

입법의견서

「국방인공지능법」 제정안에 대한 입법의견서

국제법 준수, 책임성 조항 등 법안 수정
되어야
고위험 군사AI의 금지·제한 규정 추가
국방인공지능위원회의 독립성과 민주적
통제 제도화 필요

발행일 | 2026년 5월 14일 (목)

발행 | 인공지능 책임성과 공공성 강화를 위한 시민사회

공동행동(AI시민행동)

들어가며	3
「국방인공지능법안」의 주요 쟁점과 개선 방향	4
1. 국가안보 중심에서 책임성 중심으로 목적 조항의 방향 전환	4
2. ‘자율성’, ‘인간통제’ 등 핵심 개념 정의 조항 추가	6
3. 모호한 ‘인적 개입’ 개념을 구체화하고 살상 결정 위임 불허 원칙도 명시	7
4. 국방인공지능위원회의 독립성과 민주적 통제 제도화	10
5. 개발 및 특례 조항의 제한적 적용 및 군사AI 금지·제한하는 규정 추가	12
6. 안전성 확보 의무조항 강화 및 피해 책임 범위와 기준 명확화	14
7. 특정 국방인공지능에 대한 금지 규정 필요	17
8. 영향받는 자의 권리구제와 책임자 처벌	18
나가며	19

‘인공지능 책임성과 공공성 강화를 위한 시민사회 공동행동(AI시민행동)’은 2026년 3월 31일 발족한 연대기구로 녹색연합, 디지털정의네트워크, 민주사회를 위한 변호사 모임, 참여연대, 한국여성단체연합 등 인권, 노동, 복지, 여성, 환경, 소비자, 평화 분야 전국 46개 단체들이 함께 하고 있습니다. AI시민행동은 인공지능 개발과 활용에서 책임성과 공공성 강화, 민주적 거버넌스 구축, 민주적 공론장 형성을 목표로 활동하며, 인공지능 개발과 활용을 명목으로 규제를 완화하고 그 영향받는 사람과 지역의 권리를 소홀히 취급하는 등 한쪽으로 치우친 정부 정책을 균형있는 사회정책으로 전환할 것을 요구하고 있습니다. [\[자세히보기\]](#)
(문의 : AI시민행동 사무처 참여연대 공익법센터 이지은 간사 02-723-0666)

- 지난 2026년 1월 22일「인공지능기본법」이 전면시행되었으나, 해당 법에는 국방 및 국가안보 목적의 인공지능은 적용 대상에서 제외됨. 따라서 국방인공지능에 대한 별도의 규율이 필요한 실정임¹. 애초에 기본법을 공통 규범으로 제정하고 국방분야 인공지능에 한하여 필요한 예외적인 규율을 별도로 제정하는 것이 바람직하나, 현행 기본법이 이를 다 포괄하지 못해 별도 입법이 이뤄져야 하는 상황임.
- 현재 국회에는 더불어민주당 부승찬 의원과 국민의힘 유용원 의원이 공동으로 대표 발의한 「국방인공지능법안」이 발의되어 있음. 발의된 법안에는 국방인공지능을 통제하거나 규제하기 위한 내용은 거의 없고 사실상 국방인공지능 개발과 기술 발전을 촉진한다고 볼 수밖에 없는 조항들이 다수 존재해 ‘산업 육성법’에 가깝다고 평가받고 있음.
- 전 세계적으로 인공지능을 활용한 무기 개발과 투자가 크게 증가하고, 실제 전장에 투입·사용되는 사례가 늘면서 여러가지 법적·윤리적 쟁점이 발생하고 있음. 이런 상황에서 국방인공지능 정책의 기본이 되는 법안을 제정하면서 국방인공지능에 대한 규제와 통제를 규율하는 것은 향후 당면할 위험성을 대비하기 위해서라도 매우 중요함. 특히 국방인공지능 정책은 인간 생명과 기본권에 직접적인 영향을 미칠 수 있으므로 국방인공지능 운용에 대한 기본 원칙과 방향, 금지·제한되는 인공지능에 대한 정의, 위험한 인공지능 기술에 대한 규제와 통제 방안, 국가와 기업의 책임 등에 대한 내용이 법안에 반드시 포함되어야 함.
- 국방인공지능은 유엔 헌장, 국제인도법, 국제인권법을 비롯하여 국제법 및 국내법 체제에 합치하는 방식으로 규제되어야 하며, 특히 자율 무기 체계에 관하여 최근 국제 사회가 권고 및 논의하고 있는 금지·제한의 방향을 반영해야 함.

¹ 이런 식으로 기본법에서 적용을 배제하고 별도의 특별법을 제정하는 것은 비효율적이고 법률간 일관성을 저해할 수 있으며, 오히려 규율의 공백을 야기할 수 있음. 국방분야 인공지능의 특수성에도 불구하고 일반 인공지능과 공통적으로 적용되어야 할 규범이나 절차 역시 존재함. 공통 규범을 본 법안에서 별도로 규율할 경우 중복으로 인한 비효율이 발생할 수 있고, 또는 반드시 포함되어야 하는 공통 규범이 본 법안에서 다루어지지 않을 수 있음. 실제로 본 법안은 고영향 인공지능에 해당하는 국방인공지능을 개발, 제공하는 사업자에 대한 책무성 조항, 국방인공지능의 투명성 의무 등을 전혀 포함하고 있지 않음. 따라서 국방 분야의 특수성을 고려하되 인공지능 규율의 통일성을 유지하고 규율의 공백을 최소화하기 위해서는 인공지능기본법을 모든 인공지능에 원칙적으로 적용하고, 국방 분야 인공지능에 한하여 필요한 예외적인 규율을 본 법안에서 규정하는 방식이 바람직함.

- 이에 인공지능 책임성과 공공성 강화를 위한 시민사회 공동행동(AI공동행동)은 아래와 같이 현재 발의된 법안의 문제점을 지적하는 의견서를 제출함.

