한다. 이는 시스템 설계 오류나 데이터 편향으로 인한 결과이며, 기술적 그리고 규범적 보완을 통해 해결할 수 있다. 반면, AI 음성 모사 기술이 악용되어 정치인의 허위 발언을 조작하고, 선거 기간 동안 여론을 왜곡하는 방식으로 사용된다면 이는 명백히 국가안보 위협으로 간주될 수 있다. 마찬가지로, 자율무기 시스템이 AI에 의해 독자적으로 목표를 식별하고 공격 결정을 내리는 기술이 국제규범 없이 운용된다면, 혹은 적대 국가가 AI 기반 악성코드를 통해 발전소나 금융 시스템을 마비시키는 사이버 테러를 감행한다면, 이는 AI 보안을 넘어선 국가안보의 사안으로 인식되어야 할 것이다.

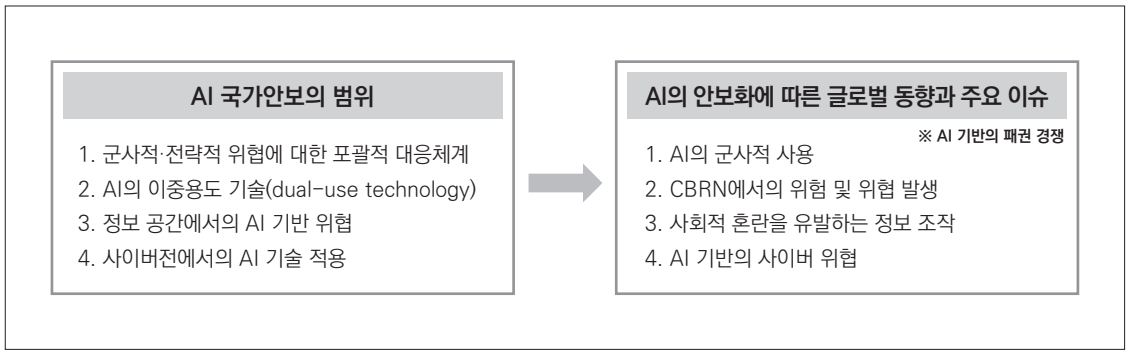
결론적으로, AI 국가안보는 인공지능 기술의 악용 가능성과 그것이 초래할 수 있는 정치·군사·사회 시스템의 교란 및 파괴 등의 위협을 포괄한다. 특히, AI가 초래할 수 있는 위험이 안전 이슈와 기술 내부의 문제를 넘어 사회 전반에 영향을 미치는 복합 안보 이슈로 확대되는 현실 가운데 유사 개념과의 명확한 구분은 향후 AI 거버넌스와 국가전략 수립의 출발점이 되어야 한다.

2. AI 국가안보의 범위

AI 국가안보의 연구 범위를 명확히 한정하는 것은 학술적 정합성과 정책적 실효성을 동시에 확보하기 위한 필수적 과제이다. 인공지능 기술은 군사, 정치, 경제, 사회, 환경 등 거의 모든 국가 기능과 연결되며, 그 영향력은 전방위적으로 확산되고 있다. 이러한 특성은 AI를 단순한 기술이 아닌, '위협의 영역'으로 간주하게 만들며, 그에 따라 국가안보 개념도 기술적, 제도적, 인식적 차원에서 확장되는 경향을 보이고 있다. 하지만 이처럼 광범위한 논의는 오히려 개념의 모호성과 적용 범위의 혼란을 초래할 수 있으며, 정책 설계와 대응 전략 수립에 있어서도 실천적 어려움을 유발할 수 있다. 따라서 AI 국가안보 개념이 내포하는 복합성과 다차원성에도 불구하고, 그 연구 범위를 전략적으로 설정하고 이론적 그리고 분석적 초점을 명확히 함으로써 국가전략 수립과 국제적 거버넌스 논의에 실질적인 기여할 수 있는 이론적 기반을 마련하는 것이 필요하다.

첫째, AI 국가안보는 인공지능 기술이 야기하는 군사적·전략적 위협에 대한 포괄적 대응체계를 포함해야 한다. AI는 감시, 정찰, 무기 운용, 사이버 작전 등 군사작전 전반에 걸쳐 활용되며, 기존의 군사력 개념을 재정의하고 있다. 자율무기 체계의 도입, 지능형 전술 판단 시스템, 적대 행위자의 AI 기반 사이버 공격은 국가의 주권과 물리적 안보를 직접적으로 위협하는 요소로 부상하고 있다. 이에 따라 AI 국가안보는 단순한 기술 통제나 무기 운용 수준의 논의에 머무르지 않고, AI의 전략적 가치, 작전 체계 내 통합 방식, 인간의 개입 수준, 윤리적 통제 장치 등 복합적 요소를 포괄하는 다학제적 분석 틀을 적용해야 한다.

둘째, AI의 이중용도 기술(dual-use technology) 특성을 고려한 생물·화학·방사능·핵(CBRN) 연계 위협 분석이 AI 국가안보 주요 축이 되어야 한다. 생성형 AI 및 파운데이션 모델의 발전은 전문가가 아닌, 일반인도 고위험 물질을 설계하거나 시뮬레이션할 수 있는 환경을 조성하고 있으며, 이는 전통적 대량살상무기(WMD) 관리 체계의 유효성을 약화시킬 수 있다. 특히, AI가 실험 설계를 자동화하거나 병원체 구조를 생성하는 데 사용될 경우 비의도적 결과마저도 국가 차원의 재난으로 이어질 수 있는 위험이 존재한다. 따라서 AI 국가안보는 기술 자체의 통제만이 아니라, 연구 환경의 투명성, 감시체계 구축, 국제적 정보 공유 및 공동 대응 방안 마



〈그림 1〉 AI 국가안보의 범위에 따른 이슈 구분

련을 포함하는 포괄적 거버넌스를 수반해야 한다.

셋째, 정보 공간에서의 AI 기반 위협, 즉 정보 조작과 인지전(cognitive warfare)에 대한 연구도 AI 국가안보의 필수 영역이다. 생성형 AI는 사실과 유사한 허위 콘텐츠를 신속하고 대규모로 생성할 수 있는 능력을 갖추고 있기 때문에 정치적 선전, 사회적 분열 조장, 민주주의 기반 훼손 등 심대한 안보 위기를 초래할 수 있다. 인지전은 사회 전반의 인지적 신뢰성을 침해하는 새로운 형태의 위협으로 진화하고 있다. 이에 AI 국가안보는 디지털 여론 공간의 구조적 안정성, 알고리즘 기반 정보 흐름 통제, 허위 정보 대응 기술 개발 및 정책적 프레임 구축 등 다양한 방면에서 연구되어야 한다.

넷째, 사이버전에서의 AI 기술 적용은 AI 국가안보 논의에서 핵심적 영역으로 부상하고 있다. AI는 기존의 인력 중심적 사이버 공격 방식을 넘어 자동화된 탐지와 적응적 공격을 수행함으로써 사이버 위협의 정밀성과 확산 속도를 비약적으로 증대시키고 있다. 이러한 특성은 군사 및 민간 인프라를 불문하고 국가의 전략적 자산 전반에 걸쳐 새로운 취약성을 노출시키며, 단순한 정보 탈취를 넘어 사회 혼란과 정치적 불안정까지 야기할 수 있다.

〈그림 1〉은 AI 국가안보의 범위에 따라 관련 이슈를 어떻게 구분해야 하는지를 보여준다. 앞서 언급한 네 개의 구분 이외에도 경제 및 무역 영역 그리고 인프라 구축의 내용을 포괄할 수 있으나, 이렇게 범위의 영역을 확장할 경우 모든 분야가 안보의 영역으로 포함될 수 있기에 상기 언급 내용으로 한정한다. 다만, 'AI 기반의 패권 경쟁'의 경우 AI 국가안보의 관점에서 모든 범위에 영향을 미치는 부분이기 때문에 별도의 그리고 최상위의 이슈로 포함되어야 한다.

III. AI의 안보화에 따른 글로벌 동향과 주요 이슈

1. AI 기반의 패권 경쟁

AI는 새로운 시대에 국가 간 패권 경쟁의 양상을 근본적으로 변화시키고 있다. 전통적으로 국제정치에서 국가의 힘은 군사력과 경제력에 의해 결정되었으나, 오늘날 AI는 이러한 하드파워(hard power)를 요소를 연결하고 증폭 시키는 전략적 자산으로 부상하고 있는 것이다. 특히, AI는 상업적인 측면에서 기술적 진보에 영향을 미치는 것은 물론이고, 정보, 군사, 외교 등 다양한 분야에서 국가의 전략적 영향력을 확장시키는 도구로 작용하고 있다. 즉, 글로벌 힘의 재편을 야기하는 중심축이 되고 있다. AI 기술의 이와 같은 특성은 기술 우위 확보를 통해 약육강식의 국제 사회에서 정치적 목적을 달성하고자 하는 국가 간 경쟁을 한층 더 심화시키고 있다. 과거에는 군사력의 증강이나 경제 제재를 통해 상대국에 대한 영향력을 행사했다면, 이제는 AI를 활용한

정보 관리, 사이버전, 지능형 무기체계 운용 등이 권력 투사의 주요 수단으로 자리잡고 있다. 이러한 기술 중심의 경쟁 구도는 갈수록 복합화되고 있으며, 불확실성과 안보 불안정을 동시에 증대시키고 있다.

