GLOBAL NK
Zoom & Connect

EAI

# 한국의 소버린 AI 추진과 대북 사이버 위협 취약성 증대 요인 분석

레이프 에릭 이슬리(이화여대)

최다빈(이화여대)

## 서론

한국의 정책결정자들은 인공지능(AI)을 매우 큰 기회로 인식하는 경향이 있으며, AI 경쟁에서 뒤처지는 것을 가장 큰 위험으로 인식한다(Kim, 2025). 사이버 분야의 경험을 인정받아 임명된 기술기업 경영자 출신 한성숙 국무총리는 정부가 공공 부문의 AI 도입을 가속화할 것이라고 밝혔다(Yonhap, 2026). 윤석열 정부는 한국을 미국, 중국과 함께 3대 AI 강국으로 도약시키겠다는 목표를 제시하고 다양한 규제·투자 정책을 추진한 바 있다(MSIT, 2024). 이재명 정부는 국가 경쟁력을 강화하고, 디지털 기술의 활용 분야를 다변화하며 국민이 한국의 여건에 맞는 AI 솔루션을 이용할 수 있도록 국내 AI 모델 개발 등을 추진하면서 "소버린 AI(sovereign AI)"라는 목표를 보다 명시적으로 채택했다(MSIT, 2025). 2026년 6월, 한국 정부는 삼성, SK, 현대 등 기업들이 반도체, AI 데이터센터, 피지컬 AI 분야에서 약속한 다년간의 국내 투자를 바탕으로, 국가 AI 생태계를 구축하고 수도권 이외 지역의 발전을 촉진하기 위한 3대 메가 프로젝트를 발표했다. 이재명

대통령은 7월 미국과 한국의 기술업계 지도자들이 참석한 샌프란시스코 AI 서밋 연설에서 3조 달러가 넘는 이 투자 계획이 한국을 "글로벌 AI 공급망의 대체 불가능한 핵심 축"으로 만들 것이며 "한국 경제와 글로벌 AI 산업을 새로운 차원으로 끌어올리는 데 기여할 것"이라고 주장했다(J. Lee, 2026).

소버린 AI는 넓은 의미에서 외부 행위자에 과도하게 의존하지 않고 AI 정책을 자율적으로 선택하는 데 필요한 국가적 역량으로 정의된다(Varela Sandoval et al., 2026). 이재명 대통령의 AI 담당 수석비서관을 지낸 하정우는 재임 당시, 한국의 소버린 AI 구상이 국내 시장의 기술적 의존을 해소하는 데 그치지 않고 미중 기술 경쟁 심화에 대응해야 하는 제3국에 AI 솔루션을 수출하는 것까지 목표로 한다고 설명했다(H. Lee, 2026). 글로벌 무역, 경제 정책, 국가안보, 문화 콘텐츠 전반에 AI가 빠르게 통합되면서 주권 논의는 데이터와 인프라를 넘어 AI 기술·서비스의 계층 구조인 AI 스택의 다른 계층까지 포괄하는 방향으로 확대되어 왔다(Stanford HAI, 2026). 소버린 AI를 둘러싼 한국 내 논의에는 독자적인 AI 스택 개발을 주장하는 이들과, OpenAI, Google, Meta, Anthropic 등 현재의 업계 선도 기업들에 대한 과도한 의존보다는 이들과의 파트너십을 선호하는 이들이 함께 참여하고 있다 (Chey Institute, 2026). 그러나 소버린 AI 옹호론자들은 한국 정부의 비용-편익 분석에 반드시 반영되어야 할 결정적 변수, 즉 북한의 사이버 위협을 간과하는 경향이 있다. 북한의 사이버 역량 확대, 한국의 취약점에 대한 집중 공략, 공격 작전에 AI를 활용하려는 노력은 한국이 소버린 AI를 추진할 때 직면하는 위험의 성격과 수준을 바꾼다.

## 소버린 AI의 예상 비용과 편익

한국은 전적으로 독자적인 AI 대안에 의존하기보다 기존의 미국 AI 기업들과 협력하고 있다. 이러한 협력 기조는 인프라 가용성, 역량 확보 비용, 보안상의 이점 등을 고려할 때 앞으로도 지속될 가능성이 크다. 미국의 거대 테크 기업들은 사이버보안에 수십억 달러를 투자하고, 세계 최고 수준의 보안 연구 인력을 고용하며, 보안이 철저한 대규모 데이터센터를 운영하고 있다. 전 세계 기반모델(파운데이션 모델)과 최첨단 모델(프론티어 모델) 공급에서 한국의 역할이 제한적이라는 점은 모델 개발에서 정면으로 경쟁하는 데 드는 모든 비용을 감수하기보다 AI 스택의 다른 계층에서 부가가치를 창출하려는 전략적 선택을 시사한다. 프론티어급 모델을 구축하려면 컴퓨팅 자원, 데이터센터 용량, 에너지, 전문 인력 등에 막대한 초기 투자가 필요하다(Kim and Kwon, 2026). 그 대신 한국 기업들은 반도체, AI 도입·운영 인프라, 산업 응용 분야에서의 강점을 활용해 왔다. NVIDIA의 AI 가속기용 고대역폭 메모리(HBM)의 주요 공급업체인 SK하이닉스와, 세계적인 메모리 생산업체이자 테슬라의 AI 칩 위탁생산을 맡은 삼성의 사례는 한국 기업들이 소버린 AI에 필요한 전체 기술 계층을 독자적으로 구축하는 데 우선순위를 두기보다 제조 전문성과 고객 맞춤형 생산을 통해 글로벌 AI 가치사슬에 편입되고 있음을 보여준다(Draudt-Véjares and Lee, 2026).