「국방인공지능법안」의 주요 쟁점과 개선 방향

1. 국가안보 중심에서 책임성 중심으로 목적 조항의 방향 전환

제1조(목적) 이 법은 국가 경쟁력의 혁신기반인 인공지능 기술의 국방분야 개발·도입 및 운용에 관한 기본사항을 정하고, 국방인공지능의 안전성·신뢰성 및 책임성 확보를 위한 관리체계를 정립함으로써 국민의 안전을 보호하고 국방력 증진과 국가안보 강화에 이바지한다.

- <국방인공지능법안> 제1조 목적 조항에는 ‘국방력 증진과 국가안보 강화’를 전면에 내세우고 ‘안전성·신뢰성 및 책임성’은 이를 보완하는 요소로 규정하고 있음. 그러나 국방인공지능 분야 기본법이 되는 법률이고 일반적으로 군사 분야에 인공지능 기술이 사용되는 것은 고위험 분야로 분류되는 민감한 사안으로 인공지능 기술이 가진 한계와 문제점에 대한 고려를 법 제정의 목적에 충실히 담는 것이 마땅함.
- 국제 시민사회는 이미 2000년대 후반부터 인공지능 무기의 위험성을 경고하고, 자율무기체계에 대한 기술적·법적·윤리적 논의의 필요성을 제기해왔음. 2014년부터 유엔 특정재래식무기금지협약(CCW)을 중심으로 자율살상무기체계(LAWS)를 비롯한 군사 영역에서의 인공지능 규제에 대한 논의를 시작했으며 이에 대한 논의가 10년 넘게 이어지고 있음. 최근에는 국제인도법과 국제인권법을 아우르는 신규 국제협약 제정을 위한 노력들도 이루어지고 있음. 안토니오 구테흐스 유엔 사무총장 역시 인공지능 무기, 특히 자율살상무기체계의 위험성에 대해 지속적으로 언급해왔으며 2026년까지 법적 구속력이 있는 국제 협약을 체결할 것을 각국에 촉구하기도 했음².
- 이러한 추세 속에 한국 정부는 지난 2024년 네덜란드와 함께 ‘인공지능의 책임있는 군사적 이용에 관한 고위급회의(REAIM)’를 공동주최하며, ‘행동을 위한 청사진(Blueprint for Action)’을 채택함. 이 문서에는 군사 인공지능을 사용할 때

² UN, <A New Agenda for Peace>, 2023.7.20

<https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-new-agenda-for-peace-en.pdf>

반드시 유엔헌장과 국제인도법, 국제인권법 등을 지켜야 한다는 내용이 담겨있음³. 뿐만 아니라 한국 정부는 2024년 유엔 총회에서 ‘군사분야 인공지능과 국제평화 및 안보에 대한 함의(Artificial Intelligence in the Military Domain and its Implications for International Peace and Security)’라는 결의안(A/RES/79/239) 채택을 공동 주도함. 해당 결의안은 국제법이 인공지능의 생애주기 전반에 걸쳐 발생하는 모든 사안에 적용된다고 확인하며, 무력사용에 대한 인간의 판단과 통제에 관한 조치를 포함한 안전장치(safeguards)를 강조하고 있음. 더구나 2023년 채택된 유엔인권이사회 결의안(A/HRC/RES/53/29)은 한국 정부 주도로 채택되었는데, 해당 결의문에는 특정 AI 활용에 있어 ‘수용불가능한 위험(unacceptable risks)’을 명시적으로 밝히고 있음.

- 문제는 한국 정부가 국제사회에서 약속하거나 결의한 안전장치, 국제법 준수 등의 내용을 국내법을 제정하는 데에는 전혀 반영하지 않는다는 것임. 위에서 언급한 원칙을 국방인공지능법에 명시하지 않는 것은 국제사회에서의 약속과 국내 이행 사이의 간극만 넓힐 뿐임. 한국이 국제사회에서 주창해오던 내용은 국내 국방인공지능 정책에서도 동일한 원칙과 기준을 적용해 반영해야 함.
- 제네바협약 제1추가의정서 제36조는 “신무기, 전투수단 또는 방법의 연구·개발·획득 및 채택에 있어서 체결당사국은 동 무기 및 전투수단의 사용이 본 의정서 및 체결당사국에 적용 가능한 국제법의 다른 규칙에 의하여 금지되는지의 여부를 결정할 의무가 있다”로 규정함. 이 조항에 따르면 국가들은 새로운 무기·수단·방법이 민간인 보호, 무차별적 공격 금지 등 국제인도법을 준수하는지, 군사적 목적에 맞는지, 민간인에게 불필요한 피해를 주지 않는지 등에 대해 사전에 검토해야 함. 국제 인도법은 기술이 급격히 발전하는 상황에서도 민간인을 전쟁으로부터 보호하기 위해 필수적이고 허용되는 무기와 방법을 제한하고 있으며, 모든 국가들은 제1추가의정서 비준 여부와 관계없이 신무기의 적법성 평가에 대한 이해관계를 갖고 있음. 신기술의 발전 속도를 고려할 때 신무기에 대한 법적 검토의 수행은 매우 중요함⁴.

● 수정 요구 사항

- 제1조 목적 조항은 ‘국방분야에서의 인공지능 기술 개발 및 활용에 관한 기본원칙을 수립하고, 실질적이고 책임있는 인간 통제와 국제법 준수를 보장함으로써 인공지능

³ 외교부, 보도자료 <2024 인공지능의 책임있는 군사적 이용에 관한 고위급회의(REAIM) 결과>, 2024.09.10

https://www.mofa.go.kr/www/brd/m_4080/view.do?seq=375378&page=1&pitem=10

⁴ 한국법제연구원, <자율살상무기의 국제법적 쟁점>, 2019.09.30

기반 무기체계의 안전성·책임성·투명성을 확보하고, 민간인 보호와 인권 존중에 부합하는 책임있는 국방인공지능 활용 체계를 확립하는 데 이바지한다’로 수정해야 함.