이와 같은 상황에서 주목할 점은 AI 기술이 전통적 의미의 강대국뿐 아니라, 중견국이나 군사 및 경제적 자원이 열위에 있는 국가들에게도 새로운 전략적 기회를 제공한다는 점이다. AI는 자원 집약적인 기술이 아니라, 알고리즘, 데이터, 컴퓨팅 역량, 그리고 인재 확보 등 비물질적 요소에 기반하기에 일정한 기술 생태계와 전략적 투자만 갖추면 중소규모 국가도 AI 경쟁에 실질적으로 참여할 수 있다. 이는 기존의 일극 혹은 다극 체제에서 벗어나 다차원적 경쟁 구도를 형성하는 데 영향을 미치고 있다.

이러한 국제적 상황에서 글로벌 패권국 미국과 패권 도전국 중국은 AI를 국가전략의 핵심에 두고 자국 중심의 패권 질서 형성을 위한 경쟁을 본격화하고 있다. 미국은 2024년 10월, 'AI 국가안보각서(National Security Memorandum on AI)'를 통해 AI 기술의 국가안보적 활용 방안을 공식화했다(The White House, 2024). 이후 2025년 1월, 트럼프 행정부 2기는 출범과 동시에 규제 완화 및 기술 우위 확보를 중심으로 한 AI 전략을 강화하였다. 특히, 기존 바이든 정부의 윤리·안전성 중심 행정명령(Executive Order 14110)을 폐기하고, '미국 AI 리더십 장벽 제거 행정명령'(Executive Order 14179)을 발표함으로써 AI를 국가 경쟁력을 좌우할 핵심 자산으로 규정했다(The White House, 2025a).

중국 역시 AI를 국가 차원의 핵심 전략으로 설정하고 있다. 2017년 '차세대 인공지능 발전계획'을 기반으로 2030년까지 AI 분야에서 강국이 되겠다는 비전을 제시한 이래, 정부 주도의 일관된 전략을 유지하고 있다(State Council of China, 2017). 특히, 2025년 3월 양회에서는 '임바디드 인텔리전스 (embodied intelligence)'를 핵심 기조로 설정함으로써 AI 기술을 물리적 환경에 융합시켜 산업, 사회, 군사 등 다양한 영역에 걸쳐 적용하려는 포괄적 전략을 강화했다(TechNode Feed, 2025). 이는 AI가 단지 첨단 기술의 영역이 아니라, 중국식 국가통치 모델의 근간이자 국제사회에서 영향력을 확대하기 위한 핵심 수단임을 보여준다.

결과적으로 AI는 기존의 국제질서를 위협하거나 대체할 수 있는 신흥 전략 권력으로서 각국은 이를 기반으로 한 새로운 질서 재편의 선두에 서기 위해 총력을 기울이고 있다. 또한, 상대적 중견국 및 약소국은 이러한 변화 속에서 기술 종속의 위험을 줄이고 기술 주권을 확보하는 동시에 AI 글로벌 거버넌스 논의에 적극 참여함으로써 새로운 국제질서에서 실질적 목소리를 낼 수 있는 기반을 마련하고자 노력하고 있다. 즉, AI는 국가의 전략적 자율성과 안보 역량 확보를 위한 핵심 기술로 인식되고 있다.

2. AI의 군사적 사용

AI 기술은 현대 전쟁의 양상을 근본적으로 재편하는 혁신적 전략 수단으로 부상하고 있다. 전장 환경이 점점 더 복잡하고 예측 불가능해지는 가운데 AI는 감시 및 정찰, 정보 분석, 전술 판단, 무기 운용 등 군사작전의 전 주기에 걸쳐 작전 효율성과 의사결정 속도를 극대화하는 수단으로 주목받고 있다. 이러한 기술적 진보는 AI가 국가안보 전략 전반을 구성하는 중심 요소로 편입되고 있음을 보여준다.

미국은 AI를 군사 분야에 본격적으로 통합하기 위한 제도적·조직적 기반을 일찍이 마련했다. 2018년 발표된 '국방전략(National Defense Strategy)'과 '2018년 국방부 인공지능 전략 요약'을 기반으로 같은 해 '합동 인공지능 센터(JAIC)'를 설립했고, 이는 국방부 산하 각 부처 및 작전 단위에 AI 기술을 통합하고 실전 배치하는 역할을 수행했다(U.S. Department of Defense, 2018; U.S. Department of Defense, 2019). JAIC는 AI 기술을 실제 작전 환경에서 운용 가능한 시스템으로 전환하는 데 초점을 맞추었으며, 민간 기업과의 협업을

통해 기술 이전과 공동 개발을 적극적으로 추진했다. 또한, AI 기술의 무분별한 사용을 방지하고 윤리적 기준을 제시하기 위한 정책과 가이드라인을 수립함으로써 '책임 있는 AI 사용'을 위한 제도적 틀까지 마련했다(U.S. Department of Defense, 2021).

이후 2022년 JAIC는 기능을 확대 및 개편하여 '디지털 및 AI 사무소(Chief Digital and Artificial Intelligence Office, CDAO)'로 전환했다. CDAO는 기존의 AI 기술뿐만 아니라, 국방부 전반의 데이터, 분석, 디지털 역량을 총괄하는 중앙 관리 기관으로서 AI 기반 작전능력 강화, 데이터 중심의 의사결정 체계 구축, 신속한 기술 도입과 실험 체계 확립 등을 주도하고 있다(U.S. Department of Defense, 2022a). 미국 국방부는 이러한 실행체계를 뒷받침하기 위해 2023년 '데이터, 분석 및 AI 도입 전략(Data, Analytics and Artificial Intelligence Adoption Strategy)'을 수립했다. 이 전략은 AI 기술을 활용하여 전장 상황에서의 정보 우위 확보, 신속한 의사결정 역량 강화, 실시간 작전 운용 효율성 극대화를 목표로 삼고 있으며, 궁극적으로는 AI가 전투 지휘체계의 핵심 구성 요소로 작동하도록 하는 것을 추구하고 있다(U.S. Department of Defense, 2023a; U.S. Department of Defense, 2023b).

중국 역시 AI를 군사 전략에 본격적으로 통합하며 '지능화 작전(intelligentized operations)'이라는 새로운 군사 개념을 수립했다. 2018년 10월 공식 발표된 이 개념은 AI를 중심으로 한 무인 무기체계, 자율 작전 플랫폼, 사이버전 수행 능력 등을 핵심 축으로 삼고 있으며, 전통적 인력 중심 작전 방식에서 자동화·자율화 중심의 군사 패러다임으로의 전환을 추구하고 있다(Ministry of National Defense of the People's Republic of China, 2019; Kania, 2019). 특히, 중국은 민간 기술의 발전을 군사 분야에 신속하게 적용할 수 있도록 '군민융합(civil-military fusion)' 전략을 전방위로 추진하고 있으며, 이를 통해 AI 기술의 군사적 활용 속도와 범위를 비약적으로 확대하고 있다(Licata, 2023; U.S. Department of State, n.d.).

이와 같은 미중 중심의 군사 AI 경쟁은 양국의 첨단 기술 패권 경쟁을 넘어 다른 국가들의 안보화 전략에도 빠르게 파급되고 있다. 이스라엘은 가자 전쟁에서 AI 기반 표적 생성 시스템 '라벤더(Lavender)'를 실전 운용하며 전쟁 수행 방식을 근본적으로 바꾸고 있다. 라벤더는 가자 주민 대부분을 감시 데이터로 분석하여 약 3만 7천 명을 잠재적 전투원으로 표시했고, 군은 이 목록을 거의 그대로 수용해 표적당 수 초 내에 남성 여부만 확인한 뒤 공습을 승인하는 등 AI 출력값을 사실상 인간의 판단처럼 취급했다. 특히, 하급 전투원으로 분류된 대상에게는 값싼 비유도탄을 사용하여 거주 건물을 통째로 파괴하고, 한 명의 표적 제거를 위해 수십 명의 민간인 희생을 사전에 허용하는 공습 기준을 적용함으로, 안보 위협으로 규정된 주체에 대한 살상 허용 범위를 AI 시스템과 결합 후 비약적으로 확대하는 새로운 형태의 안보화 전략을 드러내고 있는 것이다(Abraham, 2024).

우크라이나는 러시아와의 전면전에서 국가 생존을 최우선 안보 의제로 설정하고, 전력 및 인구 열세를 극복하기 위해 전투원 대신 AI 기반 무인·자율 시스템이 전장에서 임무를 수행하는 데 초점을 맞추었다. 이를 구체화하는 수단으로 AI 자율 기능을 탑재한 드론과 장거리 타격 드론 등을 비대칭 전력으로 대량 운용하는 전략을 선택하고 있다. 이러한 시스템들은 아직 목표 탐지, 식별, 감시정찰 등 일부 기능에서만 부분적 자율성을 구현하는 등 인간의 통제를 유지했다. 이는 제한된 인적·물적 자원을 효율화하면서도 전장의 데이터 처리 속도와 타격 성공률을 비약적으로 높이는 방향으로 점진적·단계적 자율전 수행 능력을 발전시키고 있음을 보여준다(Bondar, 2025).

