

www.eai.or.kr

EAI 워킹페이퍼

군사 부문: AI 시대 남북관계의 리스크와 기회

김양규(국방대학교)

군사 부문: AI 시대 남북관계의 리스크와 기회

김양규

국방대학교 교수

I. 서론

‘인간의 지능이 요구되는 과업을 수행할 수 있는’ 인공지능(AI)은 말단의 지엽적 첨단기술이 아닌 인간 사회 전체의 향방을 바꿀 수 있는 범용기술이다. 군사 분야에서의 활용도 단순히 공상과학소설에 등장하는 자율무기체계에 탑재된 로봇 군단에 활용될 가능성이 있는 미래 기술이 아닌, 현재 전장에서 감시정찰 및 인식, 군사적 판단, 지휘통제, 억제제 작동 방식과 조건을 모두 바꿀 수 있는 기술이다. 따라서 AI의 군사적 효용은 특정 무기체계 하나를 자동화하는 데 그치지 않는다. 방대한 데이터를 빠르게 처리하고, 이질적 센서 정보를 융합하며, 표적을 식별하고, 작전 방안을 생성하며, 지휘관의 결심을 보조한다. 이러한 점에서 AI는 모든 영역에서 활약하는 “가능케 하는 기술(enabler)”이자 “전력증강자(force multiplier)”이다. 특히 정확성과 투명성 혁명이 진행되는 현대전에서 누가 먼저 보고, 누가 먼저 판단하며, 누가 먼저 타격하는지를 결정하는 ‘속도’가 핵심 변수로 부상하고 있고, AI는 바로 이 속도에서 큰 잠재적 이점을 제공한다. 그러나 그 효과는 데이터의 품질, 통신망의 생존성, 인간-기계 상호작용, 적의 전자전 및 기만 능력에 따라 달라지며, 오류와 과신을 동시에 증폭시킬 수 있다(김양규 2024; 김양규 2026b).

한반도에서 이 변화는 더욱 위험한 함의를 갖는다. 한반도는 지리적 중심이 짧고, 공격무기들의 비행시간이 짧으며, 남북 간 군사적 적대성이 높고, 군과 군 사이 통신선과 위기관리 채널은 불안정하다. 여기에 북한의 핵전력 고도화, 전술 핵 운용 가능성, 재래식·핵 이중 열세, 한국의 3축체계와 첨단 재래식 정밀타격 능력이 동시에 작동하고 있다. 이런 조건에서 AI는 단순히 군사 효율성을 높이는 기술이 아닌 위기 시 오판과 선제사용 유인을 증폭시켜 원하지 않는 확전의 연쇄반응을 일으킬 수 있는 변수로 작동한다.

이러한 맥락에서 본 장에서 다루고자 하는 핵심 질문은 다음과 같다. 첫째, 북한은 AI를 어떤 방식으로 군사·안보 영역에 활용하려 하고 있는가? 둘째, AI는 남북 군사환경과 전략안정성을 어떻게 변화시키는가? 셋째, 북한의 군사 AI 미래 옵션은 대중(對中)·대러(對露) 기술협력, 제재회피형 자체 활용 및 장기적 남북 위험감축 경로 가운데 어떤 방향으로 전개될 가능성이 큰가? 넷째, 군사 부문에서 남북협력이 현실적으로 어렵다면, 한국은 어떤 방식으로 전략적 안정성을 도모해야 하는가?

각각의 이슈와 관련하여 본 장은 네 가지 주장을 제시한다. 첫째, 현재 북한이 추진 중인 AI 전력화 방향을 고려할 때 향후 위협은 북한이 단기간에 미국이나 중국 수준의 초정밀 군사 AI 체계를 보유할 가능성에서 찾기는 어렵다. 북한의 GPU 및 메모리 칩의 한계에서 비롯되는 컴퓨팅 파워의 명확한 한계, 전력 불안정 공급, 최첨단 독자 모델 개발의 어려움, 안정적인 전문인력 확보의 제약 등을 고려할 때 그럴 가능성은 매우 낮고 이런 위협을 과장할 필요는 없다. 보다 현실적인 위협은 북한이 제한된 연산 및 데이터 환경에서도 운용 가능한 경량화된 AI, 중국 또는 러시아와의 협력 하에 고도화될 가능성이 있는 감시, 식별, 추적 기술, 음성 및 영상 합성을 통한 사이버 공격, 자동화 및 무인기 기술을 방어보다 공격에 활용하고, 특히 한국이 가장 취약한 부분을 찾아내서 교란을 시도하는 비대칭 전력으로 활용하는 데 있다. 북한은 1990년대 후반부터 패턴 인식, 음성 처리, 데이터 최적화 연구를 축적해 왔으며, 최근에는 안면인식, 다중 객체 추적, 음성합성 경량화, 역양 식별 등 군사 및 사이버 전용 가능성이 높은 기술 영역에서 진전을 보이고 있다는 점에 주목할 필요가 있다(김민정 2025, 1-5).

둘째, AI는 한국의 전략 비핵무기(SNNW)와 재래식 대군사타격(counterforce) 능력을 크게 제고할 가능성이 크다. AI 기반 ISR, 디지털 표적화, 지휘통제 자동화, 작전결심 지원은 한국의 킬체인과 장거리 정밀타격 능력을 고도화할 수 있다. 그러나, 이 능력이 무제한적으로 추구될 경우 북한은 이를 핵 무력의 무력화 또는 참수작전의 전조로 인식할 수 있다. 문제는 이러한 변화가 현재 상당 부분 정착기에 들어선 북한의 공세적 핵 교리 및 한미동맹 전력에 대비해 볼 때 여전한 재래식·핵 능력 이중 열세 구조와 결합될 경우 북한의 “먼저 쓰지 않으면 잃는다(use-it-or-lose-it)”라는 불안을 자극할 수 있다 (김양규 2026).

셋째, 군사영역에서 남북협력은 매우 비현실적이다. 북한은 AI를 국가 주권, 체제 생존, 무력 현대화와 결부된 전략 자산으로 인식하고 있으며(북한 AI 연구 TF 2026), 한국은 AI 기반 전략적 비핵무기와 정밀타격 능력으로 북한의 핵 능력 증강에 대응하려 하고 있다(국방부 2022). 즉, 당분간 남북간 군비 경쟁 국면이 지속되고 강화되는 것은 피하기 어려운 엄중한 현실이다(Kim

2025). 따라서 AI의 군사적 이용, 즉 ISR, 지휘통제(C2), 무인체계, 사이버 방어 및 공격, 표적식별과 관련된 남북협력은 기술 이전, 제재 위반, 군사정보 노출, 이중용도 악용 위험 때문에 실현 가능성이 극히 낮다. 그렇기 때문에 군사 분야에서는 남북간 “협력 가능성”을 논의하기보다는 오히려 이러한 어려운 상황 속에서도 어떻게 한반도에서 전략적 안정성을 담보하여 다른 영역에서 남북 협력의 물꼬를 틀게 만들 수 있는가에 달려있다.

넷째, AI 시대 한반도의 전략적 안정성을 위한 현실적 경로는 무제한적 AI 기반 공격 및 대군사타격 경쟁이 아닌 관리된 형태의 군비경쟁이다. 한국은 전쟁 발발을 억지하면서도 북한의 핵 선제사용 유인을 낮추는 방향으로 군사력을 설계해야 한다. 선제타격 자동화보다 대량응징보복(KMPR), 지휘통제 회복탄력성, 조기경보 신뢰성, 확증보복 능력 강화 영역에 AI의 군사적 활용 노력을 집중해야 한다. 최후의 보루로서 대군사타격 능력은 완전히 포기할 수 없지만, 엄격한 정치적 승인, 의미 있는 인간 통제, 충분한 숙고시간, 다중 정보확인, 감사 가능성, 시간 및 공간적 완충장치가 작동하는 마지막 안전장치로 한정되어야 한다.

II. 북한의 군사 AI 활용과 안보적 위험

1. 북한 AI 담론의 변화: 기술 추격에서 국가전략으로

AI 시대 남북 군사력 변화 방향을 분석하기 위해 가장 먼저 살펴보아야 할 것은 북한이 군사영역에서 AI 기술을 어떻게 인식하고 있는가의 문제이다. 복수의 연구들은 최근 북한의 AI 담론이 단순한 새로운 기술 소개나 산업정책 차원을 넘어 국가전략 담론으로 진화하고 있다고 지적한다. 『노동신문』 보도에 따르면 북한은 AI를 국가 주권, 기술 독점, 군사 혁신, 체제 존립과 결부된 필수 과제로 새롭게 정의하는 경향이 보인다. 특히, 2024년 이후 북한 매체는 해외 AI 경쟁, 러시아의 국가 주도 AI 전략, AI의 군사적 파급력을 반복적으로 보도하며, AI를 선택 가능한 정책 수단이 아닌 제대로 대응하지 않을 경우 전략적 불이익을 초래할 수 있는 국가적 과제로 위치시키는 모습을 볼 수 있다(북한 AI 연구 TF 2026, 1-4).