- 한국이 2024년 REAIM 정상회의와 유엔총회 결의안에서 약속한 바를 국내법적으로 이행할 수 있도록 >국제인도법 및 국제인권법 준수, >제네바협약추가개정서 제36조에 따른 무기검토 의무, >민간인 보호 등은 ‘국방인공지능 개발과 활용의 원칙’으로 별도의 조문으로 명문화해야 함.

2. ‘자율성’, ‘인간통제’ 등 핵심 개념 정의 조항 추가

제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

1. “인공지능”이란 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」 제2조제1호에 따른 인공지능을 말한다.
2. “인공지능기술”이란 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」 제2조제3호에 따른 인공지능기술을 말한다.
3. “국방인공지능”이란 군사적 목적을 달성하기 위하여 무기체계, 전력지원체계 및 국방정보시스템 등에 도입되거나 연계되어 운용되는 인공지능을 말한다.
4. “국방인공지능시스템”이란 국방인공지능을 활용하여 다양한 국방 분야에서 실제 및 가상환경에 영향을 미치는 예측, 추천, 결정 등의 결과물을 추론하는 군사적인 목적의 인공지능 기반 시스템을 말한다.
5. “국방인공지능기술”이란 국방인공지능을 구현하기 위하여 필요한 하드웨어·소프트웨어 기술 또는 그 활용 기술을 말한다.
6. “국방인공지능사업”이란 국방 목표 달성을 위하여 인공지능기술을 연구개발, 도입, 운용 또는 활용하는 일체의 활동을 말한다.

- 제2조(정의) 제3호는 ‘국방인공지능’을 “군사적 목적을 달성하기 위하여 무기체계, 전력지원체계 및 국방정보시스템 등에 도입되거나 연계되어 운용되는 인공지능”이라고 포괄적으로 규정하고 있음. 한계가 있다고 지적되고 있는 「인공지능기본법」 조차 위험 수준에 따라 ‘고위험 인공지능’을 별도로 정의하고 있는 것과 비교할 때 개념 정의 부분이 많이 부족함. 특히 국방인공지능 규범 논의의 핵심 개념인 ‘자율성(autonomy)’, ‘인간통제(human control)’, ‘AI 기반 무기체계(AI-enabled weapon system)’ 등에 대한 정의조차 규정하지 않은 것은 큰 문제임.
- 한편, 제2조 제4호는 ‘국방인공지능시스템’을 “국방인공지능을 활용하여 다양한 국방 분야에서 실제 및 가상환경에 영향을 미치는 예측, 추천, 결정 등의 결과물을 추론하는 군사적인 목적의 인공지능 기반 시스템”이라고 정의하고 있음. 이는

인공지능기본법 상 '인공지능시스템'의 정의와 유사하지만 "다양한 수준의 자율성과 적응성"을 가진다는 요소를 포함하지 않으며, 현재 국방인공지능시스템의 핵심 기능으로 논의되는 탐지, 식별, 추적, 표적화, 타격과 같은 기능을 충분히 반영하지 못하는 한계가 있음.

- 나아가 제2조 정의에서 고영향 인공지능에 해당하는 국방인공지능을 개발, 제공하는 인공지능사업자, 고영향 인공지능을 이용하는 이용자(국방부 및 산하기관), 그리고 영향받는 자의 개념이 부재함. 이는 이 법안이 국방인공지능 사업자와 국방부 등의 책무와 국방인공지능의 영향을 받는 사람들의 권리 구제에 대해 제대로 다루고 있지 않기 때문임. 국방인공지능이 인간의 존엄성과 기본권을 침해하지 않는 방식으로 개발·활용되도록 하기 위해서는, 관련 주체의 책임과 영향받는 자의 권리를 규정하기 위한 정의 규정이 필요함.

● 수정 요구 사항

- 제2조 정의에서 국방인공지능 규범 논의의 핵심 개념인 '자율성(autonomy)', '인간통제(human control)', 'AI기반 무기체계(AI-enabled weapon system)' 등에 대해 별도로 정의해야 함. 또한, 이 법안에서 강조하는 '안전성', '신뢰성', '책임성'은 무엇인지에 구체적인 설명이 있어야 함.
- 인공지능기본법을 준용하여 국방인공지능 사업자 및 이용자의 책무와 영향받는 자 권리 구제를 위해 국방인공지능사업자, 이용자, 영향받는자의 정의를 규정할 필요가 있음.⁵

3. 모호한 '인적 개입' 개념을 구체화하고 살상 결정 위임 불허 원칙도 명시

제3조(기본방향) ① 국방인공지능 정책은 국방인공지능기술을 개발하고 활용하여 국가 안보 및 국방력 강화를 최우선으로 고려하는 한편, 국방인공지능시스템의 안전성·신뢰성·책임성을 확보하여야 한다.

② 국방부장관, 병무청장, 방위사업청장(이하 "국방부장관등"이라 한다)은 다음 사항과 관련하여 책임 있는 국방인공지능 개발 및 활용을 추구하고, 기술 개발 및 활용 과정에서 발생할 수 있는 위험을 체계적으로 식별·분석·완화·관리하여야 한다.

1. 국방에 관련된 군정 및 군령과 그 밖에 군사에 관한 사항
2. 방위력개선사업, 군수물자의 조달 및 방위산업의 육성과 그 밖에 방위사업에 관한 사항

⁵ 이 역시 인공지능기본법이 기본 적용되고 특별한 규율이 필요한 부분만 본 법안에서 규율하는 방식을 취했다면, 별도의 정의규정을 두지 않고 인공지능기본법에 의존할 수 있었으나, 현재는 국방인공지능에 인공지능기본법이 적용되지 아니하므로, 본 법안에서 이 개념들을 정의할 필요가 있음.