독일 역시 러시아 위협을 NATO 차원에서 구조적 그리고 장기적 안보위협으로 재정의하는 가운데 전자전 환경에서도 생존성을 확보하도록 설계된 AI 기반 드론 체계를 통해 동맹 중심의 집단방위 전략을 보완하고 있다. 독일 뮌헨 기반 국방기술 기업 헬싱은 우크라이나에 최대 100km 사거리의 전기추진 정밀타격 무

기 HX-2 드론 6,000대를 공급할 예정인데, 이 플랫폼은 온보드 AI를 탑재해 강도 높은 전자전 교란에도 대응할 수 있도록 설계되었다. HX-2는 헬싱의 정찰-타격 통합체계인 알트라(Altra)에 연동돼 단일 조종자가 다수 기체를 동시 운용하는 군집 형태의 작전도 가능하다(Bandouil, 2025).

이러한 국가별 대응은 AI 군사화가 각국이 직면한 위협을 어떻게 존재적·구조적 위험으로 규정하느냐에 따라 전략이 달라진다는 점을 보여준다. 결국 AI는 전통적 무기체계의 대체 수준을 넘어 작전 지휘체계, 전략 기획, 군사교리, 동맹구조까지 재편하는 변환의 촉매로 작용하고 있다. 동시에 AI 무기체계의 실전 배치는 자동화된 살상 판단의 윤리성, 책임 소재 불명확성, 오용 가능성 등 새로운 안보 딜레마도 심화시키고 있다. 즉, AI 군사화는 기술 경쟁이 아니라, 각국이 자국 안보를 재정의하고 정책 및 군사 구조를 변화시키는 전면적 안보화 과정으로 이해할 필요가 있다.

이처럼 AI의 군사적 활용은 자국의 측면에서 군사작전의 효율성을 비약적으로 향상시키는 긍정적 효과를 가져오고 있으나, 동시에 갈등의 확산과 오판에 따른 무력 충돌 가능성을 높이는 복합적인 위험 요소로 작용할 수 있다. AI가 무기체계나 작전 판단에 직접적으로 관여할 경우 인간의 정치적·전략적 판단이 배제되거나 약화되며, 결과적으로 전쟁 발발의 문턱을 낮출 수 있다는 점에서 심각한 안보적 도전으로 부상하고 있는 것이다(Simmons-Edler et al., 2024). 역사를 돌이켜 보면, 국가 간 전쟁의 많은 사례는 단순한 전략적 충돌이 아닌 정보의 왜곡, 의도하지 않은 신호 해석, 지휘체계의 오류 등으로 인해 발생하기도 했다. 이러한 오판(miscalculation)은 많은 경우 전쟁을 방지하려던 정치적 노력조차 무력화시켰다(Levy, 1983). 기존에는 국가가 정치적 목적을 달성하기 위해 제한된 범위의 군사력을 사용하려는 경향이 강했으나, AI 기반의 정밀 타격 시스템이나 자율 무기체계를 보유하게 되면 상대적으로 '손쉬운 승리' 가능성을 과신하게 되어 오히려 무력 사용의 유혹이 커질 수 있다.

AI가 인간의 개입 없이 전투 상황을 분석하고 자동으로 대응 결정을 내리는 시스템에 적용될 경우 상황 판단에 오류가 발생했을 때도 이를 조기에 제어하거나 수정하기 어려운 문제가 생길 수 있다. 실시간 데이터를 기반으로 한 AI의 판단이 완벽히 객관적일 것이라는 기대와 달리 데이터의 편향성이나 입력 오류는 치명적인 판단 착오로 이어질 수 있다. 이로 인하여 불필요한 무력 충돌이나 민간인 피해, 그리고 국제 규범을 위반하는 군사 행동으로 확산될 수 있는 우려가 제기된다. 더 나아가 AI가 핵무기와 같은 전략 무기의 사용 결정 과정에 개입하는 시나리오는 가장 심각한 위험 중 하나로 지적된다. 핵무기는 인류 전체에 영향을 미치는 치명적인 무기이기 때문에 철저한 인간 중심의 통제와 판단이 필수적이다. 하지만 AI가 핵 위협을 인식하거나 선제 사용을 결정하는 판단 주체로 작용할 경우 단 한 번의 오작동이나 정보 오해가 돌이킬 수 없는 재앙으로 이어질 수 있다(Geist & Lohn, 2018). 특히, 일국의 핵 선제 사용이 '도미노'처럼 연쇄적인 핵 대응을 유발할 경우 국제질서 전체가 붕괴될 수 있는 시나리오도 배제할 수 없다.

이러한 위험성과 더불어 현재 세계 각국은 AI를 접목한 신형 무기체계를 경쟁적으로 개발하면서 'AI 기반 군비경쟁(AI arms race)'이 본격화되고 있다(Reed, 2024). 이는 단순히 기술 우위를 확보하기 위한 경쟁을 넘어서 상대국의 기술 진전에 대한 불신과 위기감이 군사적 긴장을 증폭시키는 방향으로 작용하고 있다. AI 무기의 배치가 증가할수록 시스템의 오작동, 해킹, 의도되지 않은 충돌 등의 위험 역시 기하급수적으로 증가하게 된다. 즉, AI의 군사적 활용은 현대 안보의 핵심 전략 요소이지만, 그 이면에는 군사력 사용의 가능성을 확대하는 심각한 부작용이 공존하고 있다. 따라서 AI 기반 무기체계의 개발과 배치에는 기술적 검증뿐 아니라, 윤리적 기준, 정치적 통제, 그리고 다자적 협의에 기반한 국제 규범의 정립이 절대적으로 필요하다. AI가 인류의 안전을 지키는 도구가 되기 위해서는 기술이 아닌 인간이 중심이 되는 통제 구조 속에서 운용되어야

한다. 이는 향후 AI 군사화 시대를 준비하는 데 있어 가장 핵심적인 과제가 될 것이다.

3. CBRN에서의 위험 및 위협 발생

AI 기술이 화학, 생물, 방사능, 핵 분야와 결합하면서 새로운 유형의 안보 위협이 점차 현실화되고 있다(Bengio et al., 2025). 2025년에 발간된 International AI Safety Report는 AI 기반의 CBRN을 안보의 영역으로 규정하고 있다. 이로 인하여 무기 개발, 고위험 물질 설계, 비국가 행위자의 악용 가능성 등 과거에는 제한적이었던 위협 요소들이 한층 다양하고 복합적으로 나타나고 있다(Barrett et al., 2024). 특히, 생성형 AI나 파운데이션 모델의 발전은 정보 접근성과 기술 적용의 문턱을 크게 낮추며, CBRN 무기화의 가능성을 현저히 높이고 있다는 우려를 낳고 있다.

이에 2024년 7월, 미국 국립표준기술연구소(NIST) 역시 '생성형 인공지능 프로파일' 보고서를 통해 유사한 문제를 지적했으며, 보고서는 AI가 특정 화학물질이나 병원체의 구조 예측·생성을 지원할 수 있음을 경고하고 있다(U.S. National Institute of Standards and Technology, 2024). 2025년 발표된 NIST의 후속 분석인 Managing Misuse Risk for Dual-Use Foundation Models 초안에서는 '이중용도 파운데이션 모델'의 잠재적 위험을 보다 구체적으로 제시하면서, 민간 목적의 모델이 고위험 실험 또는 무기 설계에 악용될 경우 심각한 보건·안보 위기가 발생할 수 있음을 경고했다(U.S. National Institute of Standards and Technology, 2025). 특히 이러한 기술은 고급 연구시설 없이도 비교적 저렴하고 간단한 환경에서 사용 가능하므로 테러 단체나 비국가 행위자들의 접근성이 높아 위협 수준이 더욱 심각하다고 평가된다.

이러한 기술 발전은 의도된 군사적 악용뿐 아니라, 과학적 실험 과정에서 발생할 수 있는 비의도적 결과로도 이어질 수 있다. 예를 들어, AI 시스템이 실험 설계를 수행하는 과정에서 예상치 못한 독성 화합물이나 위험 병원체가 생성될 가능성이 존재한다. 이는 생명공학 연구의 속도를 높이는 장점과 동시에 안전성 측면에서는 치명적인 결과로 이어질 수 있는 잠재적 리스크를 내포한다. 더욱이 인터넷에 공개된 데이터베이스, 오픈소스 AI 툴, 저비용의 컴퓨팅 자원 등이 결합되면, 글로벌 차원의 보건·안보 위협으로 전이될 가능성도 배제할 수 없다. 특히, 테러리스트나 극단주의 세력이 공개된 AI 툴을 활용하여 독성 화학물질 또는 유전자 변형 바이러스를 설계할 수 있는 상황도 고려되고 있다. 결론적으로, AI와 CBRN 기술의 융합은 기존 안보 틀을 넘어서는 새로운 도전 과제를 제기하고 있다. 이로 인하여 단순한 기술 규제만으로는 대응이 어려우며, 이중용도 기술에 대한 사전적 관리, 투명한 연구 체계, 위험 식별 및 대응 시스템 구축, 그리고 국제사회 차원의 감시·공조 체계가 반드시 함께 마련되어야 한다. AI 기술은 인류 발전에 기여할 수 있는 도구이지만, 그 오남용이 방지되지 않는다면, CBRN 영역에서는 상상 이상의 재난을 초래할 수 있다. 따라서 AI-CBRN 연계 위험은 향후 글로벌 안보 정책의 핵심 의제로 반드시 포함되어야 할 사안이다.