이러한 담론 변화는 AI 기술 변화에 대응하는 정책 전환 노력을 내부적으로 정당화하려는 과정으로 이해할 수 있다. 앞서 지적한 바와 같이 AI는 경제성장, 과학기술 발전, 군사 현대화, 체제 생존을 동시에 상징할 수 있는 범용기술이다. 북한 당국은 노동당 제9차 대회를 통해 “수중발사형의

대륙간탄도미사일종합체와 각이한 인공지능무인공격종합체들, 유사시 적국의 위성을 공격하기 위한 특수자산과 적의 지휘중추를 마비시키기 위한 매우 강력한 전자전무기체계들, 더욱 진화된 정찰위성”을 첨단 기술군의 혁신 과제로 제시하며 공식 당대회에서 최초로 AI를 북한이 추구하는 신형무기 가운데 하나로 언급하였다(신범철 2026, 2-3).

이를 통해 북한은 AI 기술 분야 개발에 자원을 집중하고 사회적 동원을 이끌어내고자 할 것으로 전망된다. 공식적인 담론 수준에서 AI는 “잘하면 성장할 수 있는 기술”이면서 동시에 “뒤처지면 회복 불가능한 손실을 초래할 수 있는 기술”로 제시된다. 이는 AI 개발을 위한 자원 배분, 인재 동원, 군사 부문 우선순위 설정을 정당화하는 정치적 언어가 된다(북한 AI 연구 TF 2026).

특히 주목할 점은 북한이 AI를 군사 분야에서 “화약과 핵무기에 이은 제3의 혁명”으로 소개하는 외부 평가를 인용하고 있다는 것이다. 당국자에 의한 공식 선언이 아닌 외부 담론의 인용 형식을 취하고 있지만, 이는 AI 기반 무기체계 도입이 선택적 군사 혁신이 아닌 현대전의 변화 속에서 불가피한 과제라는 인식을 드러낸다. 2024년 이후 김정은이 직접 무인기와 인공지능 기반 무력 현대화를 강조한 점 또한 AI가 북한 국방전략의 주변부가 아닌 핵심 과업 영역에 이미 진입하였음을 보여준다(북한 AI 연구 TF 2026, 3-5).

2. 북한 AI 기술역량: 첨단 AI 모델이 아닌 경량화·최적화 AI

담론 수준에서 북한이 AI 역량을 강조하고 있다고 해서 이를 두고 북한이 미국이나 중국이 벌이고 있는 수준의 AI 기술 전략경쟁에 뛰어들고자 하는 것으로 생각하는 것은 매우 큰 오류이다. 북한은 GPU와 같은 대규모 컴퓨팅 인프라, 독자적인 AI 모델, 글로벌 오픈소스 생태계 활용에서도 하드웨어 접근, 국제 협업 및 기술정보 획득의 구조적 한계를 보인다. 따라서 글로벌 AI 경쟁 무대에서 선진국들과 상당한 격차가 존재한다. 이러한 제약을 고려할 때, 현재 북한이 집중하고 있는 것은 제한된 연산환경에서도 작동하는 경량화된 AI(small AI), 알고리즘 최적화, 실시간 감시, 식별, 추적 기술인 것으로 보인다(김민정 2025). 다만 공개자료는 북한의 비공개 군사연구를 직접 보여주지 못하므로, 이 평가는 북한의 전반적 역량에 대한 확정적 판단이 아니라 공개적으로 관찰 가능한 기술 방향에 대한 제한적 평가로 이해해야 한다.

북한은 1990년대 후반부터 패턴 인식, 음성 처리, 수리적 최적화 등을 중심으로 AI 연구 기반을 축적해 왔고, 2010년대 이후 김일성종합대학과 김책공업종합대학 등에서 AI 관련 연구조직을 확충한 것으로 보인다. 공개자료상 확인되는 최근 연구는 안면인식 함수 개선, 다중 인물 추적, 음성합성 경량화, 한국어 억양 식별 등이다. 이 연구들은 비록 민간 분야 연구로 제시되었지만, 군사적 전용 가능성이 크다. 안면인식과 다중 객체 추적은 제한된 영상 품질에서도 표적 식별과 이동 경로 예측을 가능하게 하며, 음성합성과 억양 식별은 신원 위장, 지휘통신 교란 등 사이버 공격의 효율성을 높일 수 있다(김민정 2025, 3-8).

이러한 특성은 북한식 AI 활용의 방향을 시사한다. 북한은 단기간에 첨단 군사 의사결정 AI 지원 체계나 전장 모델을 구축하기 어렵다. 그러나 저비용 및 비대칭적 용도에 한정할 시 상당한 수준의 안보위협을 가할 수 있다. 예컨대 AI 기반 피싱, 딥페이크, 악성코드 자동화, 암호자산 탈취, 신원 위장, IT 노동자 네트워크 운영, 감시카메라 기반 내부통제, 제한적 표적식별, 무인기 영상처리 등은 북한의 제약 조건에도 시도해 볼 수 있다. 따라서, 향후 북한이 AI 기반 자동화 기술을 활용한 사이버 공격을 감행할 경우 암호자산 탈취, 자금세탁, 위장 신원 생성의 속도와 규모가 비약적으로 증가할 가능성이 있다(김민정 2025, 3).

3. 북한 군사 AI의 우선 적용 영역

최근 보고서 분석을 중심으로 살펴볼 때, 북한의 군사 AI 활용은 다음 네 영역에서 집중적으로 나타날 가능성이 크다(김민정 2026; 윤훈희 2026; 신범철 2026).

첫째, 무인기와 자율무기 영역이다. 김정은은 2024년부터 무인기 개발에 AI를 적극 도입할 것을 지시했고, 2025년에는 무인장비와 인공지능기술 분야를 무력 현대화 건설에서 최우선적으로 중시해야 한다고 강조했다. 2026년 2월 9차 당대회에서도 “각이한 인공지능무인공격종합체들”이 상용 무기체계에 도입될 것을 주문했다. 현재 북한은 전략정찰 무인기, 다목적 공격형 무인기, 전술급 자폭형 무인기 등 다양한 무인체계를 공개하고 있으며, 여기에 AI 역량이 결합될 경우 표적 인식, 경로계획, 군집 운용, 영상분석, 전자전 회피에 활용될 수 있다.

둘째, 사이버 및 정보전 영역이다. 북한은 이미 사이버 작전, 암호자산 탈취, 원격 IT 노동자 네트워크, 사회공학 기법에서 세계적 수준의 역량을 보여준 바 있다. AI를 적극적으로 활용할 시 기존의 사이버 작전 활동 속도와 규모가 더 강화될 수 있다. 피싱 메일의 개인화, 음성 및 영상

답페이지, 악성코드 개발 보조, 취약점 탐색, 가짜 신분증 생성, 암호자산 세탁 자동화는 북한이 상대적으로 낮은 비용으로 큰 효과를 얻을 수 있는 영역이라고 볼 수 있다.

셋째, 감시·정찰 및 표적획득 영역이다. 북한의 군사정찰위성, 중·고도 정찰 무인기, 감시카메라, 영상처리, 다중 객체 추적 기술은 전장 상황인식 향상에 기여할 수 있다. 물론 현재 미국이 전장에서 구현하고 있는 수준의 의사결정지원 모델과 비교할 수 있는 수준은 아니다. 북한이 보유한 인공위성의 중간 수준에 불과한 해상도(Van Diepen 2026), 대규모 컴퓨팅 인프라 활용 능력에서 명확한 한계 및 이로 인한 데이터 처리 능력 제한 등을 고려할 때 북한이 독자적으로 다음 절에서 설명하는 현대 AI의 군사적 활용의 핵심 영역인 디지털 표적화 능력을 갖추기는 매우 어렵다. 현재로서는 AI 능력보다도 정기적인 감시망을 구축하여 실질적인 상황 인식 능력을 향상시킬 수 있는 복수의 위성 네트워크 구축이 북한에게는 훨씬 더 절실한 상황이다.

넷째, 핵 전력 운용 보조 영역이다. 미중을 비롯한 대부분의 나라에서 아직 핵무기 지휘통제(NC3)에 AI가 직접 통합되었다고 공개적으로 확인된 사례는 제한적이며, 핵 의사결정체계의 보안성과 불투명성을 고려하면 실제 통합 수준을 외부에서 확인하기 어렵다. 북한 역시 단기간에 핵 사용 결정 자체를 AI에 위임할 가능성은 낮다. 그러나, 핵 운용환경은 핵탄두와 투발수단만으로 구성되지 않는다. 핵 역지가 작동하기 위해서는 정찰, 감시, 표적획득, 지휘통제, 통신보호, 조기경보, 피해평가 등이 필요하고 무엇보다 2차 공격능력(second-strike capability) 담보를 위해 보복 확증성을 제고해야 한다(김양규 2024). 북한의 2022 핵무력정책법이 핵무력의 구성요소로 핵탄두 및 운반수단뿐 아니라 지휘·조종체계, 운용과 갱신을 위한 인원, 장비, 시설을 포함하고 있다는 점에서 핵 운용환경과 사이버-AI의 간접적 연계는 중요하다.