3. 징집 및 소집과 그 밖의 병무행정에 관한 사항

③ 국방인공지능을 개발·활용함에 있어 자동화된 결정이 사람의 생명 또는 신체에 중대한 위해를 발생하게 할 우려가 있는 경우에는 인적 개입이 보장되도록 하여야 한다.

- 제3조에서는 국방인공지능 개발 및 이용의 기본적인 원칙을 명확하게 규정할 필요가 있으며, 유엔을 비롯한 국제 사회가 오랜 논의를 통해 형성해온 다음과 같은 원칙이 포함되어야 함.
- 첫째, 국제인도법(IHL) 및 국제인권법을 준수하여야 한다는 원칙임. 이 원칙은 <치명적 자율무기체계>에 대한 유엔 총회 결의안(A/RES/79/62) 등 대부분의 국제 규범에서 재확인하고 있음. 둘째, 무력 사용에 따른 결과와 법적, 도덕적 책임은 기계에 전가될 수 없으며, 인공지능 시스템의 전 수명 주기에 걸쳐 군사 지휘계통 내 인간이 책임을 져야 한다는 원칙임. 이를 위해 의미있는 인간의 통제 및 판단(Meaningful Human Control and Judgment)이 보장되어야 함. 셋째, 군사 분야 인공지능 시스템은 안전성과 신뢰성을 갖춰야 한다는 원칙임. 이를 위해 철저한 시험, 평가 메커니즘의 구축, 적절한 안전장치의 마련, 추적가능성과 설명가능성의 확보가 명시되어야 함.
- 제3조 제1항은 국방인공지능 정책이 “국가 안보 및 국방력 강화를 최우선으로 고려”해야 한다고 하면서 그 다음에야 안전성·신뢰성·책임성을 언급하고 있음. 법안의 중심 가치가 인간의 존엄과 기본권이 아니라 안보 효율성에 놓여 있다고 해석할 수밖에 없음. 이는 안전성·신뢰성·책임성이 충분히 확보되지 않은 국방인공지능의 도입을 정당화하는 근거로 작용할 우려가 있으므로, 이 문구를 삭제할 필요가 있음.
- 제3조(기본방향) 제3항은 “자동화된 결정이 사람의 생명 또는 신체에 중대한 위해를 발생하게 할 우려가 있는 경우에는 인적 개입이 보장되도록 하여야 한다”고 명시하고 있음. 그러나 ‘중대한 위해’가 무엇인지, ‘우려’에 대한 판단 기준과 주체는 누구인지, ‘인적 개입’의 주체, 시점, 방식은 무엇을 의미하는지에 대해 정의되어 있지 않음. 인간이 단순히 최종 승인권을 갖는 구조라면 이는 실질적 통제라기보다 형식적 승인에 그칠 가능성이 높음.
- 이미 전장 현장에서는 인공지능의 표적 추천을 인간이 거의 그대로 승인하는 자동화 편향(automation bias) 문제가 발생하고 있음. 우크라이나 전장에서 운용된 AI 기반 표적 데이터 융합 시스템은 알고리즘이 표적 후보를 대량으로 생성하고 위험 점수를

산출, 공격 우선 순위를 정렬하는데, 이 과정에서 인간의 판단 범위는 이미 알고리즘에 의해 구조화되기 때문에 다른 결정을 내리기 어려움. 이스라엘이 팔레스타인 가자지구 집단학살에 활용한 AI 기반 의사결정시스템 ‘라벤더(Lavender)’와 ‘가스펠(Gospel)’은 개인의 행동 패턴을 분석해 ‘위험 점수’를 매겨 표적으로 자동분류하는데, 이스라엘군은 실제 단 20초만에 타격 승인을 결정하고 있다고 알려짐.

- 스톡홀름국제평화연구소는 “공식적인 공격 승인은 인간 운용자에게 있지만, 잘 조정된 통제 구조가 없다면 표적 결정이 기능적으로 자동화되는 쪽으로 기울어질 수 있다”고 지적한 바 있음⁶. 인간이 최종 승인권을 갖고 있다 하더라도 알고리즘이 제시한 표적 리스트와 위험 점수가 이미 결정을 구조화하기 때문에, 인간은 사실상 기계의 결정을 승인하는 ‘고무도장(rubber stamp)’ 역할에 불과하다는 것임.
- 또한, 이 법안은 ‘인적 개입(intervention)’을 보장한다고 명시하며, 마치 인간이 개입하여 자동화된 결정의 폐해를 줄일 수 있는 것처럼 서술하고 있음. 그러나 이는 도리어 AI가 자율적으로 작동할 수 있는 범위를 더 넓게 허용할 여지를 줄 뿐임. 앞서 언급했듯이 인공지능이 추천한 표적을 인간이 거의 그대로 승인하고 있는 상황에서, 구체적 요건 없이 ‘인적 개입’이라고만 명시하는 것은 형식적인 승인을 반복하는 것일뿐임.
- 이에 국제사회에서는 단순한 ‘인적 개입’을 넘어 AI가 내리는 결정의 결과와 과정에 대해 인간이 실질적으로 책임지고 조정할 수 있는 능력을 확보하는 ‘의미있는 인간통제(Meaningful Human Control)’를 논의해왔음. 그리고 지금은 AI 체계와 임무 특성, 전장 환경, 민간인 위험도 등 맥락에 따라 인간 통제 수준을 결정하는 ‘맥락에 적합한 인간통제(CAHJC, Context-Appropriate Human Control)’를 중심으로 논의가 진행되고 있음. 인간 지휘관이 경험과 맥락에서 도출하는 통합적 판단을 AI가 대체할 수 없다는 인식론적 한계가 ‘맥락에 적합한 인간 통제’ 요건의 핵심 근거임.
- 한편, 제3항은 자동화된 결정에 있어서 인적 개입의 보장을 ‘사람의 생명 또는 신체에 중대한 위해를 발생하게 할 우려가 있는 경우’로 한정하고 있음. 그러나 국방인공지능은 비단 AI 기반 무기체계만 의미하지 않으며, 국방정보시스템 및 병무행정 등 다양한 영역에서 활용될 수 있음. 이 과정에서 인공지능이 사람의 분류, 선별, 예측, 평가에 개입할 경우 직접적 살상 여부와 무관하게 심각한 권리침해가