4. 사회적 혼란을 유발하는 정보 조작

2019년 미국은 국제사회에서 확산되고 있는 악의적 정보 조작에 대응하기 위해 '해외 악의적 영향 센터(Foreign Malign Influence Center, FMIC)'를 설립했다. FMIC는 해외 행위자에 의해 조직적으로 제작 및 유포되는 허위 정보와 조작된 콘텐츠를 감시하고, 이로 인한 정치사회적 영향을 체계적으로 분석하며 대응하는 역할을 담당하고 있다. 특히, FMIC는 악의적인 의도를 지닌 허위 정보의 흐름을 사전에 차단하거나, 그 영향력을 최소화하기 위한 정책적 조치를 주도하고 있으며, 이는 허위 정보가 국가안보에 미치는 심대한 영향을 반영한 선제적 대응으로 평가할 수 있다(DeVine, 2024).

허위 정보의 생산과 유포는 단순히 잘못된 사실을 전달하는 차원을 넘어 사회의 정치적, 경제적, 문화적 네

트워크에 심대한 변화를 초래할 수 있는 구조적 위협으로 작용하고 있다. 현대 사회는 디지털 네트워크를 기반으로 정치적 의사결정 과정이 이루어지고 사회적 신뢰가 형성된다. 이와 같은 네트워크에 대한 인위적 조작은 여론의 왜곡, 사회적 분열, 정치체제에 대한 신뢰 상실로 직결되며, 궁극적으로는 민주적 거버넌스의 근간을 위협하는 결과를 초래한다. 특히, AI 기술의 발전은 이러한 정보 조작 전략의 속도와 정밀성을 비약적으로 증대시켰다. AI는 대규모 데이터를 신속하게 분석하고 사실처럼 보이는 허위 정보를 대량 생성할 수 있는 능력을 제공함으로써 허위 정보의 확산 과정을 자동화하고 그 파급력을 극대화하는 데 기여하고 있다. 이러한 기술적 변화는 현대 전쟁 양상에도 깊은 변화를 초래하고 있으며, 하이브리드 전쟁의 개념이 AI 시대에 재해석되고 있다(Moy & Gradon, 2023).

생성형 AI의 발전은 허위 정보 조작 전략을 더욱 정교화했다. 과거에는 가짜 뉴스(fake news)나 왜곡된 영상(deepfake)을 제작하는 데 상당한 시간과 자원이 소요되었으나, 현재는 생성형 AI를 통해 신속하고 대규모로 사실과 구별이 어려운 허위 콘텐츠를 생산할 수 있게 되었다. 악의적 행위자들은 이를 이용하여 특정 국가나 조직에 대한 부정적 인식을 확산시키는 전략을 구사하고 있으며, 이는 단순한 이미지 훼손을 넘어 헌법적 가치에 반하는 이념을 주입하거나 역사적 사실을 왜곡하는 데까지 이르고 있다. 특히, 이러한 조작은 대상 사회의 내부 갈등을 심화시키고, 체제 전복적 분위기를 조성하는 방향으로 유도되고 있다. 이러한 변화 속에서 최근 안보 분야에서는 '인지전(cognitive warfare)'이라는 개념이 부각되고 있다. 인지전은 인간의 사고, 감정, 행동 등 인지적 요소를 직접적인 공격 목표로 삼는 새로운 형태의 전쟁을 의미한다. 전통적 물리적 충돌이 아닌, 정보와 인식의 조작을 통해 상대방의 결정을 오도하고 사회적 연대와 신뢰를 해체하는 것을 목표로 한다. 인지전은 허위 정보, 심리전 등을 복합적으로 활용하여 목표 사회의 인지 구조를 변형시키고, 궁극적으로 정치적, 사회적 불안을 심화시키는 데 집중한다(NATO Allied Command Transformation, n.d.).

정보 조작을 통한 정치사회적 혼란 전략은 과거에도 존재했지만, AI 기술과 결합되면서 그 위협의 범위와 파급력은 전례 없는 수준으로 확대되었다. 이에 따라 단순한 정보 차단이나 사후적 검열 조치로는 이러한 위협에 효과적으로 대응하기 어렵다는 한계가 명확히 드러나고 있다. 현대 사회는 실시간 정보 순환 구조를 기반으로 움직이고 있으며, 허위 정보가 유포된 후의 수습은 대체로 사회적 신뢰의 회복에 실패하거나, 상당한 시일과 비용을 요구된다. 따라서 향후 허위 정보 및 인지전에 효과적으로 대응하기 위해서는 생성형 AI 기반 허위 정보의 탐지, 분석, 차단을 위한 기술적 역량을 강화하는 것은 물론, 허위 정보의 위험성을 조기에 식별하고 사회 전반에 대한 인식 제고를 통해 면역력을 높이는 전략이 병행되어야 한다. 더 나아가 국가 차원에서는 허위 정보 유포를 조직적으로 수행하는 해외 행위자에 대한 적극적 대응체계를 구축하고, 국제적 협력을 통해 정보 공간의 투명성과 신뢰성을 제고하려는 노력이 절실히 요구된다.

5. AI 기반의 사이버 위협

AI의 발전에 따라 사이버의 위협은 더욱 증가하고 있다. 기존에는 사람에 의한 사이버 공격과 방어가 진행되었다면, 현재는 발전된 AI가 스스로 공격할 지점을 파악하여 취약 지점을 공격할 수 있다. 이에 미국은 사이버 안보 차원의 대응을 강화함으로써 AI 관련 리스크를 선제적으로 통제하려 하고 있다. 사이버 분야에서는 2022년 미국 국방부가 '책임 있는 AI 전략(DoD Responsible AI Strategy)'을 수립하여 AI 기반 시스템의 보안성과 신뢰성을 핵심 과제로 명시하였다(U.S. Department of Defense, 2022b). 이어 2023년 발표된 '국가 사이버보안 전략(National Cybersecurity Strategy)'에서는 AI 시스템을 포함한 핵심 인프라 보호를 위한 보안 체계 강화를 강조하고 있다(The White House, 2023). 이는 AI 시스템이 군사 및 민간 부문 모

두에서 필수적 자산으로 자리잡음에 따라 이를 대상으로 한 복합적 사이버 위협에 대응하기 위한 제도적 기반을 마련하려는 시도로 평가할 수 있다.

한편, 중국 역시 AI 안보 역량 강화를 위해 적극적인 조치를 취하고 있다. 사이버 안보 분야에서는 2024년 2월 중국 공업정보화부(Ministry of Industry and Information Technology)가 2026년까지를 목표로 한 데이터 보안 강화 계획을 발표하였다(Ministry of Industry and Information Technology of the People's Republic of China, 2024). 해당 계획은 랜섬웨어 공격을 모의한 비상 대응 훈련 및 데이터 보안 관련 교육 프로그램을 포함하고 있으며, 이는 AI 시스템을 포함한 디지털 인프라의 사이버 복원력을 강화하려는 전략적 접근으로 해석된다. 경제적 차원에서는 AI 연산에 필수적인 반도체 기술에 대해 미국 및 서구 국가에 대한 기술 의존도를 최소화하고, 자국 내 첨단 반도체 산업의 자립도를 제고하기 위한 정책적 노력을 강화하고 있다. 이러한 행보는 기술적 자립을 통해 경제안보를 확보하고 외부 압박에 대한 구조적 취약성을 감소시키려는 전략적 의도를 반영한다(Lin & Huang, 2025).

AI 국가안보와 관련하여 부상하고 있는 핵심 위협 중 하나는 AI 시스템을 대상으로 한 사이버 공격의 고도화이다. AI 기술이 통합된 인프라 시스템은 공격자에게 새로운 공격 표면을 제공하며, AI를 활용한 사이버 공격은 자동화, 적응성, 대규모 확산 가능성, 그리고 지속적 학습을 통한 정밀화 등의 특성을 지닌다(Karthikeyan, 2024). 이는 전통적 사이버보안 체계로는 대응이 어려운 새로운 차원의 복합적 위협을 초래하고 있으며, AI 시스템의 보안은 현대 국가안보 전략에서 필수적 고려요소로 자리잡고 있다. 특히, AI 기반 인프라에 대한 공격은 단순한 정보 탈취를 넘어 사회적 혼란과 정치적 불안정을 초래할 수 있기에 국가 전체의 기능적 안정성에 대한 심대한 위협으로 작용한다. 이처럼 AI 기술은 사이버 영역에서 새로운 위협의 수단으로 부상하고 있으며, 주요국을 사이버 영역에서 자신의 핵심 이익을 유지하기 위해 선제적으로 전략을 수립하고 있다.

6. 한국의 AI 국가안보 현황 평가와 구조적 도전요인

앞서 살펴본 세계 각국의 AI 안보화 흐름은 한국의 현재 위치를 다시 바라보게 하는 중요한 기준을 제공한다. 한국은 디지털 인프라, 반도체, 기술 활용 능력에서 세계적 강점을 갖고 있지만, 정작 AI를 국가안보의 핵심 요소로 다루는 논의는 아직 충분히 자리 잡지 못했다. 그동안 AI 정책은 주로 산업 성장, 디지털 경제 활성화 같은 경제적 관점에서 추진되어 왔고, 군사, CBRN, 정보, 사이버를 아우르는 'AI 국가안보'라는 개념을 정교하게 세우고 체계화하는 작업은 시작 단계에 있다. 이 간극은 지금의 글로벌 환경에서 한국이 전략적으로 뒤처질 위험성과 함께, 국가 차원의 근본적인 점검이 필요하다는 사실을 보여준다.