그러나 이러한 협력에는 의존이라는 대가가 따른다. 현재의 구조에서는 학습 데이터, 모델 아키텍처, 규제 이행 등 핵심적인 사안에 대한 결정 권한이 외국 기업들에게 있으며, 이로 인해 한국의 기업과 기관은 국내 AI 솔루션을 개발·도입하는 과정에서도 종종 소비자의 지위에 머무르게 된다. 소버린 AI에 관심을 가진 국가들은 미국 정부가 "규칙을 바꾸거나, 접근을 제한하거나, 기술 의존을 지렛대로 활용할" 가능성에 대비하고자 한다(Chavez, 2026). 2026년 6월, 트럼프 행정부는 국가안보를 이유로 수출통제를 발동하여 모든 외국인의 Anthropic 최상위 모델

접근을 차단했고, 이로 인해 Anthropic은 하룻밤 사이에 모든 사용자에 대해 Claude Fable 5와 Mythos 5의 서비스를 중단해야 했다(The Economist, 2026b). 이보다 불과 몇 주 전에는 Anthropic과 미 국방부 간의 이용 권한 분쟁으로 해당 기업의 제품이 정치적 쟁점이 된 뒤, 트럼프 대통령이 AI 기업들에 안전성 점검을 위해 정부가 새로운 프론티어 모델에 사전 접근할 수 있도록 자발적으로 협조하라고 지시한 바 있다(Froman, 2026). Fable 5와 Mythos 5에 대한 미국 내외 이용자의 접근은 약 2주 만에 재개되었지만, 이후 Anthropic은 OpenAI, Meta와 함께 신규 모델이 테스트 환경을 벗어나 인터넷상의 실제 시스템을 사이버공격한 사례를 공개했다(CSA, 2026). 한국은 이러한 문제에 대응하고자 AI 기본법을 제정함으로써 규제 측면에서 선도적 역할을 보여주었다. 그러나 한국의 전략과 법적 체계는 여전히 국가안보에 대한 충분한 고려 없이 산업 진흥에 초점이 맞춰져 있다(Bae and Kim, 2025). 더욱이 추진 중인 플랫폼 관련 입법과 데이터 국내 저장 의무화 조치는 한미 간 기업 협력에 걸림돌이 되고, 한국의 AI 산업을 글로벌 클라우드 인프라 및 국경 간 협력으로부터 고립시킬 위험이 있다(Girishankar, 2025).

소버린 AI는 한국 정부 기관과 기업들이 규제, 데이터 거버넌스, 윤리 기준을 자율적으로 정하고 관리하기를 원하는 이들에게 매력적으로 다가온다. 한국의 여론 주도층 또한 외국 모델에 대한 의존 증가가 토큰 및 API 이용료의 해외 지출을 늘려 한국의 서비스 수지 적자를 확대시키고 있다는 우려 속에서, 국내 AI 역량의 필요성을 국가 경제 주권의 관점에서 규정하고 있다(Han, 2026). 이론적으로 국내에서 개발한 모델은 국내 경제에 상당한 편익을 창출함과 동시에 수출 가능성도 열어줄 수 있다. 한국의 관련 주체들은 미국이나 중국 플랫폼에 대한 과도한 의존을 피하고자 하는 개발도상국에 AI 솔루션을 제공하는 것을 목표로 하고 있다(Kharpal, 2025). 그러나 소버린 AI 정책에는 상당한 비용 또한 수반될 수 있다. 한국 정부는 2028년까지 5만 2천 개, 2030년까지 26만 개의 GPU를 확보하겠다고 공약했지만(Jie, 2025), 업계 전문가들은 특히 에너지

제약을 고려할 때 한국이 약속한 데이터센터를 실제로 건설할 수 있을지에 대해 회의적인 시각을 표명해 왔다. 한국전력공사 (KEPCO)의 송전망 구축 사업 가운데 절반 이상이 장기간 소요되는 인허가 절차로 인해 지연되고 있다(Draudt-Véjares and Lee, 2026). 더욱이 정부가 소버린 AI에 막대한 재원을 투입하면 전통적인 국방 분야나 고령화에 대응하기 위한 보건・연금에 배정할 예산이 줄어들 수 있다.