⁶ 스톡홀름국제평화연구소(SIPRI), <자율무기체계와 인공지능 기반 의사결정지원체계의 군사 표적 선정>, 2025.6 https://www.sipri.org/sites/default/files/2025-12/ko_laws_v.pdf

발생할 수 있음. 이에 따라 국방인공지능은 생명과 신체에 대한 권리 뿐만 아니라, 개인정보, 차별금지, 적법절차, 표현의 자유, 이동의 자유, 난민과 이주민의 권리, 병역행정에서의 평등권 등 다양한 기본권에 영향을 줄 수 있음. 따라서 국방인공지능의 개발 및 이용과 관련된 모든 생명 주기에서 책임성과 책무성을 가져야 하며, 국방인공지능의 자동화된 결정에 의해 영향받는 자의 권리가 침해될 우려가 있을 경우, 영향받는 자의 설명요구권과 인적개입을 요구할 권리가 보장되어야 함.

● 수정 요구 사항

- 제3조 기본방향에서 첫째, 국제인도법(IHL) 및 국제인권법을 준수하여야 한다는 원칙, 둘째 국방인공지능의 개발 및 이용의 전 주기에서 안전성 · 신뢰성 · 책임성을 확보하고, 영향받는 자의 설명요구권과 인적개입 요구권을 보장해야 한다는 원칙을 명문화 해야 함. 셋째 ‘인적개입’이라는 모호한 표현을 ‘맥락에 적합한 인간의 책임있는 판단과 통제’ 등으로 대체하고, 인간 통제의 구체적인 구성요소를 법률에 명문화해야 함. 또한, 인간에 대한 살상 결정을 기계에 위임하지 않는다는 원칙이 명시되어야 함.

4. 국방인공지능위원회의 독립성과 민주적 통제 제도화

제6조(국방인공지능위원회) ① 국방인공지능의 발전 및 국방력 강화와 국방인공지능사업의 진흥과 관련된 주요 정책 사항을 심의하고 의결하기 위하여 국방부장관 소속으로 국방인공지능위원회(이하 “위원회”라 함)를 둔다.

② 위원회는 위원장 1명과 부위원장 1명을 포함하여 25명 이내의 위원으로 구성한다.

③ 위원회의 위원장은 국방부장관이 되고, 부위원장은 국방부차관이 된다.

④ 위원회의 위원은 다음 각 호의 사람이 된다.

1. 대통령령으로 정하는 관계 중앙행정기관의 차관 또는 차관급 공무원
2. 병무청 차장
3. 방위사업청 차장
4. 합동참모본부 전략기획본부장, 육 · 해 · 공군 참모차장 및 해병대 부사령관
5. 국방부의 인공지능에 관한 업무를 담당하는 차관보
6. 한국국방연구원장, 국방과학연구소장, 국방기술품질원장, 국방기술진흥연구소장
7. 국방과학기술 및 인공지능 분야에 관한 전문지식과 경험이 풍부한 사람 중에서 방위사업청장의 의견을 들어 국방부장관이 위촉하는 사람

제7조(위원회의 기능) 위원회는 다음 각 호의 사항을 심의 · 의결한다.

1. 기본계획의 수립 및 변경에 관한 사항
2. 제3조에 따른 기본방향에 기반한 국방인공지능 관련 정책 및 제도의 수립 및 개선에 관한 사항
3. 국방 전략 및 작전 수행 개념에 부합하는 국방인공지능 연구개발 전략 수립에 관한

사항

4. 국방인공지능기술 개발 및 활용을 위한 전략 및 자원 확보 방안에 관한 사항
5. 국방인공지능기술 개발 및 국방 분야 적용에 저해되는 규제의 발굴 및 개선에 관한

사항

6. 민간 인공지능기술 도입 등 국방 분야 인공지능 활용 촉진에 관한 사항
7. 국방인공지능 관련 국제 협력에 관한 사항
8. 국방인공지능의 안전성 및 신뢰성 확보 등에 관한 사항
9. 국방인공지능 관련 군사적 및 사회적 변화 양상과 정책적 대응에 관한 사항
10. 이 법 또는 다른 법률에서 위원회의 심의사항으로 정한 사항
11. 그 밖에 위원회의 위원장이 필요하다고 인정하여 위원회의 회의에 부치는 사항

- 법안 제6조와 제7조는 국방인공지능 거버넌스 체계인 국방인공지능위원회의 구성과 기능에 대해 명시하고 있음. 국방인공지능위원회는 국방인공지능 관련 주요 정책 및 제도 개선 등을 심의·의결하는 기구인데 국방부 장관을 위원장으로 하고 위원은 군, 정부 중심 인사들로 구성하고 있음. 그러나 국방 분야에서의 인공지능 기술이 갖는 파급력과 위험성을 고려할 때 군, 정부 관계자 중심으로 위원회가 구성되면 논의가 국방인공지능 사업의 추진과 기술 도입 확대에만 초점을 맞출 수밖에 없음.
- 국방인공지능이 국제법 준수와 민간인 보호 문제와 직결된 만큼 다양한 관점에서 위험을 검토하고 통제할 수 있는 거버넌스 체계가 필요함. 시민단체 활동가, 국제법 전문가, 평화학 전문가 등 다양한 분야의 외부 전문가 참여를 보장하도록 위원회 구성을 개선해야 함. 또한 관계 중앙행정기관의 참여 범위에 개인정보보호위원회, 국가인권위원회 등을 포함할 필요가 있음.
- 한편, 제7조(위원회의 기능) 제8호는 국방인공지능의 안전성 및 신뢰성 확보를 위원회의 기능으로 규정하고 있지만, 동시에 제5호는 국방인공지능기술 개발 및 국방분야 적용에 저해되는 규제의 발굴 및 개선에 관한 사항을 위원회가 심의·의결하도록 규정하고 있음. 위원회 자체가 ‘국방인공지능의 발전 및 국방력 강화와 국방인공지능사업의 진흥’ 정책을 심의, 의결하는데 목적을 두고 있기에 사실상 위원회는 규제 완화를 중심으로 운영될 가능성이 높음. 더구나 위원회 의결 사항에 대한 공개나 국회 보고 의무도 명확히 규정되어 있지 않아 정책 결정 과정에 대한 민주적 통제가 이뤄지기 어려운 구조적 한계가 있음.
- 국제 규범 역시 군사 분야 인공지능 거버넌스에 있어 학계, 산업계, 시민사회 등 다양한 이해관계자의 참여 중요성을 강조하고 있음. 유엔 총회 결의안(A/RES/79/239)은 “국가들이 군사 분야의 인공지능 활용이 평화와 안보에