AI 기반 패권 경쟁의 관점에서 보면 한국은 매우 독특한 위치에 서 있다. 미국과의 확고한 동맹, 중국과의 깊은 경제적 연결이라는 이중 구조 속에서 한국은 기술 공급망의 핵심이자 중견국 외교의 중심축을 모두 담당하고 있다. 그러나 이러한 복합적 위치는 오히려 한국이 장기적인 AI 국가안보 전략을 세울 때 더 정교한 균형 감각을 요구한다. AI가 지정학적 갈등의 핵심 변수가 되면서 기술 의존, 공급망 위험, 동맹 선택이 서로 얽히기 때문에 단순히 산업 경쟁력 강화 차원의 대응만으로는 충분하지 않다.

앞에서 확인한 네 가지 안보 영역인, 군사, CBRN, 정보 조작 및 인지전, 사이버전을 한국에 비추어 보면 분야별 발전 속도가 서로 다르게 진행되어 왔음을 알 수 있다. 군사 분야에서는 첨단 무기체계 도입이 빠르게 이루어졌지만, AI 무기체계 운영에서 인간 통제 원칙이나 윤리 기준을 명확히 제도화하는 논의는 아직 부족하다. CBRN 영역 역시 국제규범 참여 경험은 풍부하나, AI와 결합했을 때 새로 등장하는 위험을 다룰 전담 접근은 아직 미비하다. 정보 공간에서는 허위 정보 및 인지전에 대응하는 체계가 여러 기관에 나뉘어 있어 통합성과 일관성이 떨어지고, 사이버 분야에서도 AI가 만들어내는 새로운 공격 및 방어 구조를 반영한 전략적 대비는 시작

단계에 머물러 있다. 이러한 점검은 한국이 놓인 현실을 더욱 명확하게 보여준다.

이처럼 한국의 AI 국가안보 관련 제도와 정책을 종합해 보면, 기술력과 산업 경쟁력에 비해 전략, 조직, 거버넌스 측면이 상대적으로 느린 속도로 발전해 온 불균형이 드러난다. 정책은 여러 부처로 분절되어 있고, 고위험 AI 분야에 대한 위험 평가나 시나리오 분석도 제도화되지 못했다. 특히, 국제 규범 논의에서 일관된 국가전략을 뒷받침할 지식 플랫폼이나 연구 생태계는 아직 충분히 형성되지 않았다. 하지만 이러한 한계는 동시에 앞으로 보완을 통해 도약할 수 있는 공간이며, 한국이 중견국으로서 국제 AI 질서 형성에 기여할 수 있는 새로운 전략적 기회이기도 하다.

결국 글로벌 동향 분석은 한국이 어떤 방향으로 나아가야 하는지를 자연스럽게 보여준다. 한국은 외부 환경 변화에 수동적으로 대응하기보다 자국의 기술 및 안보 구조를 냉정하게 진단하고 이에 맞는 능동적 전략을 구축해야 한다. 다음 Ⅳ장에서는 이러한 진단을 토대로 글로벌 시각의 AI 국가안보 정책 분석 필요성, 군사 영역에서의 책임 있는 활용 원칙, AI-CBRN 위협에 대응하는 국제 규범 참여 등 구체적인 정책적 함의를 제시하고자 한다. 이를 통해 한국이 AI 시대의 새로운 안보 질서에서 책임 있는 규범 설계자이자 전략적 중견국으로 자리매김할 수 있는 방향을 명확히 제안할 것이다.

Ⅳ. 정책적 함의

AI의 안보화가 가속화되는 국제정치 환경은 기술·군사·경제·사회 전반의 구조적 변화를 동반하며, 이 과정에서 AI는 국가안보의 주변적 요소가 아니라 전략적 중심축으로 자리하게 되었다. AI는 군사작전의 자동화, 사이버 공간에서의 공격·방어 구도 변화, 정보조작을 통한 사회 불안정 심화, 그리고 CBRN 위협의 고도화 등 다양한 영역에서 안보적 효과를 생성하며 국가 존립과 직결되는 변수로 작용하고 있다. 따라서 AI 기술을 어떻게 관리하고 규범화하며 국제협력 속에서 제도화할 것인지는 단순한 기술정책의 범주를 넘어 미래 국제질서 재편의 핵심 요인으로 기능한다. 이러한 맥락에서 AI 기반 기술안보는 기술·군사·정치·경제·사회가 모두 결합된 복합적 안보의 문제이며, 정책 논의 또한 국내 제도 정비를 넘어 글로벌 권력 구조 속에서의 전략적 위치를 고려한 방향으로 설정될 필요가 있다.

무엇보다 AI 안보화는 특정 기술의 위험성만을 다루는 차원이 아니라, 어떤 정치적 담론과 권력 관계 속에서 위협이 구성되고 제도화되는지를 분석하는 작업이 병행될 때 비로소 타당한 정책 방향성을 도출할 수 있다. AI는 기존 권력 분포를 재편하며 강대국, 중견국, 글로벌 사우스 등 다양한 행위자에게 서로 다른 기회와 제약을 부여하고 있으며, 이러한 변화의 구조적 함의를 해석하는 작업은 한국의 전략적 선택을 설계하는 데 필수적이다. 이러한 문제의식에 기반할 때, 한국의 AI 국가안보 전략은 다음 세 가지 차원에서 보다 구체적이고 실행 가능한 정책적 방향을 설정할 필요가 있다.

1. 글로벌 시각에서 AI 국가안보 정책 분석의 필요성 : 글로벌 패권, 정보 조작, 사이버 위협

AI의 발전은 단순한 기술혁신을 넘어 국가안보의 구조적 변화를 촉발하고 있다. 군사작전 자동화, 초정밀 사이버 공격, 실시간 정보 조작, CBRN 위협 증폭 등 AI는 국가의 생존과 직결되는 핵심 변수로 기능하고 있으며, 기존 안보 체계로는 감당하기 어려운 복합적 위험을 만들어내고 있다. 그럼에도 AI를 기술적 성취나 산업 경쟁력 차원에 한정하여 이해하려는 관성은 심각한 전략적 오류를 야기할 수 있다. 특히, AI를 둘러싼 글로벌 경쟁은 기술 우위 확보를 넘어 국제질서의 규범·표준·거버넌스 주도권을 결정짓는 방향으로 확장되고 있으며, 미국과 중국은 AI를 전략적 패권 확보의 중심축으로 삼아 군사력, 정보통제, 경제적 영향력을 강화하고 있다.

AI 기술은 국경의 의미를 약화시키고 사이버 공간과 인지 전장에서 초국가적 위협을 증폭시키며 이러한 환경은 후발국이나 중견국에게 기술 종속의 리스크를 높

인다. 따라서 개별 국가의 기술 육성이나 규제 논의만으로는 충분하지 않으며, 글로벌 차원에서 AI 국가안보 이슈를 체계적으로 분석하고 대응하는 전략적 시각이 필수적이다. 주요국의 AI 안보 전략을 비교 및 모니터링함으로써 자국의 취약성을 정확히 파악하고, 군비경쟁, 정보 조작, 사이버 위협 등 초국가적 위험에 대비한 국제 협력 체계를 마련해야 한다. 특히, 생성형 AI와 이중용도 기술이 초래할 수 있는 위험을 사전에 식별하고 이를 관리할 글로벌 규범 수립 과정에 적극 참여하는 것이 중요하다.

이 과정에서 한국은 단순한 규범 수용자가 아니라, 중견국으로서 기여할 수 있는 현실적 기회를 갖고 있다. 한국은 반도체, 통신, 디지털 인프라 등 데이터 기반 산업에서 글로벌 공급망의 핵심축을 담당하고 있으며, 이는 기술 패권 경쟁 속에서 전략적 발언권을 확보할 수 있는 중요한 기반이 된다. 더 나아가 한국이 안전한 AI 생태계 구축을 위해 보유한 기술 인프라와 정책 역량을 연계한다면, 미·중 경쟁 구조 속에서도 전략적 자율성을 유지하면서 국제 규범 형성 과정에서 능동적 역할을 수행할 수 있다.

이러한 관점에서 한국은 단계별·우선순위 기반 접근을 통해 보다 실천적이고 구체적인 정책 방향을 확보할 필요가 있다. 첫째, 단기적으로 주요국의 AI 안보 전략과 위협 동향을 상시적으로 분석하여 글로벌 수준에서의 위험 식별·대응 기반을 확보해야 한다. 둘째, 중기적으로 동맹국 및 민주주의 국가들과 AI 기반 정보조작 및 사이버 위협에 대한 공동 대응 체계를 마련하고, 글로벌 플랫폼 기업과의 협력 메커니즘을 제도화하여 실질적 정보 공유·탐지·대응 기반을 확보해야 한다. 셋째, 장기적으로 국제 규범과 거버넌스 수립 과정에서 한국이 기여할 수 있는 분야를 중심으로 국제 기여 모델을 구축함으로써 책임 있는 기술 주체로서의 위상을 강화해야 한다.

정보 조작 분야에서도 생성형 AI는 딥페이크와 실시간 콘텐츠 조작을 통해 민주주의 체제의 기반을 흔들 위험을 증폭시키고 있으며, 이는 개별 국가가 홀로 대응하기 어려운 초국가적 위협이다. 따라서 한국은 정보조작 대응 기술 고도화뿐 아니라, 글로벌 플랫폼 기업과의 실시간 위험 정보 공유 체계, 국제 규범 제정 과정에서의 참여를 강화해야 한다.

사이버전 영역에서는 AI가 통합된 국가 핵심 인프라가 새로운 전장이 되고 있으며, 생성형 AI 기반 공격은 초연결 환경에서 국경을 초월해 확산된다. 이에 한국은 국가 기간시설을 중심으로 AI 기반 공격 시뮬레이션과 위기 대응 훈련을 제도화하고, 동맹국과의 사이버 위협 탐지 네트워크를 구축하여 대응 능력을 강화해야 한다.