한국 기업들이 소버린 AI 솔루션을 빠르고 저렴하게 구축하도록 압박을 받는다면, 가중치가 공개된 오픈웨이트(open-weight) 모델을 개발의 기반으로 활용할 수 있다. 이는 단순한 추측이 아니다. 과학기술정보통신부는 네이버의 하이퍼클로바X(HyperCLOVA X)가 알리바바의 오픈소스 AI 시스템인 Qwen을 자사 모델에 포함시킨 사실이 드러나자, 정부의 핵심 소버린 AI 모델 선정 사업에서 네이버 클라우드를 탈락시킨 바 있다(G. Lee, 2026). 중국 기업들은 대규모 지식 증류(distillation)를 통해 미국 프론티어 모델의 능력을 체계적으로 추출한 뒤, 이를 누구나 내려받을 수 있는 오픈웨이트 모델로 공개해 왔다(Dunnmon, Narayan, and Saad-Falcon, 2026). 중국의 프론티어 AI 모델은 더 낮은 가격을 제시하면서 미국 모델과의 격차를 좁혀가는 것처럼 보일 수 있으나, 복잡한 작업을 완수하는 데 더 많은 토큰이 필요하다는 점을 고려할 때 비공개 평가 지표들은 실제 격차가 공개 벤치마크가 시사하는 것보다 더 크다는 점을 보여준다(The Economist, 2026a). 더욱이 Cisco와 CrowdStrike는 정치적으로 민감한 용어가 일반적인 코딩 프롬프트에 포함될 경우, 가장 널리 사용되는 중국 오픈웨이트 모델 중 하나인 DeepSeek-R1이 보안 취약점이 있는 코드를 생성할 확률이 50퍼센트 높아진다는 사실을 발견했다(Kassianik and Karbasi, 2025; Stein, 2025). 중국의 오픈소스 또는 오픈웨이트 솔루션에 의존한다는 사실을 드러내지 않는 소버린 AI 모델은 AI 주권을 달성하는 것이 아니라, 하나의 의존을 투명성이 낮고 잠재적으로 더 위험한 또 다른 의존으로 대체하는 셈이 되며, 이는 북한 해커들이 활용할 수 있는 공격 표면(공격에 악용될 수

있는 접점)을 넓힐 수 있다. 국내 AI 스택에서 불투명한 대외 의존을 줄이는 것은 한국 내 자산의 대북 유출을 억제하는 데 도움이 될 수 있으며, 유출된 자산의 상당 부분은 북한 정권의 무기 프로그램 자금으로 사용되고 있다(UN Panel of Experts, 2024).

## 북한과 그 밖의 관련 사이버 위협

미국 국가정보국장실(ODNI)의 2026년 연례 위협 평가 보고서는 북한의 사이버 프로그램을 "정교하고 기민하다"고 평가하며, 인공지능이 "사이버 영역의 위협을 가속화할 가능성이 크다"고 경고했다(Office of the Director of National Intelligence, 2026). 파이브 아이즈(Five Eyes) 정보기관 공동 성명 또한 AI로 위력이 증대된 사이버공격이 각국 정부와 기업이 대비하지 못한 임박한 위협이라고 경고한 바 있다(NCSC, 2026). 한편 소버린 AI 추진은 해커들에게 더 많고 어쩌면 더 손쉬운 표적을 제공할 수도 있다. 마이크로소프트 워드의 대안으로 개발된 한국의 문서작성 프로그램인 아래아한글(HWP)은 각급 정부 기관, 핵심 산업, 학계 전반에서 사용되어 오며, 2013년부터는 한국의 네트워크에 침투하려는 북한 해커들의 주요 공격 표적이 되어 악용되었다(Choi, 2025). 디지털 자립이라는 같은 논리에 따라 구축된 소버린 AI 모델은 이러한 취약성을 훨씬 더 큰 규모로 재현할 위험이 있다. 북한이 사이버 공격 역량을 확보하려는 주된 동기는 자금 확보인 것으로 보인다. 실제로 북한의 사이버 활동 상당수는 가상자산(암호화폐) 탈취와 연관되어 있다. 북한과 연계된 해커들은 2025년 한 해 동안 약 20억 달러 상당의 가상자산을 탈취했으며, 이는 그해 전 세계 가상자산 탈취액의 약 60퍼센트에 해당한다(Chainalysis, 2025). 불법적인 사이버 활동은 김정은 정권 전체 수입의 30퍼센트가량을 차지하는 것으로 추정된다(Multilateral Sanctions Monitoring Team, 2025). CrowdStrike의 2026년 금융서비스 위협 동향 보고서에 따르면,

FAMOUS CHOLLIMA 그룹은 AI로 만들어낸 신원 정보를 활용해 가상자산 거래소, 핀테크 기업, 소매은행에 침투함으로써 공격 횟수를 두 배로 늘렸다(CrowdStrike, 2026).

북한 정부는 사이버, 경제, 군사 분야에 AI 역량을 접목하려는 노력을 지속해 왔다. 2025년 8월, Anthropic은 북한 공작원들이 자사의 Claude 모델을 이용해 미국 포춘 500대 기업에 원격근무 직원으로 부정 취업했으며, 가짜 신원을 만들고 AI의 도움을 받아 코딩 평가를 통과했다고 확인했다(Anthropic, 2025). 2026년 5월, 구글은 북한의 해킹 그룹 APT45가 인공지능을 활용해 표적의 사이버보안 사각지대를 찾기 위해 수천 건의 프롬프트를 반복 입력하며 앞선 분석 결과를 다음 분석에 활용했다고 보고했다. 이는 국가의 지원을 받는 사이버 위협 행위자가 AI를 활용해 새로운 취약점을 대규모로 발견하고 악용한 최초로 확인된 사례였다(Kang, 2026).