미치는 영향을 이해하고 해결할 수 있도록 지원하는 과정에서, 유엔 기구, 국제 및 지역기구 뿐만 아니라 국제적십자위원회(ICRC), 학계, 시민사회, 기술계 및 민간 부문의 기여를 인정하고, 다자간 이해관계자 접근법(multi-stakeholder approach)의 중요성을 강조”하고 있음. REAIM 글로벌 위원회도 5대 핵심 권고사항 중 하나로 “군사 분야 내 AI의 책임감 있는 통합을 위하여 영구적이고 포괄적이며 다중이해관계자 및 다자간 대화(a permanent, inclusive, multi-stakeholder, and multilateral dialogue) 체계를 구축할 것”을 제시하고 있음.

- 수정 요구 사항

- 위원회 구성은 군, 정부 중심 구조에서 벗어나 시민사회와 외부 전문가, 관계 중앙행정기관 등의 참여를 확대하고, 외부 위원이 전체 위원의 3분의 1 이상이 되도록 함.
- 위원회 기능은 단순한 정책 심의 수준을 넘어 국방인공지능의 위험 관리와 책임성을 강화할 수 있는 방향으로 강화해야 함. 특히 제7조 제8호의 ‘국방인공지능의 안전성 및 신뢰성 확보 등에 관한 사항’은 단순한 심의, 의결로 충분하지 않으며, 국방인공지능의 ‘반인도적 사용 가능성’이나 국제법 위반 가능성을 예방하기 위한 검토와 관리 기능을 명시적으로 포함해야 함.
- ‘인공지능 기술의 국방 분야 개입·도입 및 운용 현황’과 위원회 활동이 정기적으로 국회에 보고될 수 있도록 국회 보고를 의무화하고, 국회의 요구가 있을 경우 관련 정책과 사업 현황을 보고하도록 규정하여 민주적 통제가 가능하도록 해야 함.
- 국방인공지능의 위험성과 책임 문제를 고려할 때, 위원회에 인공지능기술의 반인도적 적용이나 국제법 위반 가능성에 대해 조사 권한을 부여하고, 위원회와는 별도로 국방인공지능의 개발 및 운용과정에 대한 독립적인 감독 기능을 수행하는 감독관 제도를 도입해야 함.

5. 개발 및 특례 조항의 제한적 적용 및 군사AI 금지·제한하는 규정 추가

제14조(인공지능기술의 적용) ① 국방부장관등은 「방위사업법」 제3조제3호 및 제4호에 따른 무기체계 및 전력지원체계와 「국방정보화 기반조성 및 국방정보자원관리에 관한 법률」 제2조제7호에 따른 국방정보화사업(이하 “무기체계등”이라 함)에 인공지능기술을 적용하여 도입 및 운용하는 경우로서 신속한 전력화가 필요하거나 기술적 혁신성이 인정되는 등 대통령령으로 정하는 요건을 갖춘 경우에는 국방인공지능기술 분야의 특성을 고려하여 무기체계등의 획득방법 및 절차를 별도로 정할 수 있다.

② 제1항에 따른 무기체계의 도입 및 운용을 위한 세부 기준과 절차, 그 밖에 필요한 사항은 국방부령으로 정한다.

제16조(국방인공지능 학습데이터의 관리 및 보안 등) ① 국방부장관등은 국방인공지능의 성능 향상 및 신뢰성 확보를 위하여 필요한 학습데이터의 생산·수집·관리 및 공유를 촉진하기 위한 시책을 수립하여야 한다.

② 국방부장관등은 학습데이터의 효율적인 관리, 품질 확보 등의 업무를 전문적으로 수행하기 위하여 대통령령으로 정하는 바에 따라 전담기관을 지정할 수 있다.

③ 제1항에 따른 시책 수립 및 제2항에 따른 전담기관의 지정, 그 밖에 필요한 사항은 대통령령으로 정한다.

제17조(민간 인공지능기술의 도입 및 활용) ① 국방부장관등은 무기체계등에 적용 가능한 민간 인공지능기술을 적극적으로 발굴하고 도입하여야 한다.

② 국방부장관등은 무기체계등의 소요 결정 시 민간 인공지능기술의 적용 여부가 체계적으로 검토·반영될 수 있도록 합동참모의장과 협의하여 관련 기준 및 절차를 마련하여야 한다.

제18조(국방인공지능의 활용 촉진) ① 국방부장관등은 국방인공지능의 활용을 전 국방 분야로 확산하고 관련 산업 및 기술과의 연계를 강화하기 위하여 필요한 시책을 수립하여 추진하여야 한다.

② 국방부장관등은 국방인공지능의 개발을 지원하기 위하여 필요한 경우에는 「국방과학기술혁신 촉진법」에 따른 국방연구개발사업에 국방인공지능 관련 연구개발과제를 우선적으로 반영하여 추진할 수 있다.