결국 AI는 기술의 영역을 넘어 권력, 질서, 국제 규범의 문제로 확장되었다. AI를 기술혁신의 하위 영역으로만 다루는 순간 국가안보 차원의 전략적 대응 기회를 상실하게 된다. 한국은 중견국으로서 전략적 자율성을 유지하면서도 글로벌 AI 안보 거버넌스 구축 과정에서 기여할 수 있는 공간을 확장해야 한다. 이러한 접근은 한국이 AI 시대의 책임 있는 국가안보 행위자로 자리매김하기 위한 필수적 전략이자, 국익과 국제적 신뢰를 동시에 확보하기 위한 실천적 경로이다.

2. 군사적 측면에서 윤리 및 안전에 대한 논의의 중요성

AI 기술은 군사 분야에서 무기체계의 자동화, 감시 및 정찰, 전장 관리, 사이버전 등 다양한 영역에 적극적으로 활용되고 있다. AI는 뛰어난 효율성과 신속한 의사결정 지원 능력 덕분에 군사적 가치가 크게 주목받고 있지만, 동시에 그 활용이 초래할 수 있는 예상치 못한 부작용에 대한 우려도 커지고 있다. 특히, AI 기반 무기체계가 인간의 통제를 벗어나 독자적으로 판단하고 행동할 가능성이 존재하는 만큼 군사적 환경에서 AI를 적용하는 데 있어 더욱 신중하고 체계적인 논의가 필요하다. 무엇보다 오늘날의 전쟁은 단순한 무력 충돌을 넘어 국제법과 인도주의적 기준을 준수하는 윤리적 고려가 필수적인 시대에 접어들었다. 민간인 보호, 전투원과 비전투원의 구분, 무력 사용의 비례성 등은 국제인도법이 강조하는 핵심 원칙이며, AI가 개입하는 군사작전 역시 이러한 법적 및 윤리적 기준을 충족해야만 한다. 따라서

군사 분야에서 AI를 도입할 때는 기술적 효율성만을 중시할 것이 아니라, 그 운용이 윤리적 정당성과 국제 규범에 부합하는지를 면밀히 검토하는 것이 필수적이다.

군사적 의사결정 과정에 AI가 관여하는 경우 그 결정의 과정과 결과에 대한 투명성, 그리고 책임 소재를 명확히 규명하는 것이 중요하다. 인간의 생명과 직결되는 치명적 무력 사용과 같은 분야에서는 AI가 아닌 인간이 최종 결정을 내리는 인간 통제를 반드시 확보해야 하며, 이를 소홀히 여길 경우 AI의 오류나 오판이 심각한 윤리적 그리고 전략적 문제로 이어질 수 있다. 만약 윤리적 고려 없이 AI가 군사적으로 활용된다면, 작전 실패에 그치지 않고 국제법 위반이나 국제 사회로부터의 도덕적 정당성 상실로 이어질 위험이 존재한다.

또한, AI 기술의 특성상 군사적 긴장을 예상보다 빠르게 고조시키는 에스컬레이션(escalation) 위험 또한 무시할 수 없다. AI는 전장 상황에 신속히 대응하는 능력을 갖추었지만, 데이터 편향, 오작동, 적대적 공격(adversarial attack) 등의 요인으로 인해 잘못된 판단을 내릴 가능성도 존재한다. 다수의 국가나 세력이 AI 시스템을 자동화하여 상호 대응하는 상황에서는 인간의 직접적인 개입 없이도 충돌이 순식간에 확산될 수 있는 위험이 현실화될 수 있다. 이러한 가능성은 군사 AI 개발과 운용 과정 전반에 걸쳐 철저한 안전성 검증, 윤리적 통제, 그리고 국제적 규범에 기반한 협력 체계가 필수적임을 강하게 시사한다.

특히, 한국과 같은 중견국은 AI 군사기술에서 미국 및 중국과 같이 대규모 투자가 가능한 국가들과는 다른 전략적 조건에서 대응해야 한다는 점에서 현실적 역량을 고려한 단계별 접근이 필요하다. 우선, 단기적으로는 한국군에 적용 가능한 AI 군사윤리 기준과 안전성 검증 프로토콜을 마련하여, '인간 통제(Meaningful Human Control)'의 범위와 실행 절차를 명확히 설정하는 것이 중요하다. 중기적으로는 동맹국 및 파트너 국가와의 협력을 통해 군사 AI 위험평가, 적대적 공격 방어기술, 데이터 편향 검증 등 기술적 안전 분야에서 공동 연구 체계를 구축함으로써 독자적 자원 투자의 한계를 보완할 수 있다. 장기적으로는 한국이 중견국으로서 비교우위를 갖는 국제 규범 설계, 신뢰구축 메커니즘, 다자적 거버넌스 협력 분야에서 주도적 역할을 수행하며 군사적 AI 활용의 윤리·안전 기준을 설정하는 글로벌 논의에서 실질적 기여를 확대할 필요가 있다.

결국 군사적 AI 활용은 단순히 기술적 우위 확보를 위한 수단이 아니라, 인간성, 윤리성, 국제규범 준수라는 다층적 관점에서 종합적으로 접근해야 한다. 이러한 통합적 시각을 바탕으로 할 때에만 군사적 측면에서 AI는 현대 전쟁의 복잡성과 윤리적 요구를 동시에 충족시키며, 책임감 있는 방향으로 발전할 수 있을 것이다. 한국 역시 중견국의 전략적 위치를 활용하여 기술적, 제도적, 국제적 대응을 단계적으로 추진한다면, 군사 AI 시대에 필요한 윤리 및 안전 거버넌스를 구축하는 데 실질적 역할을 수행할 수 있을 것이다.

3. CBRN 영역에서 새로운 글로벌 질서 형성에 적극 참여

AI 기술의 급속한 발전은 CBRN 영역에서도 기존 국제 통제 체계의 근본적 재검토를 요구하고 있다. 생물무기금지협약(Biological Weapons Convention, BWC)과 화학무기금지협약(Chemical Weapons Convention, CWC) 등 기존 글로벌 규범은 전통적인 형태의 무기 개발과 사용을 통제하는 데 중점을 두고 있으며, AI 기반의 신종 위협을 다루는 조항은 부재한 상황이다. AI 기술이 CBRN 무기의 개발, 운용, 확산 방식에 본질적인 변화를 초래할 가능성이 현실화되고 있음에도 현행 규범 체계는 새로운 기술 환경을 충분히 반영하지 못하는 한계에 직면해 있다.

이러한 변화에 대응하여 미국은 AI 기반 CBRN 위협을 포괄하는 새로운 글로벌 질서 구축을 목표로 적극적인 정책적 준비를 진행하고 있는 것으로 판단된다. AI 기술이 야기할 수 있는 군사적 그리고 비군사적 CBRN 위협을 체계적으로 분석하고, 이를 통제할 새로운 규범 마련을 선도하려는 노력을 통해 향후 국제 규범 형성 과정에서 전략적 주도권을 확보하려는 움직임이 뚜렷

하게 나타나고 있다. 이는 단순한 위협 관리 차원을 넘어 기술 경쟁의 우위를 글로벌 규범 설정이라는 방식으로 제도화하려는 시도로 볼 수 있으며, 향후 미국이 AI-CBRN 관련 국제 규범의 주요 방향성을 주도할 가능성을 높이고 있다.

국제적 흐름 속에서 한국 역시 CBRN 영역의 새로운 글로벌 질서 형성 과정에 능동적으로 참여해야 할 필요성이 명확해지고 있다. AI 기술이 결합된 CBRN 위협에 대한 규범이 마련될 때까지 소극적으로 대응하거나 후발주자로 남을 경우 국가 이익과 기술 주권이 충분히 반영되지 못할 위험이 존재하기 때문이다. 특히, 신흥 기술을 규율하는 국제 규범은 초기 설계 단계에서의 논의 구조가 최종 형태를 결정짓는 경우가 많아 일단 구조가 확립된 이후에는 이를 수정하거나 이의를 제기하기가 극히 어려워지는 특성이 있다. 따라서 규범 수립 초기 단계부터 적극적으로 입장을 표명하고 논의 구조에 깊숙이 개입하는 것이 필수적이다.

이를 위해서는 우선 AI 기반 CBRN 기술이 야기할 위험성, 윤리적 그리고 법적 쟁점, 기술적 악용 경로에 대한 체계적 분석 능력을 확보하고, 이를 바탕으로 국제 사회에서 전문성과 신뢰성을 갖춘 발언권을 구축해야 한다. 한국은 생명공학, 정보보안, 원자력 안전 분야에서 축적한 기술력을 보유하고 있는 만큼 이를 활용해 AI-CBRN 위험 평가 기준, 안전성 검증 프로토콜, 모니터링 체계 등 기술 기반 규범 요소를 제안하는 역할을 수행할 수 있다. 동시에 국내적으로는 관련 기술 개발과 규범 논의의 연계를 강화하고, 정책 당국–학계–산업계 간 협력을 통해 대응 역량을 통합하는 체계를 구축할 필요가 있다. 한 예로, 국가 차원의 AI-CBRN 리스크 통합 분석 플랫폼 구축, 전문가 풀(pool) 운영, AI 오용 사례 연구 프로그램 등을 추진함으로써 국제 협상에서 활용 가능한 지식 기반을 확충할 수 있다.