4. 현실적인 안보위협: 북한 군사 AI의 비대칭 효과

북한의 군사 AI는 독립적으로 전쟁의 판도를 바꾸기보다 기존 북한의 비대칭 전략을 증폭시키는 방향으로 작동할 가능성이 크다. 이는 전력증강자로 기능하는 AI의 일반적인 속성 때문이기도 하다. 북한은 한미동맹에 비해 여전히 재래식 전력에서 상당한 열세이며, 핵전력에서도 미국의 확장억제와 비교하면 구조적으로 매우 큰 폭의 열세에 있다. 2025년 기준으로 미국의 실전 배치된 핵탄두가 1,770기이며, 해체를 기다리는 퇴역탄두까지 포함한 총재고는 5,177기다. 북한은 공개정보에 기초한 저장탄두 추정치가 약 50기인 것으로 평가된다(Kristensen and Korda 2025). 물론 이러한 수량

비교만으로 실제 억제력의 격차를 판단하기는 어려우며, 전력의 생존성, 투발수단, 경보태세와 동맹구조를 함께 고려해야 한다. 그러나, 북미 간 핵무기 역량의 압도적 격차로 인해 북한이 제한된 자원을 비대칭적으로 활용하여 상대의 취약점을 공격하는 전략을 선호해 왔다는 것은 주지의 사실이다. AI는 북한의 이러한 전략에 부스터 역할을 한다.

첫째, 가장 현실적인 위협은 AI를 북한의 사이버 비대칭성 역량 강화에 활용하는 것이다. 사이버 공격은 낮은 비용, 귀책(attribution)의 어려움, 비대칭 효과, 제재 회피 수익 창출을 동시에 제공한다. 경량화된 AI로도 사이버 공격의 성공률을 높이고, 공격 자동화를 가능하게 하며, 감시, 정찰 및 자료 분석과 정보전 콘텐츠 생산을 가속하는데 활용할 수 있다. 따라서, 사이버전 차원에서 군사 AI의 활용이 가장 위협적이라고 볼 수 있다.

둘째, AI는 북한의 무인기 등 저가 무기체계의 정밀공격 능력을 높일 수 있다. 북한이 미국이나 중국과 같은 수준의 대규모 컴퓨팅 인프라나 우주 기반 ISR 체계를 확보하는 것은 불가능하지만, 중국 및 러시아와 협력하여 저가 무인기와 영상처리 경량화 AI, 전자전 회피 알고리즘, 자폭 드론 운용을 결합하면 현재 가지고 있는 비대칭 전략의 대남 군사 위협을 제고할 수 있다.

셋째, AI는 핵과 비핵 수단의 경계를 허물며 전략적 안정성을 훼손하는 '핵 얽힘(entanglement)'의 위험을 증폭시킨다(김양규 2024; 윤정현 2026). AI 기반의 전략적 비핵무기(SNNW)는 사이버 개입, 데이터 조작, 기만 신호 등을 통해 신호를 왜곡시킨다. 이러한 정보 환경에서는 상대의 행동이 방어인지 공격인지 명확히 구분하기 어렵고, 의사결정 시간이 극도로 압축되어 과잉 대응이나 선제공격의 구조적 압박이 커지게 된다. 특히, 조기경보 및 지휘통제 체계가 취약한 북한은 신호의 불확실성 속에서 '기다림'보다 '선제 대응'을 생존을 위한 합리적 선택으로 수용할 위험이 높다. 이는 특히 재래식 및 핵 전력 차원에서 한미동맹에 비해 열세에 있는 북한 입장에서 더욱 그러하다(Larsen 2023; Kim 2025).

따라서, 북한이 추진중인 AI의 군사적 활용 위험은 '북한이 곧 AI 강국이 되어 기존과는 수준이 다른 차원의 안보위협을 제기한다'는 식의 과장된 전망이 아닌, 제한된 AI 역량이 북한의 기존 비대칭 군사전략, 사이버 작전, 핵 운용환경과 결합할 때 발생하는 복합위험으로 이해해야 한다.

III. AI 시대 한반도 군사환경 변화와 남북관계

1. AI의 군사적 성격: 정확성, 속도, 상황인식

북한이 획득할 수 있는 제한적 수준의 군사AI가 아닌 현재 진행되고 있는 기술혁신의 기회가 충분히 활용된다면 어떤 변화를 가져오는 것일까? 한국은 인구 대비 AI 특허에서 세계 최고 수준을 보이고, AI 연구, 산업 및 컴퓨팅 인프라에서 북한보다 현저히 유리한 위치에 있다(Sajadieh 2026).

최고 수준의 AI가 군사적으로 발휘하는 효과는 세 가지로 요약할 수 있다. 첫째는 정확성이다. AI는 영상, 신호, 위성, 드론, 통신자료를 분석해 표적 식별과 타격 정밀도를 높인다. 둘째는 속도다. AI는 탐지, 분류, 판단, 방책 생성, 타격결심까지 걸리는 시간을 단축한다. 셋째는 상황인식이다. AI는 서로 다른 출처의 데이터를 융합하여 지휘관이 전장 상황을 더 빠르고 넓게 파악하도록 돕는다(김양규 2026).

이 세 요소는 한국의 3축체계와 결합될 때 특히 큰 효과를 낼 수 있다. 킬 체인(Kill Chain)은 북한의 미사일 발사 징후를 탐지하고 선제적으로 타격하는 개념이며, 한국형 미사일방어(KAMD)는 북한 미사일을 요격하는 방어체계다. 대량응정보복(KMPR)은 북한이 핵을 사용할 경우 지휘부와 핵심 군사시설을 보복 타격하는 대량응정보복 체계이며, AI는 이 세 축 모두를 강화할 수 있다. 그러나 AI가 가장 큰 유혹을 제공하는 영역은 킬체인과 재래식 대군사타격(conventional counterforce)이다(Bowers and Hiim 2021; Kim 2025b)인데 북한의 이동식 발사대, 지휘통제 노드, 미사일 기지, 핵 관련 시설을 더 빨리 탐지하고 타격할 수 있다고 믿게 만들기 때문이다.

문제는 이러한 능력이 억제를 강화하는 동시에 위기 불안정성을 키울 수 있다는 데 있다. 한국 입장에서는 AI 기반 정밀타격 능력이 북한 핵 위협을 억제하고 피해를 줄이는 합리적 대응으로 보일 수 있다. 그러나 북한 입장에서는 그것이 자신들의 핵 능력을 선제적으로 제거하려는 시도로 보일 수 있다. 특히 북한이 재래식 전쟁에서 패배하면 정권 붕괴로 이어질 수 있다고 인식한다면, 위기 초기에 제한적 핵 사용 또는 선제 핵 사용을 고려할 유인이 생긴다(Kim 2025b).

2. 전략적 비핵무기와 AI 기반 대군사타격의 부상

최근 전략 안정성 논의에서 중요한 개념은 전략적 비핵무기(SNNW)이다. 이는 핵무기는 아니지만 상대의 핵심 전략자산, 지휘통제, 조기경보, 핵전력 생존성에 영향을 주어 핵무기와 유사한 전략적 효과를 창출할 수 있는 비핵무기체계를 의미한다. 극초음속 미사일, 장거리 정밀타격, 사이버 공격, 우주 및 대우주 능력, 전자전, 무인체계, AI 기반 표적식별이 여기에 포함될 수 있다 (윤정현 2026, 1-4; Zala 2024).

AI는 동 무기체계의 효과를 증폭시킨다. 과거 전략적 비핵무기가 물리적 타격 능력에 주로 의존했다면, AI 기반 SNNW는 표적 탐색, 우선순위 설정, 실시간 추적, 위협평가, 타격결심 지원을 통한 전략적 효과를 낸다. 즉, AI는 비핵무기가 핵전력의 생존성을 위협할 수 있다는 인식을 강화한다. 이 점에서 AI 기반 SNNW는 단순한 첨단 재래식 무기가 아닌 핵 역지의 전제조건을 흔드는 기술적 환경이다.