③ 국방부장관등은 제2항에 따라 개발된 국방인공지능의 신속한 전력화 및 군사적 활용을 지원하기 위하여 필요한 경우 관련 규제의 전부 또는 일부를 적용하지 아니하고 실증할 수 있도록 하는 규제특례 및 관련 절차가 간소화될 수 있도록 하는 방안이 시행될 수 있도록 적극 지원하여야 한다.

④ 국방부장관등은 국방인공지능의 개발 및 활용 촉진을 위하여 필요한 경우 민간 기업, 대학, 연구기관 등과의 협력을 강화하고 기술 교류 및 공동 연구개발을 지원할 수 있다.

- 법안 제3장은 국방인공지능기술의 개발 및 특례를 명시하고 있는데 ‘신속전력화’, ‘기술혁신성’ 명분으로 규제나 절차를 우회할 수 있는 방법을 명문화함으로써 안전성 확보나 인간 통제, 책임 구조를 법적으로 보장하지 못하는 심각한 문제가 있음.
이러한 ‘개발 및 특례조항’은 「인공지능기본법」에도 없는 규정임.

- 제14조(인공지능기술의 적용) 제1항은 “신속한 전력화가 필요하거나 기술적 혁신성이 인정되는 등 대통령령으로 정하는 요건을 갖춘 경우... 무기체계등의 획득방법 및 절차를 별도로 정할 수 있다”고 명시하고 있는데, ‘신속한 전력화 필요’나 ‘기술적 혁신성 인정’ 등 기준 자체도 매우 추상적이고 모호할 뿐 아니라 누가, 어떤 기준으로 판단하는지 명확하지 않음. 게다가 무기체계등의 획득방법 및 절차를 별도로 정하게 함으로써 안전성 검증이나 책임성 등에 대한 검증을 생략하는 등 자의적으로 절차를 간소화할 우려가 있음. 국방인공지능기술 적용에서 국제법 준수, 인간 통제, 제네바협약 제1추가의정서 제36항에 따른 무기 검토 의무를 포함한 구체적 요건과 기준을 마련하는 것이 필요함.
- 제16조(국방인공지능 학습데이터의 관리 및 보안 등)는 학습데이터의 생산·수집·관리 및 공유를 촉진하기 위한 시책 수립을 규정하고 있음. 그러나 인공지능 학습데이터는 시스템의 성능과 한계를 결정짓는 핵심 요소이며, 특히 국방인공지능의 경우 그 영향은 국가 안보뿐 아니라 시민의 생명과 기본권에까지 미칠 수 있음. 그럼에도 불구하고 본 법안은 “학습데이터의 생산·수집·관리 및 공유를 촉진하기 위한 시책을 수립”하도록 하고 있을 뿐, 학습데이터의 관리와 관련하여 구체적인 요건이나 기준을 제시하지 않고 있음. 즉, 개인정보 및 민감정보 처리의 한계, 데이터 출처의 적법성, 민간 데이터의 군사적 활용 기준, 데이터의 정확성 및 대표성 확보, 데이터 편향 점검, 독립적 감독 체계 등 필수적인 요소들이 규정되어 있지 않음. 국방 분야의 특성상 데이터가 비공개적으로 관리되는 경우가 많다는 점을 고려할 때, 법률 수준에서 최소한의 데이터 보호 및 남용 방지 요건을 명확히 규정할 필요가 있음.
- 제17조(민간인공지능기술의 도입 및 활용) 제1항은 “국방부 장관은 무기체계 등에 적용 가능한 민간 인공지능기술을 적극적으로 발굴하고 도입하여야 한다”고 명시하고 있음. 그러나 민간 인공지능 기술은 본래 군사 목적이 아닌 상업용으로 개발된 것이므로, 이를 군사적 목적으로 적극 도입하게 될 경우 생성형 AI, 안면인식 기술, 자율주행 기술 등도 쉽게 군사적 감시나 타격 시스템으로 전환될 수 있는 위험이 존재함. 즉, 이중용도(dual-use) 기술의 악용 가능성이 높음. 따라서 민간의 인공지능을 국방 목적으로 도입할 때 어떠한 원칙과 절차가 필요한지 엄격하게 규정될 필요가 있음. 그러나 본 법안은 이에 대해 적절한 규정을 두고 있지 않으며, 단지 민간 인공지능기술을 적극적으로 적용하기 위한 목적의 기준 및 절차 마련만을 규정하고 있을 뿐임. 민간의 인공지능을 도입할 때에는 인공지능기본법의 고성능

인공지능(제32조)과 고영향 인공지능(제34조)에서 규정하고 있는 것 이상의 안전성, 신뢰성, 책무성 의무를 사업자에게 부과할 필요가 있음.

- 제18조(국방인공지능의 활용 촉진) 제3항 역시 “국방인공지능의 신속한 전력화 및 군사적 활용을 지원하기 위하여 필요한 경우 관련 규제의 전부 또는 일부를 적용하지 아니하고...”라고 명시하고 있는데, 이는 ‘신속한 전력화’, ‘군사적 활용’을 명분으로 사실상 규제 회피를 법으로 명문화한 것이기에 이 또한 심각한 문제가 있음. 더구나 ‘필요한 경우’라고 표현하여 규제 회피의 요건을 포괄적으로 규정함으로써 국방부장관 등이 자의적으로 권한을 남용할 우려가 있음.