또한, 한국은 중견국으로서의 위치를 활용하여 더욱 균형적 그리고 포괄적 규범 체계를 지지하는 외교 전략을 전개할 수 있다. 미국 중심의 규범화 움직임을 단순히 수용하기보다는 EU·영국·호주·일본 등 기술 동맹국뿐 아니라, ASEAN 및 중동 등 신흥 기술 수요국과도 협력 채널을 다변화하는 중재자 및 연결자 역할 수행이 가능하다. 이를 통해 한국은 투명성 및 검증 중심 규범, 오용 방지 원칙, 책임 있는 AI 활용 기준 등 국제사회가 공감할 수 있는 분야에서 실질적 기여를 확대할 수 있으며, 중견국 간 협의체 구성, 다자 기술 작업반 참여, 국제표준 기구(ISO/IEC)의 관련 표준 제정 등에서도 한국의 입지를 강화할 수 있다.

이처럼 AI 기술과 CBRN 위협이 결합한 새로운 국제 환경은 단순한 기술적 그리고 산업적 문제를 넘어 글로벌 거버넌스와 규범 경쟁이라는 전략적 과제로 전환되고 있다. 한국이 중견국으로서 갖는 기술 역량과 외교적 신뢰를 기반으로 단계적 접근을 수행한다면, 국가안보를 강화함과 동시에 첨단 기술 분야에서의 국제적 위상과 발언권을 확대하는 계기를 마련할 수 있을 것이다.

V. 결론

본 연구에서는 AI의 안보화(securitization)라는 관점을 바탕으로 AI 기술이 국가안보의 범주에 본격적으로 편입되고 있는 글로벌 동향을 종합적으로 분석했다. 21세기 이후 AI는 군사, 사이버, 정보, CBRN 등 다양한 안보 영역에서 핵심 전략 자산으로 기능하고 있으며, 이로 인하여 국제질서의 재편 및 국가 간 권력 구도의 변화에 중대한 영향을 미치고 있다. 특히, 미국과 중국을 중심으로 한 AI 패권 경쟁은 기술적 우위를 넘어 글로벌 규범 설정과 전략 질서의 주도권 확보를 목표로 치열하게 전개되고 있으며, 이에 따라 동맹 구조, 무기체계, 정보 환경, 기술 표준 등 다수의 안보 지형이 재구성되고 있다.

AI의 군사적 활용은 기존 무기체계의 자동화·지능화를 넘어 작전 계획, 정보 분석, 무력 사용 결정 등 전투지휘 전반에 걸친 구조적 변화를 유발하고 있다. 이는 작전 효율성 증대라는 긍정적 효과와 함께 인간의 판단 배제, 갈등 및 위기 확산의 가능성, 책임 소재 불명확성 등 심각한 전략적·윤리적 문제를 수반한다. 동시에 AI와 CBRN 기술의 융합은 고위험 물질의 개발 및 운용이

더욱 용이해지는 환경을 조성하며, 비국가행위자의 접근 가능성 확대, 의도치 않은 실험 결과 등 복합적 위협을 증대시키고 있다. 더 나아가 생성형 AI의 발전은 정보 조작 및 심리전에 악용됨으로써 민주주의 기반을 훼손하고 사회적 분열을 심화시키는 새로운 형태의 인지전을 현실화시키고 있다.

이러한 변화는 국가안보 정책의 재정립을 요구한다. 첫째, AI 기술은 더 이상 국내 기술진흥 또는 산업정책의 하위범주가 아닌, 국가 생존과 직결된 전략 영역으로 인식되어야 하며, 이를 반영한 통합적 정책체계가 구축되어야 한다. 둘째, AI 기반 군사기술의 도입은 국제법과 인도주의 원칙에 부합하는 윤리적 기준을 바탕으로 설계되어야 하며, 인간 중심 통제 구조의 제도화를 통해 AI 오작동이나 무력 오남용에 대한 구조적 대응체계를 갖추는 것이 시급하다. 셋째, AI-CBRN 위협에 대응하기 위한 글로벌 규범 형성 과정에 적극적으로 참여함으로써 한국은 기술 종속의 위험을 줄이는 동시에 국제규범 설계자로서의 전략적 입지를 확보해야 한다. 넷째, 허위 정보와 인지전 확산에 대응하기 위해 AI 기반 탐지 기술 고도화, 사회적 정보 면역력 강화, 플랫폼 책임 강화 등 다층적 접근이 필요하다.

향후 연구는 다음과 같은 방향으로 확장될 수 있다. 첫째, AI 기술과 국가안보의 연계성을 반영하는 통합적 정책 프레임워크 개발을 통해 기술, 외교, 군사, 법제 간의 연계성을 확보할 필요가 있다. 둘째, AI와 국제규범 간의 접점을 중심으로 주요국의 규범 전략을 비교 분석하고, 한국의 전략적 대응 방안을 도출하는 연구가 요구된다. 셋째, AI 기술의 군사·비군사 활용에 있어 윤리성과 책임성을 확보하기 위한 글로벌 거버넌스 모델 개발과 적용 가능성에 대한 탐색이 필요하다. 마지막으로 AI 시대에 AI 안보화에 대한 메커니즘을 구체적으로 분석함으로써 정책결정자들을 위한 미시적인 측면에서의 시사점을 제공할 필요가 있다.

■ References

김법연 (2024). 생성형 AI의 법적 문제와 규제 논의 동향. 〈정보화정책〉, 31권 4호, 3-33. https://doi.org/10.22693/NIAIP.2024.31.4.003

김병운 (2016). 인공지능 동향분석과 국가차원 정책제언. 〈정보화정책〉, 23권 1호, 74-93. https://doi.org/10.22693/NIAIP.2016.23.1.074

민병원 (2006). 탈냉전시대의 안보개념 확대: 코펜하겐 학파, 안보문제화, 그리고 국제정치이론. 〈세계정치〉, 5권, 13-62.

민병원 (2012). 안보담론과 국제정치: 안보개념의 역사적 변화를 중심으로. 〈평화연구〉, 20권 2호, 203-239.

안진우·노상우·김태환·윤일웅 (2020). 인공지능 분야 국방 미래기술에 관한 실증연구. 〈한국산학기술학회논문지〉, 21권 5호, 409-416. https://doi.org/10.5762/KAIS.2020.21.9.458

이근욱 (2009). 자유주의 이론과 안보: 모순된 조합인가 새로운 가능성인가? 〈국제정치논총〉, 49권 5호, 33-53.

이재은 (2013). 국가안보 환경의 변화와 국가위기관리: 포괄적 안보 개념 하에서의 국가위기 유형. 〈Crisisonomy〉, 9권 2호, 177-198.

장기영 (2024). 인공지능과 미래 안보에 대한 국제정치 이론적 전망: 국가 및 사회 수준과 개인 수준을 중심으로. 〈국제정치논총〉, 64권 1호, 7-41. https://doi.org/10.14731/kjir.2024.03.64.1.7

전웅 (2004). 국가안보와 인간안보. 〈국제정치논총〉, 44권 1호, 25-49.

정민섭·남궁승필·박상혁 (2020). 신흥안보 창발과 미래 사회 및 자연환경 변화예측. 〈The Journal of the Convergence on Culture Technology〉, 6권 2호, 327-331. https://doi.org/10.17703/JCCT.2020.6.2.327.

조은일 (2024). 신기술은 어떻게 국제안보를 변화시키는가: 인공지능, 드론, 극초음속의 군사적 활용과 국제안보에 대한 연구. 〈국제지역연구〉, 33권 3호, 71-98. https://doi.org/10.56115/RIAS.2024.9.33.3.71

황성수·신용호 (2019). Mobility 신산업 동향 및 쟁점, 그리고 정부의 역할: O2O, 승차공유, 택배, 물류 분야의 전망 및 규제연구를 중심으로. 〈정보화정책〉 26권 2호, 3-23. https://doi.org/10.22693/NIAIP.2019.26.2.003

Abraham, Y. (2024). Lavender: *The AI machine*

directing Israel's bombing spree in Gaza. +972 Magazine. https://www.972mag.com/lavender-ai-israeli-army-gaza/

Asia Pacific Task Force (2025). *The age of AI in U.S.-China great power competition: Strategic implications, risks, and global governance.* Beyond the Horizon. https://behorizon.org/the-age-of-ai-in-u-s-china-great-power-competition-strategic-implications-risks-and-global-governance/

Bandouil, S. (2025). *German company to manufacture 6,000 AI-powered drones for Ukraine.* The Kyiv Independent. https://kyivindependent.com/german-company-to-manufacture-6-000-ai-powered-drones-for-ukraine/

Barrett, A. M., Jackson, K., Murphy, E. R., Madkour, N. & Newman, J. (2024). Benchmark early and red team often: A framework for assessing and managing dual-use hazards of AI foundation models. *arXiv preprint arXiv:2405.10986.* https://doi.org/10.48550/arXiv.2405.10986