그 이유는 핵 역지의 핵심이 단순히 핵무기의 보유 여부가 아닌, 상대의 선제공격 이후에도 보복할 수 있는 2차 공격능력(second-strike capability)의 신뢰성에 달려 있기 때문이다(Kim 2025a). 냉전기 상호확증파괴(MAD)의 기본 논리는 어느 한쪽도 상대의 핵전력을 완전히 제거할 수 없으며, 선제공격을 하더라도 감당하기 어려운 보복을 피할 수 없다는 상호 취약성 인식(mutual vulnerability)에 기초했다. 다시 말해, 핵전력의 안정성은 탄두 수 자체보다 그 전력이 얼마나 생존 가능하며, 상대가 그 생존성을 얼마나 신뢰하는가에 의해 결정된다(Jervis 1989; Lieber and Press 2017). 그런데 AI 기반 감시정찰, 실시간 표적추적, 정밀타격, 사이버·전자전 능력이 결합될 경우, 상대의 이동식 발사대, 지휘통제 노드, 조기경보 체계, 핵 운용 인프라를 더 빠르고 정확하게 찾아낼 수 있다는 기대가 커진다. 이때 실제로 완전한 무력화가 가능한지의 여부보다 더 중요한 것은, 위기 상황에서 양측이 그렇게 믿을 수 있다는 점이다. 강한 쪽은 피해제한(damage limitation)의 유혹을 느끼고, 약한 쪽은 “기다리면 잃는다”는 압박을 받게 된다. 따라서 AI 기반 SNNW는 핵무기를 직접 대체하지 않더라도, 핵 역지가 작동해 온 핵심 조건인 2차 공격능력의 생존성과 그에 대한 상호 신뢰를 약화시킬 수 있다.

한국의 경우 SNNW는 북한 핵 위협에 대응하는 매력적 옵션이다. 한국은 독자 핵무장을 추진하지 않는 조건에서 북한의 핵·미사일 위협을 억제해야 한다. 따라서 한국은 감시정찰, 정밀타격, 미사일방어, 사이버·전자전, 지휘통제 자동화에 투자할 수밖에 없다. 그러나 기존 연구들은

SNNW의 실현 가능성을 낮게 보고 있다. 보워(Ian Bowers)와 힘(Stålthane Hiim)은 한국의 3축체계를 비핵국가가 핵보유 적국을 상대로 첨단 재래식 능력에 의존해 억제력을 시도하는 독특한 사례로 분석한다. 한국은 미사일 방어, 정밀 탄도·순항미사일, C4ISR 능력을 결합하여 북한 핵 위협을 억제하고, 유사시 피해를 제한하려 한다. 이 전략은 미국의 확장억제 신뢰성에 대한 불안, 북한의 핵 능력 고도화, 한국의 전략적 자율성 추구하고 맞물려 있다 (Bowers and Hiim 2021, 7-13).

보워와 힘은 전략이 세 가지 딜레마를 낳는다고 지적한다. 첫째, 작전적으로 매우 어렵다. 한국은 짧은 시간 안에 북한 전역의 이동식 발사대, 미사일 기지, 지휘통제 노드, 핵 관련 시설을 탐지·추적·타격해야 한다. 이는 기술적으로도 어렵고, 위기 상황에서 정보 오류와 오판 가능성이 크다. 둘째, 북한은 한국의 대군사타격 능력을 수동적으로 받아들이지 않을 것이다. 북한은 이동식 발사대, 지하시설, 기만표적, 분산배치, 전술 핵, 위임형 지휘통제, 조기 사용 교리 등으로 대응할 가능성이 크다. 셋째, 이 전략은 위기 안정성을 약화시킬 수 있다. 북한은 한국의 선제타격 가능성을 우려해 먼저 핵을 사용해야 한다는 압박을 느낄 수 있다 (Bowers and Hiim 2021, 21-34).

설사 AI 기술 혁명으로 SNNW의 실현가능성을 높일 수 있다고 하더라도, 보워와 힘이 지적한 딜레마 문제는 해결되지 않는다. AI가 탐지, 추적 및 타격 시간을 줄일수록 한국은 더 신속한 대응능력을 갖게 된다. 그러나 북한 입장에서는 자신의 의사결정 시간이 더 줄어든다. 이 경우 북한은 SNNW를 단순한 '비핵' 무기로 보지 않을 수 있다. 북한의 핵전력과 지휘통제 체계가 한국의 AI 기반 SNNW에 의해 마비될 수 있다고 판단하면, 북한은 이를 사실상 핵 대군사타격에 준하는 위협으로 해석할 가능성이 크다. 즉, 한국의 탐지능력이 높아질수록 북한의 은폐능력은 낮아지고, 북한은 핵전력이 제거되기 전에 먼저 사용해야 한다는 압박을 받을 수 있다. 이 점에서 AI는 억제력 강화와 위기 불안정성 증대를 동시에 낳는다.

이러한 점에서 AI 기반 SNNW의 부상은 한반도에서 매우 특수한 위험을 낳는다. 한국은 북한의 핵 능력 증강에 대응하기 위해 AI 기반 비핵전략무기를 발전시키려 할 것이고, 이는 한국 입장에서는 독자 핵무장 없이 억제력을 강화하는 합리적 선택으로 보일 수 있다. 그러나 북한 입장에서는 한국의 이러한 능력 발전이 자신의 핵 억지력을 무력화하고 정권 생존을 위협하는 공격적 군사혁신으로 인식될 가능성이 크다. 결국 AI 기반 SNNW는 한국에게는 “핵을 갖지 않고 북한 핵을 억제하는 수단”이지만, 북한에게는 “핵을 쓰기 전에 제거당할 수 있다는 불안”을 증폭시키는 요인이 될 수 있다. 이 인식의 비대칭성이 바로 AI 시대 한반도 군사환경을 불안정하게 만드는 핵심 메커니즘이다.

3. 북한의 이중열세와 선제사용 유인

한국의 급격한 군사 AI 개발 시도에 대해 북한이 선제적 핵 사용으로 대응할 수 있다는 지적이 다소 과도한 위협인식이라고 지적할 수는 있다. 그러나 이에 대해 라슨(Even Hellan Larsen)은 핵전력과 재래식 전력 모두에서 비대칭성이 존재하는 상황에서 의도적 핵 선제사용이 합리적 선택이 될 수 있음을 게임이론적으로 설명한다. 강한 행위자는 핵 비대칭과 대군사타격 기술 발전을 활용하여 피해제한을 추구할 수 있고, 약한 행위자는 재래식 패배가 정권 붕괴로 이어질 경우 핵무기를 조기에 사용하여 상대의 의지를 꺾거나 전쟁 양상을 바꾸려 할 수 있다 (Larsen 2023, 849-856).

북한은 이 논리가 적용되기 쉬운 사례다. 북한은 한미동맹에 비해 재래식 전력에서 열세이고, 핵전력에서도 미국의 압도적 핵우위와 한국의 정밀타격 능력에 노출되어 있다. 2022년 핵무력정책법을 통해 공개된 북한의 공세적 핵교리는 바로 이러한 열세를 보완하기 위한 전략으로 이해할 수 있다. 즉, 북한은 재래식 전쟁에서 밀릴 경우 조기 핵 사용 가능성을 열어 두어 한국과 미국의 군사행동을 억제하려 하는 것이다.

문제는 AI 시대에 북한의 위협 인식이 더 악화될 수 있다는 점이다. 한국의 AI 기반 ISR과 SNNW가 강화될수록 북한은 자신의 핵전력 생존성이 줄어든다고 인식할 수 있다. 한국은 이를 방어적 억제수단으로 보더라도, 북한은 이를 정권 종말 공격의 전조로 해석할 수 있다. 기존 연구가 지적하듯, 한반도에서 한국의 재래식 대군사타격은 핵 문턱 아래에서 작동하는 것처럼 보이지만, 북한에게는 사실상 핵 대군사타격과 기능적으로 유사하게 인식될 수 있다 (Kim 2025b, 20-22).

4. 남북관계의 구조 변화: 협력보다 위협관리

따라서 AI 시대 남북 군사관계는 단순한 군사력 증강 경쟁이 아닌 인식과 시간의 경쟁으로 바뀔 것으로 전망된다. AI는 더 많은 정보를 제공하지만, 반드시 더 나은 판단을 보장하지 않는다. 오히려 위기 상황에서는 AI가 제공하는 빠른 정보, 자동화된 표적식별, 알고리즘 기반 위협평가가 인간 의사결정자에게 과도한 확신을 줄 수 있다. 잘못된 데이터, 적의 기만, 사이버 침투, 스푸핑, 오분류, 신뢰도 보정 실패, 데이터 오염과 자동화 편향이 결합될 경우 잘못된 결론이 매우 빠르게 도출될 수 있다(김양규 2026a; 2026b).

이런 조건에서 남북 간 협력 가능성은 제한된다. 협력은 최소한의 신뢰, 상호 투명성, 제도적 지속성, 기술 사용의 목적에 대한 합의가 필요하다. 그러나 현재 남북관계는 적대적 두 국가론, 북한의 핵무력 법제화, 한국의 3축체계 발전, 한미 확장억제 강화, 북러 군사협력 심화, 군사통신선 단절이라는 조건 속에 있다. AI는 이 불신을 완화하기보다 더 복잡하게 할 가능성이 크다. 따라서, AI 시대 남북 군사관계는 전략적 안정성이 매우 취약해지는 국면에 접어들 것으로 보인다.