북한의 임박한 위협 외에도, 한국은 중국과 러시아가 제기하는 위협 또한 고려해야 한다. 이는 이들이 북한의 조력자이기 때문일 뿐 아니라, 이들 국가의 사이버 활동 자체도 한국을 표적으로 삼고 있기 때문이다. TAG-74와 같은 중국 정부의 지원을 받는 해커들은 한국의 학계, 정치권, 정부 기관을 표적으로 한 사이버 첩보 활동을 벌여왔으며(Insikt Group, 2023), 여기에는 지식재산 탈취와 개인정보 침해가 수반된다. 중국이 오픈소스 방식을 활용하는 세계적인 AI 개발 선도국이라는 점은 이러한 위험을 더욱 키울 수 있다(Chan, 2026). 중국은 기술 경쟁자이자 안보상의 우려 대상으로서, 한국의 전략적 판단에 영향을 미치는 또 하나의 변수다. 한국을 겨냥한 것으로 확인된 러시아의 사이버공격 건수가 상대적으로 적다고 해서 러시아를 위협적이지 않은 존재로 간주할 수는 없다. 러시아는 북한의 사이버 역량 및 허위정보 활동에 있어 점점 더 중요한 조력자가 되고 있기 때문이다. 북한은 러시아어권 사이버 범죄 생태계를 활용하고 있으며, 이에 더해 마이크로소프트의 전문가들은 국가의 지원을 받는 북한과 러시아 해킹 그룹 간의 협력 정황을 관찰한 바 있다(Hüsch and Jarnecki, 2026).

## 한국의 소버린 AI 전략 재검토

한국의 소버린 AI 논의는 인프라의 지리적 분포, 예산 제약에 따른 공격자와 방어자 간 비대칭성, 빠르게 변화하는 핀테크 산업이라는 세 측면에서 북한의 사이버 위협을 더욱 충실히 고려해야 한다. 지리적 집중은 한국에 중대한 문제다. 소버린 AI는 데이터의 국내 저장을 요구하는데, 이는 핵심 AI 인프라를 좁고 공격 표적이 되기 쉬운 영토 안에 집중시키는 결과를 낳는다. 한국의 국토 면적은 약 10만 제곱 킬로미터에 불과하며, 수도권은 군사분계선 너머 북한 포병의 사정권에 놓여 있다. 최남단 제주도도 서울에서 약 450킬로미터 밖에 떨어져 있지 않아, 한국 전역이 북한이 보유한 다양한 미사일의 사정권에 들어간다. 2025년 9월 대전에 위치한 국가정보자원관리원 데이터센터에서 리튬이온 배터리 폭발로 발생한 화재는 709개의 정부 서비스를 마비시켰으며, 완전한 복구까지 95일이 걸렸다(Ministry of the Interior and Safety, 2025). 우발적인 화재가 이 정도의 혼란을 야기할 수 있었다면, 사이버공격, 파괴공작, 드론, 미사일 등을 통한 의도적 공격은 심각한 피해를 초래할 수 있다. 이와 대조적으로 미국의 데이터센터는 훨씬 더 넓은 영토에 분산되어 있어 북한의 물리적·사이버 공격에 훨씬 덜 취약하다. 러시아와 가까운 지정학적으로 취약한 위치에 있어 전략적 종심이 부족한 에스토니아는 이러한 위험을 인식하고, 국내 인프라가 공격받거나 파괴되더라도 정부 기능이 계속 유지되도록 룩셈부르크에 "데이터 대사관"을 설립했다(O'Regan, 2026). 높은 역량을 갖추고 활발하게 활동하는 사이버 적대세력에 직면해 있음에도 불구하고, 한국에는 이에 상응하는 조치가 마련되어 있지 않다. 한국의 개인정보 보호법(PIPA, 2023) 제28조의8 및 제28조의9는 개인정보의 국외 이전을 제한하고 있으며, 전자정부법, 국가정보원법 및 관련 보안 규정들 또한 정부 시스템에 대한 추가적인 법적 제약을 부과하고 있다.

더욱이 제한된 예산으로 개발된 소버린 AI 모델은 AI를 활용하는 적대세력을 상대할 때 공격자와 방어자 간 비대칭성으로 인해 불리해질 수 있다. 과학기술정보통신부와 한국인터넷진흥원(KISA)은 공개적으로 이용 가능한 AI 모델을 활용해 한 한국 기업이 개발 중인 소버린 AI 모델에 대한 모의공격을 수행해 약 10분 만에 취약점 7개를 발견했다(Seo, 2026). 한국의 소버린 AI 모델들은 OpenAI, Google, Anthropic의 모델에 비해 훨씬 적은 예산과 작은 이용자 기반을 바탕으로 개발되고 있다. 만약 공격 작전에 프론티어 AI 모델을 활용하는 북한 해커들의 표적이 된다면, 한국 기업들은 불리한 위치에 놓이게 될 것이다. 한국은 대규모 사이버 침해사고를 반복적으로 겪고 있다. 사고 건수는 연간 26퍼센트씩 증가하고 있으며 2025년까지 신고된 누적 건수는 2,383건에 달했다(Korea Internet & Security Agency, 2026). 라자루스 그룹(Lazarus Group)의 진화 양상은 북한 해커들이 한국의 소프트웨어 생태계에 대한 깊은 이해를 어떻게 활용하는지를 보여준다. 예를 들어 "싱크홀 작전(Operation SyncHole)"에서 공격자들은 한국의 온라인 뱅킹 및 정부 웹사이트에서 널리 사용되는 보안 도구를 악용하여 금융, IT, 반도체, 통신 부문에 걸쳐 최소 6개 기관・기업의 시스템에 침투했다 (Kaspersky, 2025).