Bengio, Y., Mindermann, S., Privitera, D., Besiroglu, T., Bommasani, R., Casper, S., Choi, Y., Fox, P., Garfinkel, B., Goldfarb, D., Heidari, H., Ho, A., Kapoor, S., Khalatbari, L., Longpre, S., Manning, S., Mavroudis, V., Mazeika, M., Michael, J., Newman, J., Ng, K. Y., Okolo, C. T., Raji, D., Sastry, G., Seger, E., Skeadas, T., South, T., Strubell, E., Tramèr, F., Velasco, L., Wheeler, N., Acemoglu, D., Adekanmbi, O., Dalrymple, D., Dietterich, T. G., Felten, E. W., Fung, P., Gourinchas, P.-O., Heintz, F., Hinton, G., Jennings, N., Krause, A., Leavy, S., Liang, P., Ludermir, T., Marda, V., Margetts, H., McDermid, J., Munga, J., Narayanan, A., Nelson, A., Neppel, C., Oh, A., Ramchurn, G., Russell, S., Schaake, M., Schölkopf, B., Song, D., Soto, A., Tiedrich, L., Varoquaux, G., Yao, A., Zhang, Y.-Q., Albalawi, F., Alserkal, M., Ajala, O., Avrin, G., Busch, C., Carvalho, A. C. P. L. F. de, Fox, B., Gill, A. S., Hatip, A. H., Heikkilä, J., Jolly, G., Katzir, Z., Kitano, H., Krüger, A., Johnson, C., Khan, S. M., Lee, K. M., Ligot, D. V., Molchanovskyi, O., Monti, A., Mwamanzi, N., Nemer, M., Oliver, N., López Portillo, J. R., Ravindran, B., Pezoa Rivera, R., Riza, H., Rugege, C., Seoighe, C., Sheehan, J., Sheikh, H., Wong, D. & Zeng, Y. (2025). *International AI safety report.* arXiv preprint arXiv:2501.17805. https://doi.org/10.48550/arXiv.2501.17805

Bondar, K. (2025). *Ukraine's future vision and current capabilities for waging AI-enabled autonomous warfare.* Center for Strategic and International Studies. https://www.csis.org/analysis/ukraines-future-vision-and-current-capabilities-waging-ai-enabled-autonomous-warfare

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., Ó hÉigeartaigh, S., Beard, S. J., Belfield, H., Farquhar, S., Lyle, C., Crootof, R., Evans, O., Page, M., Bryson, J., Yampolskiy, R., & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228.* https://doi.org/10.48550/arXiv.1802.07228

Department for Science, Innovation and Technology & AI Security Institute. (2025). *Tackling AI security risks to unleash growth and deliver plan for change.* GOV.UK. https://www.gov.uk/government/news/tackling-ai-security-risks-to-unleash-growth-and-deliver-plan-for-change

DeVine, M. E. (2024). *The Intelligence Community's Foreign Malign Influence Center.* Congressional Research Service. https://www.everycrsreport.com/reports/IF12470.html

Geist, E. & Lohn, A. J. (2018). *How might artificial intelligence affect the risk of nuclear war.* RAND Corporation. https://www.rand.org/

pubs/perspectives/PE296.html

Hannas, W. C. & Chang, H. M. (2021). China's 'new generation' AI-brain project. *PRISM, 9*(3), 18-33. https://www.jstor.org/stable/48640743

Horowitz, M. C. (2018). Artificial intelligence, international competition, and the balance of power. *Texas National Security Review, 1*(3), 37-57. http://hdl.handle.net/2152/65638

Jouvenet, P. (2025). *How China competes with the U.S. in artificial intelligence.* Eurasia Business News. https://eurasiabusinessnews.com/2025/01/25/how-china-competes-the-u-s-in-artificial-intelligence/

Kania, E. B. (2019). China's pursuit of military advantage through cognitive science and biotechnology. *PRISM, 8*(3), 83-101. https://www.jstor.org/stable/26864278

Karthikeyan, S. P. (2024). Rising threat of AI-driven cybersecurity attacks: Implications for national security. *International Journal for Research in Applied Science and Engineering Technology, 12*(9), 756-762. https://doi.org/10.22214/ijraset.2024.64042

Kwet, M. (2019). Digital colonialism: US empire and the new imperialism in the global South. *Race & Class, 60*(4), 3-26. https://doi.org/10.1177/0306396818823172

Levy, J. S. (1983). Misperception and the causes of war: Theoretical linkages and analytical problems. *World Politics, 36*(1), 76-99. https://doi.org/10.2307/2010176

Licata, N. R. (2023). *China's military-civil fusion strategy: A blueprint for technological superiority.* Foreign Policy Research Institute. https://www.fpri.org/article/2023/12/chinas-military-civil-fusion-strategy-a-blueprint-for-technological-superiority/

Lin, L. & Huang, R. (2025). *China's Huawei develops new AI chip, seeking to match Nvidia*. Wall Street Journal. https://www.wsj.com/tech/chinas-huawei-develops-new-ai-chip-seeking-to-match-nvidia-8166f606

Ministry of Industry and Information Technology of the People's Republic of China (2024). *China unveils plan to better protect data of industrial firms.* State Council of the People's Republic of China. https://english.www.gov.cn/news/202402/26/content_WS65dc9644c6d0868f4e8e45b9.html

Ministry of National Defense of the People's Republic of China (2019). China's national defense in the new era. *State Council Information Office.* https://english.www.gov.cn/archive/whitepaper/201907/24/content_WS5d3941ddc6d08408f502283d.html

Morgan, H. (2022). Conducting a qualitative document analysis. *The Qualitative Report, 27*(1), 64-77. https://doi.org/10.46743/2160-3715/2022.5044

Moy, W. R. & Gradon, K. T. (2023). *Artificial intelligence in hybrid and information warfare: A double-edged sword.* In F. Cristiano, D. Broeders, F. Delerue, F. Douzet & A. Géry (Eds.), Artificial intelligence and international conflict in cyberspace (pp. 47-74). London: Routledge.

NATO Allied Command Transformation (n.d.). *Cognitive warfare*. NATO. https://www.act.nato.int/activities/cognitive-warfare/

Pohle, J. & Thiel, T. (2020). Digital sovereignty. *Internet Policy Review, 9*(4). https://doi.org/10.14763/2020.4.1532

Reed, V. (2024). *Global AI arms race: Military tech escalates worldwide*. AI Competence. https://aicompetence.org/global-ai-arms-race-military-tech-escalates/

Simmons-Edler, R., Badman, R., Longpre, S. & Rajan, K. (2024). AI-powered autonomous weapons risk geopolitical instability and threaten AI research. *arXiv preprint arXiv:2405.01859.* https://doi.org/10.48550/arXiv.2405.01859

State Council of China (2017). *A next generation*

artificial intelligence development plan. In DigiChina, Stanford University (Trans.). https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

TechNode Feed (2025). *Embodied intelligence appears in government work report for the first time at NPC.* TechNode. https://technode.com/2025/03/06/embodied-intelligence-appears-in-government-work-report-for-the-first-time-at-npc/

The White House (2023). *National cybersecurity strategy.* https://bidenwhitehouse.archives.gov/oncd/national-cybersecurity-strategy/

The White House (2024). *Memorandum on advancing the United States' leadership in artificial intelligence, harnessing artificial intelligence to fulfill national security objectives, and fostering the safety, security, and trustworthiness of artificial intelligence*. https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2024/10/24/memorandum-on-advancing-the-united-states-leadership-in-artificial-intelligence-harnessing-artificial-intelligence-to-fulfill-national-security-objectives-and-fostering-the-safety-security/

The White House (2025a,). E*xecutive order 14179: Removing barriers to American leadership in artificial intelligence.* https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/

The White House (2025b,). America's AI Action Plan. https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf

U.S. Department of Defense (2018). *Summary of the 2018 National Defense Strategy of the United States of America: Sharpening the American military's competitive edge.* https://media.defense.gov/2020/May/18/2002302061/-1/-1/1/2018-NATIONAL-DEFENSE-STRATEGY-SUMMARY.PDF

U.S. Department of Defense (2019). *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to advance our security and prosperity.* https://media.defense.gov/2019/feb/12/2002088963/-1/-1/1/summary-of-DoD-AI-Strategy.pdf

U.S. Department of Defense (2021). *Implementing responsible artificial intelligence in the Department of Defense.* https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF

U.S. Department of Defense (2022a). *Memorandum on the initial operating capability of the Chief Digital and Artificial Intelligence Officer.* https://media.defense.gov/2022/Feb/02/2002931807/-1/-1/1/MEMORANDUM-ON-THE-INITIAL-OPERATING-CAPABILITY-OF-THE-CHIEF-DIGITAL-AND-ARTIFICIAL-INTELLIGENCE-OFFICER.PDF

U.S. Department of Defense (2022b). *Responsible artificial intelligence strategy and implementation pathway.*https://media.defense.gov/2022/Jun/22/2003022604/-1/-1/0/Department-of-Defense-Responsible-Artificial-Intelligence-Strategy-and-Implementation-Pathway.PDF

U.S. Department of Defense (2023a). *Data, analytics, and artificial intelligence adoption strategy.* https://media.defense.gov/2023/Nov/02/2003333300/-1/-1/1/DOD_DATA_ANALYTICS_AI_ADOPTION_STRATEGY.PDF

U.S. Department of Defense (2023b). *Deputy Secretary of Defense Kathleen Hicks announces publication of Data, Analytics, and AI Adoption Strategy.* https://www.defense.gov/News/Releases/Release/Article/3577857

U.S. Department of State (n.d.). *The Chinese Communist*

Party's military-civil fusion policy. https://2017-2021.state.gov/military-civil-fusion/

U.S. National Institute of Standards and Technology (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile.* https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

U.S. National Institute of Standards and Technology (2025). *Managing misuse risk for dual-use foundation models (2nd Public Draft).* https://doi.org/10.6028/NIST.AI.800-1.2pd

Yang, Z. (2025). *How Chinese AI startup DeepSeek made a model that rivals OpenAI.* Wired. https://www.wired.com/story/deepseek-china-model-ai/