IV. 북한의 군사 AI 기술획득 및 활용 경로: 대중(對中)·대러(對露) 협력과 남북협력 비교

1. 북한의 선택지는 무엇인가?

북한의 군사 AI 미래 옵션은 (1) 대중(對中)·대러(對露) 기술협력, (2) 제재회피형 저비용 자체 AI 활용, (3) 남북협력 경로 셋으로 구분된다. 이들은 상호 배타적인 시나리오가 아닌 동시에 진행될 수 있는 기술획득 및 활용 경로다. 다만, 직접적인 군사 AI 협력이라는 의미에서 남북협력 경로의 실현 가능성은 현시점에서 극히 낮다.

군사 분야에서 AI 활용에 관한 남북협력이 어려운 이유는 한반도 군사환경의 구조 변화에서 찾을 수 있다. NPT 조약국인 한국은 현재 핵을 개발하지 않은 상태에서 북한의 점증하는 핵 위협을 AI를 통해 정밀성 및 대응 속도 차원에서 크게 강화된 전략적 비핵무기와 재래식 대군사타격 능력을 통해 대응하고자 한다. 상기 언급한 바와 같이 북한은 AI를 경량화된 형태로 무인기에 탑재하여 사용하거나, 부분적으로는 사이버공격, 감시정찰, 핵 운용 보조, 지휘통제 정보화에 활용하려 할 것으로 보인다. 양측 모두 AI를 협력의 기술이 아닌 상대의 군사적 이점을 상쇄하기 위한 경쟁의 기술로 인식하고 있다. 한국 입장에서는 한미동맹 강화를 보조할 만한 대안이 군사적 AI 활용 말고는 현실적으로 없는 상황이며, 북한은 AI의 군사적 활용을 점점 체제생존을 위한 조건으로 인식하고 있다. 이런 조건에서 군사영역의 남북 AI 협력은 정책적으로도, 전략적으로도, 제도적으로도 성립하기 어렵다.

2. 시나리오 A: 대중(對中)·대러(對露) 협력을 통한 비대칭 AI 군사화

남북간 군비경쟁이 뚜렷하게 진행중인 상황에서 AI 분야에서 북한에게 가장 현실적인 외부 협력 경로는 중국과 러시아다. 중국은 AI 연구, 감시통제기술, 드론, 데이터 처리, 사이버 역량, 민군 융합 생태계에서 세계적 수준의 역량을 보유하고 있다. 한 가지 중요한 변수는 중국이 북한의 전통적인 기술협력 파트너지만, 대북제재와 한반도 불안정 회피라는 이해관계 때문에 첨단 군사 AI의 직접 이전에는 러시아보다 신중할 가능성이 크다는 점이다. 따라서 북중협력과 북러협력은 동일한 기술이전 경로로 간주해서는 안 된다.

러시아는 우크라이나 전쟁을 통해 드론, 전자전, 정찰·타격 결합, 전쟁 시뮬레이션, 미사일 기술, 위성·우주발사체 기술의 실전 경험을 선도적으로 축적하고 있다. 북한 역시 우크라이나 파병을 통해 러시아와 군사협력을 강화해 왔고, 미사일, 정찰위성, 재래식 군수생산, 무인기, 전자전 분야에서 직간접적 이익을 얻었다. AI 분야에서도 연구인력 교류와 이중용도 기술 획득의 기회가 확대된 것으로 보이지만, 구체적인 군사 AI 기술이 어느 수준까지 이전되었는지는 공개자료만으로 확인하기 어렵다.

최근 보고서들은 북러 군사협력 이후 북한의 군수공장이 풀 가동 상태에 들어갔고, 러시아의 식량, 유류 및 원자재 지원과 실전 테스트 경험이 북한의 재래식 무기와 군수 생산능력 향상에 기여할 수 있다고 본다(조장원 2025, 2-4). 2026년의 전망도 북한이 2025년 핵 무력 실전운용 고착화와 재래식 전력 현대화를 동시에 추진했고, 우크라이나 파병 이후 러시아와의 협력을 통해 재래식 전력 현대화와 군수협력 네트워크를 강화하고 있다고 평가한다(이호령 2026, 12-16).

북한이 대중(對中)·대러(對露) 협력을 통해 얻으려는 것은 현재 미중이 경쟁하는 수준의 초거대 AI 모델이나 최첨단 반도체일 필요는 없다. 데이터, 전기 등 인프라 측면에서 매우 뚜렷한 한계를 보이는 북한에게 더 필요한 것은 군사적 실용성이 큰 중간 수준의 기술이다. 예컨대 드론 영상처리, 전자전 회피, 위성영상 판독, 표적 추적, 통신교란 대응, 전술 데이터링크, 사이버 자동화, 정찰위성 운용, 미사일 유도 및 제어 알고리즘, 전쟁 시뮬레이션 등이다. 지난 9차 당대회에서 인공지능무인공격종합체를 강조한 것도 이런 맥락에서 이해할 수 있다.

중국 및 러시아와 협력을 강화하는 동 경로의 위험은 북한의 AI 군사화가 핵전력 고도화와 높은 수준으로 결합할 수 있다는 점이다. 북한은 이미 고체연료 ICBM, 전술 핵, 장거리 순항미사일, 정찰위성, 무인수중 공격정, 무인기 등을 발전시키고 있다. 김정은 스스로도 2021년 제8차

당대회에서 제시한 13개 핵·미사일 체계 중 고체 ICBM, 전술 핵, 지상공격 순항미사일, 정찰위성이 중요한 군사적 기여를 했다고 평가한 바 있다(Van Diepen 2026). AI는 이와 같은 기존 무기 체계들과 같은 수준에서 고려할 수 있는 독립적·병렬적 혁신보다 이들을 더 잘 운용하고 연결하는 보조기술이자 전력증강자 역할을 하게 될 것으로 보인다. 이 경우 남과 북은 군사 AI 분야에서 나선형(spiral model)의 군비경쟁을 벌이는 것이 불가피하다.

3. 시나리오 B: 제재회피형 저비용 자체 AI 활용

두 번째 경로는 제재회피형 저비용 자체 AI 활용이다. 이 경로는 대외 의존도를 우려할 필요 없이 현시점부터 바로 적용할 수 있기 때문에 세 시나리오 중 가장 현실성이 높다. 또한, 배타적으로 존재하는 별도 경로보다는 다른 시나리오와 동시에 진행될 것으로 보인다. 북한은 제재로 인해 고급 반도체, 클라우드 인프라, 대형 데이터센터에 접근하는 것이 현실적으로 불가능하다. 그러나 오픈소스 모델, 경량화 알고리즘, 해외 위장 IT 노동자, 사이버 범죄 네트워크, 제3국 서버, 암호자산 생태계를 활용하면 큰 비용을 들이지 않고도 현재 레거시 자산들만 활용하는 것에 비해 상당히 큰 효과를 낼 수 있다.

앞서 논의한 내용을 토대로 볼 때 본 경로에서 AI는 세 가지 방식으로 활용될 수 있다(김민정 2025). 첫째, 사이버 작전의 자동화다. 피싱, 악성코드 개발, 취약점 탐색, 가짜 신원 생성, 언어 번역, 목표별 맞춤형 사회공학 공격에 AI가 활용될 수 있다. 둘째, 정보전과 심리전이다. 딥페이크 영상, 합성 음성, 자동화된 선전 콘텐츠, 허위정보 확산, 여론 교란이 가능하다. 셋째, 내부통제와 감시다. 얼굴인식, 음성인식, 이동경로 추적, 감시카메라 분석은 체제 통제와 반체제 행위 탐지에 활용될 수 있다.

이 경로는 남북관계에도 직접적 영향을 미친다. 북한의 AI 기반 사이버 및 정보전은 한국 사회 내부의 불신을 증폭시키고, 선거, 안보, 재난 및 경제 이슈를 둘러싼 허위정보 확산을 촉진할 수 있다. 또한 금융기관, 방산업체, 연구기관, 대학, 탈북민 네트워크, 언론, 정부기관을 대상으로 한 사회공학 공격의 정교화를 낳을 수 있다. 이러한 가짜 정보가 다시 한국이 추진중인 AI 기반 정보 분석 및 의사결정 지원 루프에 활용될 수 있기에 그 피해는 가중될 수 있다. 북한 AI 위협은 군사적 전장에 국한되지 않고 남북관계 전체의 사회적 기반을 흔들 수 있는 것이다.

4. 시나리오 C: 남북협력 경로의 비현실성

북한이 자체 모델 또는 중국, 러시아와 협력할 뿐 아니라 한국과도 협력하는 것은 이론적으로는 가능하다. 특히 AI는 비군사적인 영역, 즉 재난 대응, 산림 및 수해 모니터링, 감염병 관리, 농업 생산성 향상, DMZ 생태관리, 기상정보 공유 등에 활용될 수 있다. 기초 AI 역량, 언어와 지형, 기후의 유사성 측면에서도 다른 경로 보다 한국과 직접 협력하는 것이 북한 입장에서 훨씬 효과적이라고 볼 수 있다. 그러나 군사영역에서 남북 AI 분야 협력 경로는 매우 비현실적이다.