금융 부문의 AI 도입은 북한의 사이버 작전에 악용될 새로운 취약점을 낳을 수 있다. 2025년 11월, 한국 최대의 디지털자산 거래소인 업비트(Upbit)는 솔라나(Solana) 네트워크상에서 3,060만 달러 상당의 토큰을 해킹으로 탈취당했다. 이후 한국 당국은 라자루스 그룹을 그 배후로 지목했다(Kang, 2025). 한편 주요 금융기관들은 자사 시스템에 AI를 더욱 깊이 통합하고 있다. 카카오뱅크는 "AI 네이티브 뱅크"로의 전환을 추진하고 있으며, KB국민은행, 신한은행, 하나은행, 우리은행 등 주요 금융그룹의 수장들은 모두 2026년 신년사에서 AI 및 디지털 전환 가속화에 대한 의지를 표명했다(Jun, 2026). 금융 자산 탈취가 북한의 주요 사이버 작전 목표 중 하나라는 점을

고려하면, 한국 핀테크 산업에 소버린 AI를 도입하는 것은 해당 시스템을 북한이 높은 역량을 갖추고 활발히 활동하며 공격 유인도 강한 영역에 노출시킬 수 있다.

## 결론

소버린 AI를 추진하는 국가는 한국만이 아니다. 신미국안보센터(CNAS)의 소버린 AI 지수(Sovereign AI Index)에 따르면, 2026년 6월 기준 67개국이 이러한 프로젝트를 진행하고 있다(Chavez et al., 2026). 대부분은 특정 수요에 대응하고 대외 의존 문제를 해결하는 데 목적을 둔다. 소버린 AI 모델을 처음부터 구축하는 포괄적 사업은 드물며, 상당수는 데이터센터를 중심으로 자국 내 인프라를 개선하는 데 초점을 맞춘다. 프랑스와 아랍에미리트(UAE)는 모두 AI 주권을 표방하면서도 미국 기업들과의 파트너십을 토대로 전략을 수립하고 있다. 프랑스는 1,090억 유로가 넘는 AI 인프라 투자를 발표했으며 (Élysée Palace, 2025), NVIDIA의 GPU, 실리콘밸리의 솔루션, 외국 자본에 의존해 미스트랄(Mistral)과 같은 자국의 선도 기업을 육성하고 있다. UAE는 최초의 “OpenAI for Countries” 파트너십으로 "스타게이트 UAE(Stargate UAE)"를 출범시켰으며(OpenAI, 2025), 자국의 대표 AI 기업인 G42에 마이크로소프트로부터 15억 달러의 투자를 유치했다(Microsoft, 2024). 두 나라 모두 주권을 자급자족 체제(autarky)와 동일시하지 않으며, 최첨단 컴퓨팅, 클라우드, 모델 생태계가 앞으로도 국경을 넘어 운영된다는 점을 인정한다. 그러면서도 주권은 거버넌스, 안전성 평가, AI 도입・운영 결정, 전략적 데이터에 대한 통제권으로 정의한다. 프랑스와 UAE 모두 북한이 제기하는 것과 같은 복합적 위협에 직면해 있지 않은 만큼, 한국은 자국 소버린 AI 전략의 사이버보안 아키텍처에 대해 더욱 실용적으로 접근할 필요가 있다.

한국 정부가 야심찬 소버린 AI 정책을 추진하는 것은 타당하지만, 이를 위해서는 주권을 민족주의적 관점에 얽매이지 않고 유연하게 정의해야 한다. AI 스택의 모든 계층이 전적으로

국내에서 개발·생산되고 국내에 위치해야 한다는 경직된 접근은 오히려 역효과를 낳을 것이다. 그러한 접근은 핵심 인프라를 좁고 공격 표적이 되기 쉬운 지역 안에 집중시키고, 이미 기존 디지털 시스템조차 방어하는 데 어려움을 겪고 있는 사이버보안 네트워크에 의존하는 결과를 낳을 것이다. 한국은 자국의 대표 IT 기업 육성을 우선시한 결과, 주요 사업자 한 곳에서 장애가 발생하면 국내 디지털 생태계의 상당 부분이 타격을 받을 수 있는 소수 기업 중심의 플랫폼 구조를 형성해 왔다(Merrill, 2025). 실현 가능한 소버린 AI 전략은 AI 스택의 모든 계층에서 한국의 자율적 의사결정권, 규제 영향력, 수익 창출 기회를 보장해야 한다. 이는 한국 기업들이 보안 문제가 있는 오픈웨이트 모델의 사용에 따른 위험을 줄이는 한편, 미국, 일본, 유럽 기업들과 협력할 수 있는 재량을 확대하며 달성될 수 있다. 투입되는 자원의 규모를 고려할 때, 정부는 여론뿐 아니라 고용시장과 개인정보 보호에 미치는 영향에도 유의해야 한다. 최근 한국에서 발생한 데이터 유출 사건들은 그 수가 많고 심각했으며, 쿠팡 사건의 경우 4억 달러가 넘는 과징금이 부과될 만큼 중대한 사안이었다(Lim, 2026). 법적·제도적 인프라는 기술 도입의 속도와 사이버 취약성의 확대를 따라잡는 데 어려움을 겪고 있다. 한국은 이미 오래전부터 북한의 핵무기를 홀로 억지할 수 없다는 점을 인식해 왔다(Easley, 2025). 마찬가지로 북한의 사이버 위협 역시 홀로 마주해서는 안 된다. 미국을 비롯해 가치를 공유하는 국가의 정부 및 이들 국가의 신뢰할 수 있는 기업들과 협력하는 것이 한국이 소버린 AI를 보다 안전하게 추진할 수 있는 길이다. ■