첫째, 기술의 이중용도(dual-use technology) 문제 때문이다. 무인기, 감시정찰, 영상처리, 통신망, 사이버 방어, 지휘통제, 자동화된 의사결정지원에 AI를 접목하는 기술은 민간과 군사의 경계가 매우 흐리다. 남북이 군사 AI와 관련된 협력을 추진할 경우, 그것은 즉각적으로 군사적 전용이 가능한 기술 이전 문제가 된다. 북한군의 작전능력이나 핵·미사일 프로그램에 직접 기여하는 기술지원은 유엔안보리 대북제재의 무기·기술지원 금지 및 포괄적 통제조항에 저촉될 가능성이 매우 높다. 비군사적 과학기술 협력도 통보·승인 및 전용 방지 절차를 전제로 검토되어야 한다.

둘째, 안보전략 차원에서 모순되기 때문이다. 비핵국가인 한국은 AI를 활용해 북한의 핵·미사일 위협을 탐지하고 억지하려 한다. 특히 AI 기반 감시정찰(ISR), 디지털 표적화, 지휘통제(C4I), 미사일방어, 정밀타격, 전략적 비핵무기 활용 등을 통해 북한의 핵 능력 증강에서 비롯되는 위협을 상쇄하려 한다. 이에 대한 반작용으로 북한은 당연히 AI를 활용해 한국의 감시정찰망을 회피하고, 사이버 공격과 무인기, 전술 핵 운용환경을 강화하려 할 것이다. 양측이 AI를 상대의 군사적 우위를 상쇄하는 수단으로 사용하는 조건에서 북한의 AI 능력 증강에 이바지하는 형태의 협력은 안보전략 차원에서 모순되기에 구조적으로 제한된다.

셋째, 위협 인식의 비대칭성 때문이다. 한국은 북한이 제기하는 핵 위협에 대한 방어적 대응 차원에서 AI 기반 전략적 비핵무기 활용을 추진하더라도 북한은 이를 선제타격을 통한 북핵전력 제거, 즉 1차 공격능력(first strike capability)으로 인지하여 자신의 핵전력 생존성이 크게 저하되는 공격적 능력으로 인식할 가능성이 크다. AI는 바로 이 남북간 군비경쟁의 중심축으로 자리하고 있기 때문에 남북 간 AI 군사협력은 기술협력의 문제가 아닌 억지균형 구조를 본질적으로 변화시킬 수 있는 문제다.

넷째, 정치적 조건의 부재 때문이다. 북한은 남북관계를 적대적 두 국가 관계로 규정했고, 비핵화 개념 자체를 부정하며 핵무력을 헌법적·국가전략적 권리로 빠르게 제도화하고 있다. 이번 9차 당대회 보고서에 ‘조국통일’ 항목이 전면 삭제되었을 뿐 아니라, 담론 차원에서 ‘우리민족끼리’ 대신

‘대적투쟁’이 전면에 부상하였다. 조국평화통일위원회(조평통), 조국통일민주주의전선(조국전선), 민족화해협의회(민화협) 등이 폐지되고, 대남 업무를 외무성과 국방성이 맡고 있다(북한연구실 2026). 북한이 이렇게 명시적으로 적대적 정책으로 대남 문제를 재정의하는 상황에서 군사 AI 협력은 남북 간 신뢰구축 조치보다 북한의 군사 현대화를 돕는 유화적이고 위험한 시도로 해석될 가능성이 크다.

따라서, AI 시대 남북관계에서 군사영역의 협력이 이뤄질 가능성은 극히 희박하고 거의 불가능하다고 보는 것이 현실적이다. 북한은 시나리오 A와 B가 결합된 방식, 즉 대중(對中)·대리(對露) 협력과 제재회피형 저비용 자체 AI 활용을 통해 비대칭 군사력을 강화하려 할 것이고, 한국은 AI 기반 전략적 비핵무기를 통해 북한의 핵 위협에 대응할 것이다. 양측의 AI 군사화는 협력의 기반을 만들기보다 상대의 위협 인식을 강화하고, 높은 확률로 한반도 전략적 안정성을 취약하게 할 수 있다. 이 때문에 우리가 고민해야 하는 과제의 핵심은 “협력 방안”이 아닌 “협력이 어려운 조건에서 어떻게 전략적 안정성을 설계할 것인가”가 되어야 한다. 전략적 안정이 이뤄지지 않은 상황에서는 경제, 사회, 문화 등 다른 영역에서도 남북관계 개선의 돌파구가 마련되기 어렵기 때문이다.

V. 군사 부문에서의 남북 협력 가능성과 한계: 전략적 안정성을 위한 관리된 군비경쟁

1. 협력 불가능성 이후: 무제한적 군비경쟁 vs. 관리된 군사경쟁

군사 AI영역에서 남북간 협력이 비현실적이라면 남는 선택지는 두 가지라고 볼 수 있다. 하나는 무제한적 군비 경쟁이고, 다른 하나는 전략적 안정성을 목표로 설계된 관리된 군사경쟁이다. 본 보고서는 한국이 후자를 지향해야 한다고 주장한다(Kim 2025b).

“관리된 군사경쟁”은 군비통제(arms control)를 통한 남북 쌍방의 군비 감축이나 낮은 수준에서의 남북 간 군사 AI 협력을 의미하지 않는다. 앞 절에서 살펴본 바와 같이, 현재 조건에서 그런 협력이 이뤄질 가능성은 매우 낮다. 관리된 군사 경쟁이란 한국이 현재 북한이 제기하는 위협에 대응하는 차원에서 필요한 군사력을 발전시키되, 그 방향과 운용 원칙을 한반도 전구에서 전략적 안정성을 강화시키는 방향으로 설계하는 것을 의미한다. 핵심은 북한의 핵 사용 및 한국의 정밀 선제타격 유인을 동시에 낮추는 것이다. 이를 위해서는 첫째, 북한이 핵을 사용하더라도 정권 종말을 피할 수 없다는 한국의 대북 2차 공격능력의 신뢰성, 즉 보복 확실성을 높이는 것과, 둘째,

북한이 “지금 쓰지 않으면 잃는다(use-it-or-lose-it)”는 공포에 휩싸이지 않게 지나치게 코너로 몰아넣지 않는 두 방향의 노력을 균형적으로 추진하는 것이다.

2. 무제한 AI 기반 대군사타격 경로의 위험성

물론 기술적으로 우위에 있는 한국 입장에서 무제한 AI 기반 대군사타격(counterforce) 경로는 매력적인 옵션이다. 한국은 북한보다 우월한 경제력, 기술력, 방위산업, 컴퓨팅 인프라, 데이터 역량, 동맹 기반 ISR, 우주 및 사이버 협력망을 보유하고 있다. AI는 현재의 이 기술적, 경제적, 군사적 우위를 모든 측면에서 훨씬 더 크게 만들 수 있다. 특히 군사적으로 한국은 AI로 강화된 정찰위성, 중·고도 무인기, 전자전, 사이버, 고 위력 미사일, 장거리 공대지미사일, 해상 및 수중 플랫폼, F-35 스텔스 전투기, 향후에는 원자력추진 잠수함 등을 모두 결합하여 북한 핵전력을 선제적으로 무력화할 수 있는 능력을 추구할 수 있는 잠재력이 충분하다.

그러나 압도적 우위로 이어질 수 있는 이 경로는 전략적 안정성 차원에서는 도리어 더 위험하다. 첫째, 북한의 선제적 핵 사용 유인을 높인다. 북한은 한국의 압도적 AI 기반 군사 능력이 완성될수록 자신의 핵전력 생존성이 크게 훼손된다고 느낄 수 있다. 둘째, 위기 시 의사결정 시간이 압축된다. 이번 이란 전쟁에서는 AI 기반 결심지원도구가 과거 몇 시간 또는 며칠이 걸리던 정보통합, 분석 및 방책생성 작업을 몇 분 단위로 압축한 것으로 평가된다. 별도의 실험연구에서도 인간 장교가 1~3시간에 수행한 작전분석을 일부 최신 AI 모델이 평균 56초에 완료했다.

다만 이러한 분석 및 방책생성 속도와 실제 표적식별 후 타격까지 걸린 시간은 구분할 필요가 있다(김양규 2026b). 한반도와 같이 중심이 짧은 지형에서는 이러한 시간 압축이 훨씬 더 무서운 불안정 심화 요인이 된다. 양측 모두 매우 짧은 시간 안에 모호한 신호를 해석하고 결심해야 하기 때문이다. 셋째, 오판 가능성을 키운다. 통상적 훈련, 감시정찰 강화, 사이버 이상 징후, 미사일 이동, 전자전 신호가 선제공격 준비로 해석될 수 있다. 이것이 앞서 언급한 결심 시간 단축과 상호작용하면 훨씬 더 심각한 문제가 된다. 넷째, 북한이 지금보다도 더 위험한 핵 전략, 핵 운용방식을 채택하도록 유도한다. 핵 발사권한 사전위임, 이동식 발사대 분산, 조기 핵 사용 교리, 핵·재래식 지휘체계 얹힘 등이 동시에 강화될 수 있다.

앞서 언급한 바와 같이 기존 연구들은 이미 한국의 재래식 대군사타격 전략이 북한의 선제공격 유인을 높이고, 북한이 위임형 C2를 채택하도록 압박해 우발적 또는 무단 핵 사용 위험을 키울 수 있다고 지적하고 있다(Bowers and Hiim 2021, 33-34). 이 딜레마는 AI 시대에 더 심화된다. AI는 “더 빠른 결심”을 가능하게 하지만, 핵위기에서 더 빠른 결심이 항상 더 안전한 결심은 아니다(김양규 2026b).

3. 대안: 대가치타격과 보복 확실성의 우선성

한반도 전략안정의 출발점은 북한이 핵을 사용하는 순간 정권 생존을 기대할 수 없다는 확신을 심어주는 것이다. 이는 대군사타격보다 대가치타격(countervalue), 즉, 2차 공격능력과¹ 보복 확실성의 문제다. 본고에서 대가치타격은 『2022 국방백서』가 KMPR의 대상으로 제시한 북한의 전쟁지도부와 핵심시설에 대한 응징적 억제를 의미한다. 이는 북한 체제에서 최고지도자와 전쟁지도부가 갖는 핵심적 가치를 고려한 기능적 의미의 대가치타격(countervalue)이며, 민간인과 민간시설을 직접 표적으로 삼는 고전적인 대가치타격 개념과는 구분된다(국방부 2023, 60-61).

한국은 핵보유국이 아니므로 최종적인 핵 보복은 미국의 확장억제에 의존한다. 핵협의그룹(NCG), 재래식-핵능력 통합(CNI) 등 한미 공동 핵 보복 기획과 실행체계 구축이 1차적으로 중요하다. 동시에 한국의 독자적인 능력, 즉, 재래식 고 위력 보복능력, 지휘통제 회복탄력성, 생존 가능한 보복전력, 위기 시 정부 연속성, 피해복구 능력을 통해 보복 확실성을 높일 수 있다.

기존 연구는 한반도에서 안정적 균형을 위해 한국이 보복 확실성과 지휘통제(C2) 회복탄력성을 우선하고, 거부능력 가운데 북한 핵전력의 완전한 선제 무력화를 목표로 하는 공세적 피해제한능력은 제한적으로 운용해야 한다고 제안한다. 다만 미사일방어, 수동 방호, 전력 분산과 C2 복원력 등 방어적 거부능력은 억제와 피해제한을 위해 지속적으로 강화할 필요가 있다. 핵심은 한국이 독자적으로 북한을 완전히 무력화할 수 있다는 기술적 과시가 아닌, 북한이 어떤 방식으로 공격하더라도 한국의 독자적 보복을 피할 수 없다는 인식이다(Kimb 2025, 22-25).

¹ 이하 한국의 2차 공격능력은 독자적인 핵 보복전력이 아니라, 북한의 선제공격 이후에도 작동하는 미국의 확장억제와 한국의 생존 가능한 재래식 보복능력을 포괄하는 의미로 사용한다.

동 보고서도 한국이 AI를 북한의 모든 핵 자산을 선제적으로 찾아 파괴하는 데 우선 투입되기보다, 다음 기능에 우선 활용되어야 한다고 제언한다. 첫째, 조기경보의 신뢰성 향상이다. AI는 다중 감시정찰 데이터를 통합해 오 경보 가능성을 줄이는 데 기여할 수 있다. 그러나, AI 통합 자체가 조기경보의 신뢰성을 보장하는 것은 아니므로, 적대적 기만시험, 신뢰도 보정, 독립 정보 출처 간 교차검증 및 성능저하 상황에서의 수동 운용절차를 병행해야 한다. 둘째, C2 회복탄력성 강화다. 분산된 지휘소, 예비통신망, 이동식 지휘체계, 자동화된 피해평가와 복구, 보복전력 재구성 지원에 AI를 집중적으로 활용해야 한다. 셋째, 보복전력 생존성 향상이다. AI는 핵심전력 분산배치, 기만, 은폐, 이동경로 최적화, 적 감시망 회피에 활용될 수 있다. 넷째, 위기관리 지원이다. AI는 상대의 행동을 단일한 공격 징후로 단정하기보다 여러 가능성을 확률적으로 제시하고, 인간 정책결정자가 오판을 피할 수 있도록 의사결정체계를 보조해야 한다.

4. 제한된 거부능력과 절제된 대군사타격

물론 대군사타격을 완전히 포기하는 것은 반드시 바람직한 방향은 아니다. 한미동맹의 모든 억지(deterrence) 노력과 조치에도 불구하고 북한이 실제 핵 사용을 준비하거나, 발사가 임박했거나, 핵탄두와 투발수단이 결합되어 명백한 공격 징후가 포착되는 상황에서는 한국이 최후의 수단으로 이를 방해하거나 파괴할 수 있는 능력을 보유하고 있어야 한다. 강조하고자 하는 것은 이 능력이 한국 군사전략의 중심축이 되거나 전면에 나서는 목표가 되어서는 안 된다는 점이다.

따라서 바람직한 방향은 제한적인 수준의 거부에 의한 억지능력, 곧 ‘절제된 형태의 대군사타격(bounded counterforce)’이다. 이 원칙은 다음 조건을 포함해야 한다. 첫째, 높은 승인 문턱이다. 핵 관련 대군사타격은 자동화된 군사절차가 아니라 반드시 최고정치지도자의 명시적 승인 아래서만 가능하도록 결심체계를 설계해야 한다. 둘째, 의미 있는 인간 통제(meaningful human control)다. AI는 표적을 추천할 수 있지만, 핵 관련 표적에 대한 타격결심은 절대적으로 인간이 내려야 한다. 인간이 단순히 승인 버튼을 누르는 형식적 인간통제로는 충분하지 않다. 인간 감독자에게는 AI의 추천을 거부할 실질적 권한, 충분한 숙고시간, 정보 출처와 불확실성을 확인할 수 있는 능력, 작동을 중단하거나 수동체계로 전환할 수 있는 권한이 보장되어야 한다.

셋째, 다중 정보확인이다. 단일 센서, 단일 AI 모델, 단일 정보 출처에 근거한 타격은 금지되어야 한다. 위성, 신호정보, 인적정보, 무인기, 지상감시, 동맹 정보가 상호 검증되어야 한다. 넷째, 시간 및 공간적 완충장치다. 모든 이상 징후에 대해 즉각 타격으로 반응해서는 안 되며, 위협의 단계와 지리적 위치, 발사 임박성, 지휘통제 상태를 구분한 대응이 필요하다. 다섯째, 표적 범위 제한이다. 북한 핵전력 전체, 지휘부 전체, 군사력 전체를 동시에 마비시키는 광범위한 공격은 북한에게 정권 종말 공격으로 인식될 수 있다. 제한된 거부는 임박한 핵 사용을 방해하는 목적에 한정되어야 한다. 여섯째, 동맹 조율이다. 한미 간 NCG, 확장억제전략협의체, 연합지휘체제 안에서 AI 기반 표적식별과 대군사타격 운용 원칙을 사전에 조율하고 상황 발생시 원칙에 따라 대응해야 한다.

5. 한국형 전략적 대체(strategic substitution) 전략

한국은 독자 핵무장이나 무제한 대군사타격 경쟁 대신 한국형 “전략적 대체(Strategic Substitution)”를 모색해야 한다. 이는 핵무기를 보유하지 않더라도, AI 기반 ISR, C2 회복탄력성, 정밀 보복능력, 미사일방어, 사이버 및 전자전, 수중 플랫폼, 우주 기반 감시체제를 결합해 복합전력으로 북한 핵 위협을 관리하는 전략이다(Cunningham 2025). 그러나 이 전략적 대체는 공격 우위에 대한 과신이 아니라 앞서 강조한 보복 확실성과 위기 안정성에 기초해야 한다.

한국형 전략적 대체의 핵심은 세 가지다. 첫째, 3축 체계의 핵심으로서의 대량응징보복(KMPR), 즉 북한의 핵 사용 이후에도 실행 가능하며 북한 전쟁지도부와 핵심 군사시설에 감당할 수 없는 비용을 부과하는 한국의 생존 가능한 독자적 보복능력이다. 둘째, 북한 선제공격 이후에도 작동하는 C2 생존성이다. 셋째, 북한이 한국의 통상적 군사대비를 선제공격 준비로 오인하지 않도록 하는 절제된 운용 원칙이다.