## 참고문헌


Anthropic. 2025. “Detecting and Countering Misuse of AI: August 2025.” August 27. https://www.anthropic.com/news/detecting-countering-misuse-aug-2025

Bae, Sunha, and So Jeong Kim. 2025. “AI Security Strategy and South Korea’s Challenges.” Center for Strategic and International Studies. June 20. https://www.csis.org/analysis/ai-security-strategy-and-south-koreas-challenges

Chainalysis. 2025. “North Korea Drives Record $2 Billion Crypto Theft Year, Pushing All-Time Total to $6.75 Billion.” December 18. https://www.chainalysis.com/blog/crypto-hacking-stolen-funds-2026/

Chan, Kyle. 2026. “China Is Running Multiple AI Races.” Brookings Institution. March 9. https://www.brookings.edu/articles/china-is-running-multiple-ai-races/

Chavez, Pablo. 2026. “The Sovereignty Gap in U.S. AI Statecraft.” *Lawfare*. February 16. https://www.lawfaremedia.org/article/the-sovereignty-gap-in-u.s.-ai-statecraft

Chavez, Pablo, Vivek Chilukuri, and Ruby Scanlon. 2026. “Sovereign AI’s Second Wave Is Coming Into View.” Center for a New American Security, August 20. https://www.cnas.org/publications/cnas-insights/cnas-insights-sovereign-ais-second-wave-is-coming-into-view

Chey Institute for Advanced Studies. 2026. “AI 주권 시대, 대한민국의 선택 [The Age of AI Sovereignty: South Korea’s Choice].” Chey Institute for Advanced Studies. January 14. https://www.chey.org/Kor/BooksnReports/BooksnReportsView.aspx?seq=27&pageno=1

Choi, Perry. 2025. “HWP as an Attack Surface: What Hancom’s Hangul Word Processor Means for South Korea’s Cyber Posture as a US Ally.” *38 North*. October 27. https://www.38north.org/2025/10/hwp-as-an-attack-surface-what-hancoms-hangul-word-processor-means-for-south-koreas-cyber-posture-as-a-us-ally/

Cloud Security Alliance (CSA) AI Safety Initiative. 2026. “When Test Environments Leak: Frontier AI Models Hack Real Firms.” August 7. https://labs.cloudsecurityalliance.org/research/csa-research-note-frontier-ai-models-hacking-real-systems-ev/

CrowdStrike. 2026. “2026 Financial Services Threat Landscape Report.” May 12. https://www.crowdstrike.com/en-us/resources/reports/crowdstrike-2026-financial-services-threat-landscape-report/

Draudt-Véjares, Darcie, and Seungjoo Lee. 2026. “Governing AI in the Shadow of Giants: Korea’s Strategic Response to Great Power AI Competition.” Carnegie Endowment for International Peace. April 30. https://carnegieendowment.org/research/2026/04/governing-ai-in-the-shadow-of-giants-koreas-strategic-response-to-great-power-ai-competition

Dunnmon, Jared, Avanika Narayan, and Jon Saad-Falcon. 2026. “China’s AI Heist.” *Foreign Affairs.* May 29. https://www.foreignaffairs.com/china/chinas-ai-heist

Easley, Leif-Eric. 2025. “‘Getting Asia Right’ With a More Strategic US-South Korea Alliance.” *Journal of East Asian Affairs* 38(1): 125-165. https://www.inss.re.kr/en/publications/bbs/joeaa_en_view.do?nttId=41037638

The Economist. 2026a. “China is having another AI moment.” *The Economist*. June 21. https://www.economist.com/china/2026/06/21/china-is-having-another-ai-moment

The Economist. 2026b. “Donald Trump Has Cut Off Access to the World’s Best AI Model.” *The Economist*. June 14. https://www.economist.com/business/2026/06/14/donald-trump-has-cut-off-access-to-the-worlds-best-ai-model

Élysée Palace. 2025. “Make France an AI Powerhouse.” February 11. https://www.elysee.fr/en/emmanuel-macron/2025/02/11/make-france-an-ai-powerhouse

Froman, Michael. 2026. “The AI Sovereignty Paradox at Home and Abroad.” Council on Foreign Relations. February 27. https://www.cfr.org/articles/the-ai-sovereignty-paradox-at-home-and-abroad

Girishankar, Navin. 2025. “How AI Cooperation Can Save the U.S.-ROK Trade Talks.” Center for Strategic and International Studies. July 15. https://www.csis.org/analysis/how-ai-cooperation-can-save-us-rok-trade-talks

Han, Kwang-beom. 2026. “‘AI 토큰 비용이 새 무역적자 된다’…라이너 ‘한국 AI 생존 전략은 내재화’ [‘AI Token Costs Could Become a New Trade Deficit’…Liner: ‘Korea’s AI Survival Strategy Is Internalization’].” *Edaily*, August 5. https://www2.edaily.co.kr/News/Read?mediaCodeNo=257&newsId=04890486645544368

Hüsch, Pia, and Joseph Jarnecki. 2026. “DPRK and Russian Collaboration in Cyberspace as a Driver for UK-ROK Cyber Cooperation.” *38 North*. March 4. https://www.38north.org/2026/03/dprk-and-russian-collaboration-in-cyberspace-as-a-driver-for-uk-rok-cyber-cooperation/

Insikt Group. 2023. “Multi-year Chinese APT Campaign Targets South Korean Academic, Government, and Political Entities.” *Recorded Future*. September 19. https://www.recordedfuture.com/research/multi-year-chinese-apt-campaign-targets-south-korean-academic-government-political-entities

Jie, Ye-eun. 2025. “Korea to Begin Full-Scale AI Buildout in 2026: Minister.” *Korea Herald*. December 15. https://www.koreaherald.com/article/10636733

Jun, Ji-hye. 2026. “Korea’s Top Financial Groups Pledge Overhaul as AI, Productive Finance Take Center Stage in 2026.” *Korea Times.* January 2. https://www.koreatimes.co.kr/business/banking-finance/20260102/koreas-top-financial-groups-pledge-overhaul-as-ai-productive-finance-take-center-stage-in-2026

Kang, Jae-eun. 2026. “N.K. hackers using AI to find cybersecurity blind spots, Google says.” *Yonhap*. May 12. https://en.yna.co.kr/view/AEN20260512005600320

Kang, Yoon-seung. 2025. “N. Korean hacking group Lazarus suspected behind recent crypto hacking: sources.” *Yonhap*. November 28. https://en.yna.co.kr/view/AEN20251128003952320

Kaspersky. 2025. “Kaspersky Uncovers New Lazarus-Led Cyberattacks Targeting South Korean Supply Chains.” April 24. https://www.kaspersky.com/about/press-releases/kaspersky-uncovers-new-lazarus-led-cyberattacks-targeting-south-korean-supply-chains

Kassianik, Paul, and Amin Karbasi. 2025. “Evaluating Security Risk in DeepSeek and Other Frontier Reasoning Models.” Cisco Blogs. January 31. https://blogs.cisco.com/security/evaluating-security-risk-in-deepseek-and-other-frontier-reasoning-models

Kharpal, Arjun. 2025. “South Korea Charts One-of-a-Kind Course in AI Race with U.S. and China.” *CNBC*. August 8. https://www.cnbc.com/2025/08/08/south-korea-to-launch-national-ai-model-in-race-with-us-and-china.html

Kim, Haeyoon. 2025. “The Sovereign AI Debate and Prospects of ‘Korean’ AI.” Korea Economic Institute of America. January 8. https://keia.org/analysis/the-sovereign-ai-debate-and-prospects-of-korean-ai/

Kim, J. James and Sean (Hyuk Hyun) Kwon. 2026. “From Compute to Capacity: South Korea’s Approach to Industrial AI Adoption.” Stimson Center. February 6. https://www.stimson.org/2026/from-compute-to-capacity-south-koreas-approach-to-industrial-ai-adoption

Korea Internet & Security Agency. 2026. “25 년 사이버 위협 하반기 동향 및 26 년 전망 [2025 Cyber Threat Trends and 2026 Cyber Threat Outlook Report].” February 6. https://www.kisa.or.kr/skin/doc.html?fn=20260206_153449_003.pdf&rs=/result/2026-02/

Lee, Gyu-lee. 2026. “National AI Model Project Faces Controversies over Plagiarism Claims.” *Korea Times.* January 7. https://www.koreatimes.co.kr/business/tech-science/20260107/national-ai-model-project-faces-controversies-over-plagiarism-claims

Lee, Hun-il. 2026. “하정우: ‘최대성과, GPU 26 만장 확보…독자 AI 없으면 국가안보 문제’ [Ha Jung-woo: ‘Our Biggest Achievement is Securing 260,000 GPUs…Without Independent AI, National Security is at Risk’].” *The Fact*. April 21. https://news.tf.co.kr/read/ptoday/2315267.htm

Lee, Jae-myung. 2026. “San Francisco Declaration on Artificial Intelligence.” Office of the President of the Republic of Korea. July 24. https://en.president.go.kr/president/statements-remarks/a2eIOLXS

Lim, Sun-young. 2026. “메타보다 더 세게 맞았다, 쿠팡 과징금 6247 억 ‘역대 최대’ [Hit Harder than Meta: Coupang Fined ₩624.7 Billion, the Largest Ever].” *JoongAng Ilbo*. June 12. https://www.joongang.co.kr/article/25436101

Merrill, Nick. 2025. “The South Korean Digital Paradox: How South Korea’s Internet Development Model Creates Unique Cybersecurity Vulnerabilities.” Center for Long-Term Cybersecurity, University of California, Berkeley. April. https://cltc.berkeley.edu/publication/the-south-korean-digital-paradox-how-south-koreas-internet-development-model-creates-unique-cybersecurity-vulnerabilities/