〈그림 1〉 한반도 평화를 위한 AI의 군사적 활용 방안



이는 상기 [그림 1]에서 제시한 두 갈래 경로 중 두 번째 경로와 일치한다. 무제한적 AI 기반 대군사타격 강화는 북한으로 하여금 ‘사용하거나 잃어버리거나’의 불안을 갖도록 만들지만, 보복적 성격과 표적 범위 및 운용 원칙이 명확하게 제한된 AI 기반 KMPR과 C2 생존성 강화, 절제된 대군사타격은 북한의 선제사용 유인을 낮추는 데 기여할 수 있고, 관리된 공포의 균형을 구축할 수 있다. 이것이 남북 양측이 군사적으로 경쟁을 하더라도 한반도 내 전략적 안정성을 취약하게 하지 않을 수 있는 길이다.

6. AI-핵 위험관리와 장기적 군비통제 의제

앞서 강조한 바와 같이 현재 남북 군사협력은 현실적으로 매우 어렵지만, 장기적으로는 북한을 포함한 이해당사국들 사이에서 AI-핵 위험관리와 군비통제 의제를 준비할 필요가 있다. 이는 단기적 협력 낙관론이 아닌 미래 위기관리의 최소 조건을 미리 만드는 작업이다.

한미동맹 차원에서 첫째, AI 기반 조기경보와 표적식별 체계의 오 경보 위험에 대한 한미 공동연구가 필요하다. 둘째, 한미는 북한이 어떤 신호를 선제공격 징후로 오인할 수 있는지 레드팀 분석을 실시하고 공동의 대응 매뉴얼을 구비해야 한다. 셋째, 장기적으로는 향후 남북 또는 한북미 간 위기관리 대화가 재개될 경우를 대비하여, 무인기 접근, 사이버 사고, 군사훈련 사전통보, 미사일 시험 통보, 우발충돌 방지, 통신선 복원 등 의제를 제시하고 이에 대해 준비할 필요가 있다. 넷째,

보다 포괄적으로는 AI의 군사적 사용과 관련한 국제 규범 논의에서 한국이 “의미 있는 인간 통제,” “핵무기 사용·발사 및 핵 관련 표적에 대한 최종 치명적 타격결정의 AI 단독 수행과 자율화 금지,” “오 정보 검증 절차,” “AI 기반 C2 복원력”을 주요 원칙으로 제시할 수 있을 것이다.

다만, 이러한 군비통제 의제를 현 수준의 남북관계 차원에서 협력 의제로 고려하는 것은 시기상조다. 현재 필요한 것은 남북협력 방안을 준비하는 것이 아니라 장기적으로 이러한 논의가 이어질 수 있는 근본적인 토대, 곧 전략적 안정성을 담보하는 형태의 남북 군사 경쟁 방향을 설계하는 것이다. 군비통제는 당분간 지속될 수밖에 없는 남북 군사경쟁이 장기적으로 전략적 안정 상태에 돌입하고, 양측이 상호 보복 능력의 신뢰성을 공히 인정하며, 상대를 완전히 무력화할 수 없다는 상호 좌절의 인식이 형성될 때 가능하다(Kim 2025). 따라서 한국은 먼저 AI 기반 역지력을 강화하되, 이것이 상호 안정성을 강화하는 형태로 진행될 수 있도록 하는데 역량을 집중해야 한다.

VI. 소결

북한의 AI 역량 자체는 현시점과 예측 가능한 미래에 미국이나 중국 수준의 첨단 군사 AI 능력으로 발전하여 한국에 직접적인 전략적 우위를 확보할 가능성은 매우 낮다. 북한이 가까운 미래에 미국이나 중국 수준의 군사 AI 강국이 될 가능성은 매우 희박하다. 보다 현실적인 위험은 북한이 매우 제한적인 수준의 AI 인프라에서도 활용 가능한 경량화 AI, 사이버 자동화, 무인기, 감시 및 정찰, 음성·영상 합성, 핵 운용환경 보조 기술을 기존 핵 및 비대칭 전력과 결합하는 것이다. 이 경우 북한의 군사 AI는 핵전력의 보조수단이자 사이버 및 정보전의 증폭장치, 무인기·정찰체계의 효율화 수단으로 작동할 가능성이 있다.

반대로 한국은 AI 기반 전략적 비핵무기와 재래식 대군사타격 능력을 통해 북한의 점증하는 핵능력을 효과적으로 상쇄하려 할 것이다. 한국의 입장에서 이는 합리적 대응이고, 실제 한국은 이를 추진할 수 있는 상당한 기술적·산업적 기반을 보유하고 있다. 다만 이러한 기반이 북한 핵전력 전체를 신뢰성 있게 탐지, 추적 및 무력화할 수 있음을 의미하지는 않는다. 독자 핵무장을 선택하지 않는 이상, 비핵국가로서 북한의 핵 위협에 대응하기 위해 한국은 첨단 재래식 능력과 미국 확장억제, AI 기반 ISR, C2 회복탄력성, 정밀보복 능력을 결합할 수밖에 없다. 그러나 일견 매우 타당해 보이는 동 경로가 무제한적 대군사타격 능력 확보 추구로 집중되면 북한의 핵무기 선제사용 유인을 자극하고, 한반도 전략적 안정성의 토대를 심각하게 훼손할 수 있다.

현 수준에서 가장 긴급한 과제는 당분간 남북간 군사 경쟁이 불가피함을 인정하는 것에서부터 출발해야 한다. 북한은 AI를 상대적 열세에 있는 재래식 능력 현대화와 체제 생존의 핵심적인 증폭제로 보고 있고, 한국은 AI를 북한 핵능력에 대응하기 위한 전략적 비핵무기와 대북 억지체계의 핵심 역량으로 삼고자 한다. 양측 모두 AI를 협력의 언어가 아닌 경쟁과 억지의 언어로 이해하고 있다. 더욱이 대북 국제제재, 이중용도 속성, 기술 이전 위협, 군사정보 노출, 악용 가능성을 고려할 때 남북간 군사 AI 협력은 현실적 정책 옵션이 될 수 없다.

그렇다면 AI 시대 남북관계 개선의 기회는 어디에서부터 마련되어야 하는가? 본 보고서는 순진한 협력 낙관론이 아닌 전략적 안정성을 담보하는 군사경쟁을 설계하는 것에서부터 시작해야 한다고 강조한다. 한국은 AI를 활용해 북한의 핵전력을 모두 선제적으로 제거할 수 있다는 공격 우위의 환상에 빠지는 것을 경계해야 한다. 대신 AI를 보복전력 생존성 강화를 통한 보복 확실성, 지휘통제 회복탄력성, 조기경보 신뢰성, 위기관리 등에 우선 활용해야 한다. 마지막 안전장치로서 대군사타격 능력의 속성을 고려할 때 이를 완전히 포기하는 것은 지양해야 하지만, 반드시 임박한 핵 사용에 대응하기 위한 최후의 보루로서 그 역할은 제한되어야 한다.

AI 시대 한반도 평화의 길은 더 빠른 선제타격 능력 확보 경쟁이 아닌 더 신중하게 설계된 억지력 강화 경쟁이다. 이는 어디까지나 1차 공격능력이 아닌 2차 공격 능력, 곧 보복 타격 신뢰성을 기초로 해야만 한다. 전략적 안정성은 남북이 협상 테이블에 앉아 구체적인 어젠다를 두고 협력을 모색할 때에만 가능한 것이 아니다. 1970년대 미소 데탕트가 보여주었듯이, 군사경쟁이 치열하게 진행되는 상황 속에서도 상대가 먼저 핵을 사용할 유인을 낮추고, 어떤 공격도 보복을 피할 수 없다는 인식을 공유하게 하며, 위기 시 오판을 줄이는 방식으로 설계될 때 전략적 안정성은 충분히 달성할 수 있는 목표이다. 관리된 군사경쟁을 진행하는 가운데 과거 미소가 그러했듯 새로운 형태의 협력과 관계 진전의 단초가 마련될 수 있을 것이다.■

■ 김양규_국방대학교 교수.

■ 담당 및 편집: 임재현_EAI 연구원

문의: 02-2277-0746 (ext. 209) jhim@eai.or.kr

인용할 때에는 반드시 출처를 밝혀주시기 바랍니다.
EAI는 어떠한 정파적 이해와도 무관한 독립 연구기관입니다.
EAI가 발행하는 보고서와 저널 및 단행본에 실린 주장과 의견은 EAI와는 무관하며 오로지 저자 개인의 견해를 밝힙니다.

발행일 2026년 9월 21일
979-11-7584-153-6 95340

재단법인 동아시아연구원
03028 서울특별시 종로구 사직로7길 1
Tel. 82 2 2277 1683 Fax 82 2 2277 1684

Email eai@eai.or.kr Website www.eai.or.kr