Microsoft. 2024. “Microsoft Invests $1.5 Billion in Abu Dhabi’s G42 to Accelerate AI Development and Global Expansion.” April 16. https://news.microsoft.com/source/2024/04/16/microsoft-invests-1-5-billion-in-abu-dhabis-g42-to-accelerate-ai-development-and-global-expansion/

Ministry of the Interior and Safety. 2025. "국가정보자원관리원 709 개 시스템 전체 복구 완료 [National Information Resources Services Completes Full Recovery of All 709 Systems]." December 30. https://www.mois.go.kr/frt/bbs/type010/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000008&nttId=122775

Ministry of Science and ICT (MSIT). 2024. “MSIT Presents Blueprint for Korea’s AI Innovation to Achieve AI G3 Status.” September 26. https://www.msit.go.kr/eng/bbs/view.do?bbsSeqNo=42&mId=4&mPid=2&nttSeqNo=1040&pageIndex&sCode=eng&searchOpt=ALL&searchTxt

_____. 2025. “MSIT Selects Five Elite Teams for the ‘Sovereign AI Foundation Model’ Project.” August 4. https://www.msit.go.kr/eng/bbs/view.do;jsessionid=kndLBpcdYr-gB9iqSygkTqBG4r82Uq5403hvdHoT.AP_msit_1?sCode=eng&mPid=2&mId=4&bbsSeqNo=42&nttSeqNo=1152

Multilateral Sanctions Monitoring Team. 2025. “The DPRK’s Violation and Evasion of UN Sanctions through Cyber and Information Technology Worker Activities.” October 22. https://msmt.info/Publications/detail/MSMT%20Report/4221

National Cyber Security Centre (NCSC). 2026. *The AI Shift in CyberRisk: Why Leaders Must Act Now*. June 22. https://www.ncsc.gov.uk/news/the-ai-shift-in-cyber-risk-why-leaders-must-act-now

Office of the Director of National Intelligence. 2026. *Annual Threat Assessment of the U.S. Intelligence Community.* March 18. https://www.dni.gov/index.php/newsroom/reports-publications/reports-publications-2026/4141-2026-annual-threat-assessment

OpenAI. 2025. “Introducing Stargate UAE.” May 22. https://openai.com/index/introducing-stargate-uae/

O’Regan, Ellen. 2026. “A nation on a hard drive: Inside the rise of digital embassies.” *Politico EU*. May 26. https://www.politico.eu/article/estonia-russia-luxembourg-hard-drive-data-foreign-vaults/

Personal Information Protection Act (PIPA). 2023. Act No. 19234. Korea Legislation Research Institute, Korea Law Translation Center. https://elaw.klri.re.kr/eng_service/lawView.do?hseq=62389&lang=ENG

Seo, Byeong-ju. 2026. “보안 위협 된 AI…‘독자 모델 개발·국제협력’ 폭 키우는 정부 [AI Becomes a Security Threat: Government Expands Independent Model Development and International Cooperation].” *Asia Today*. May 12. https://www.asiatoday.co.kr/kn/view.php?key=20260512010003082

Stanford Institute for Human-Centered Artificial Intelligence (HAI). 2026. *The 2026 AI Index Report.* Stanford University. April. https://hai.stanford.edu/ai-index/2026-ai-index-report

Stein, Stefan. 2025. “CrowdStrike Research: Security Flaws in DeepSeek-Generated Code Linked to Political Triggers.” CrowdStrike. November 20. https://www.crowdstrike.com/en-us/blog/crowdstrike-researchers-identify-hidden-vulnerabilities-ai-coded-software/

UN Panel of Experts. 2024. *Final Report of the Panel of Experts Submitted Pursuant to Resolution 2680*. S/2024/215. United Nations Security Council. March 7. https://undocs.org/S/2024/215

Varela Sandoval, Francisco Javier, Isabella Wilkinson, Alex Krasodomski, and Rowan Wilkinson. 2026. “How Middle Powers Can Weather US and Chinese AI Dominance: The Case for ‘Sovereign AI’ Strategies.” Chatham House. February. https://www.chathamhouse.org/sites/default/files/2026-02/2026-02-13-sovereign-ai-strategies-sandoval-et-al.pdf

Yonhap. 2026. “PM vows to speed up AI transformation in public sector.” July 9. https://en.yna.co.kr/view/AEN20260709004251315

■ 저자: 레이프에릭 이슬리(Leif-Eric EASLEY, 하버드대학교 정치학 박사)는 이화여자대학교 교수로 재직 중이며, 국제안보와 정치경제학을 강의하고 있다. 최다빈(Dabin CHOI)은 이화여자대학교 국제학 대학원 석사과정생으로, 아시아의 사이버보안과 AI를 연구하고 있다.

■ 담당 및 편집: **임재현**_EAI 연구원; **오인환**_EAI 수석연구원
문의: 02-2277-1683 (ext. 209) | jhim@eai.or.kr



발행일 2026년 9월 28일 979-11-7584-156-7 95340
“한국의 소버린 AI 추진과 대북 사이버 위협 취약성 증대 요인 분석”

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