- 수정 요구 사항

- 제3장 ‘국방인공지능기술의 개발 및 특례’는 ‘국방인공지능기술의 개발, 활용 및 규제’로 수정하고, 제14조, 제18조에서 고위험영역 인공지능 관련 사업자의 책무, 표적 선택, 교전 개시, 치명적 무력 행사와 같은 핵심 영역에서는 안전 검증을 우회할 수 없도록 특례 조항의 적용 범위를 제한해야 함.
- 제17조에서 민간 인공지능 기술을 도입하는 경우 사전에 해당 기술의 안전성, 편향성, 이중용도 위험 및 국제법 준수 여부 등 평가를 통해 도입 여부를 결정하도록 하고, 민간인공지능 기술의 군사적 활용과 관련하여 이중용도 기술 관리체계를 마련하고, 도입, 승인, 사용, 모니터링 등 단계별 관리 절차를 시행하는 것을 명시해야 함. 또한, 인공지능기본법의 고성능 인공지능(제32조)과 고영향 인공지능(제34조)에서 규정하고 있는 것 이상의 안전성, 신뢰성, 책무성 의무를 사업자에게 부과할 필요가 있음.

6. 안전성 확보 의무조항 강화 및 피해 책임 범위와 기준 명확화

제23조(국방인공지능 신뢰 기반 조성을 위한 시책의 마련) ① 국방부장관등은 신뢰할 수 있는 국방인공지능의 개발 및 활용을 위하여 다음 각 호의 시책을 마련하여야 한다.

1. 안전하고 신뢰할 수 있는 견고한 국방인공지능 개발 및 운용 환경 조성
2. 국방인공지능의 활용이 국가 안보 및 군 작전에 미치는 영향 등에 관한 전망 및 관련 법령·제도의 정비
3. 국방인공지능의 안전성 및 신뢰성 확보를 위한 기술 개발 및 적용 지원
4. 국방인공지능의 윤리적 활용 및 책임성 확보를 위한 교육 및 훈련
5. 국방인공지능 관련 기관, 연구기관, 산업계 등의 안전성 및 신뢰성 관련 자율적인 규약 제정 및 시행 지원
6. 국방인공지능 관련 단체 등의 안전성 및 신뢰성 증진을 위한 자율적인 협력, 윤리지침 제정

등 민간 활동의 지원 및 확산

7. 그 밖에 국방인공지능의 안전성 및 신뢰성 확보를 위하여 대통령령으로 정하는 사항
② 국방부장관등은 제1항 각 호의 시책을 수립·추진함에 있어 필요한 경우 관련 기관 및 전문가의 의견을 수렴할 수 있다.

제24조(국방인공지능시스템의 안전성 확보 의무) ① 국방부장관등은 안전하고 신뢰할 수 있는 국방인공지능시스템 개발 및 활용을 위한 시책을 추진하여야 한다.

② 국방부장관등은 국방인공지능시스템을 개발하거나 도입하는 기관 또는 사업자에 대하여 그 안전성을 확보하도록 하거나 필요한 지원을 할 수 있다. 특히, 전력 운용 간 안전에 중대한 영향을 미치는 국방인공지능시스템에 대해서는 안전성 확보를 위하여 다음 각 호에 해당하는 방안을 마련하도록 할 수 있다.

1. 국방인공지능 수명주기 전반에 걸친 위험의 식별·평가 및 완화
2. 국방인공지능 관련 안전사고에 대한 모니터링 및 위험관리체계 구축

③ 안전연구소는 제1항 및 제2항에 따른 시책 추진을 지원하여야 하며, 안전성 확보 대상의 범위는 대통령령으로 정한다.

제25조(국방인공지능기술 개발 및 안전한 이용 지원) ① 국방부장관등은 국방인공지능 발전을 위하여 「국방과학기술혁신 촉진법」에 따른 국방연구개발사업이 원활하게 추진되도록 필요한 시책을 추진하여야 한다.

② 국방부장관등은 국방인공지능의 안전하고 신뢰할 수 있는 이용 환경을 조성하고, 데이터 편향 문제 및 오용·남용 방지 방안을 마련하기 위하여 필요한 경우 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」 제29조에 따른 시책을 준용할 수 있다.

③ 국방부장관등은 국방인공지능기술 개발 및 활용과 관련하여 필요한 경우 국내외 연구기관, 대학, 기업 등과 협력하여 국방인공지능 관련 연구개발을 장려하고 지원할 수 있다.

- 국방인공지능법안 제1조는 ‘안전성·신뢰성 및 책임성 확보를 위한 관리 체계를 정립한다’고 명시하고 있으나, 법안의 구체적인 조항에서는 이를 실질적으로 구현하기 위한 규정은 부재함. 이 법안의 제4장은 ‘국방인공지능 안정성 확보’를 다루고 있으나, 안정성, 신뢰성 확보를 위한 법적 구속력이 있는 의무 사항, 인공지능 시스템의 오작동이나 운용 과정에서 발생할 수 있는 피해에 대해 지휘관, 개발자, 제조사 등 관련 주체 간 책임 구조, 사고 조사 절차, 기록 보존 의무 등이 명확히 규정되어 있지 않음. 특히, 현재 조항들은 대부분 “자율적인 규약”이나 “안전성을 확보하도록 하거나 지원할 수 있음” 정도에 그쳐, 사업자가 이를 지키지 않더라도 법적 책임이 발생하지 않는 구조적 한계를 가지고 있음.
- 특히 제24조(안전성 확보 의무) 제2항은 “국방부장관등은 국방인공지능시스템을 개발하거나 도입하는 기관 또는 사업자에 대하여 그 안전성을 확보하도록 하거나 필요한 지원을 할 수 있다”고 명시하고 있으나, 이는 의무 규정이 아니기에 사업자나

기관이 안전성 확보를 회피할 가능성을 있음. 위험도가 높은 시스템조차 안전 조치를 선택적으로 적용할 수 있도록 열어두고 있는 것임.

- 국방인공지능은 그 특성상 고위험 인공지능에 해당함에도 불구하고, 이 법안은 국방인공지능시스템을 제공하는 사업자의 책무에 대해서는 규율하지 않고 있음. 예를 들어, 위험관리방안의 수립, 인공지능의 학습데이터와 최종결과와 관련된 투명성 요건, 국방인공지능시스템을 이용하는 이용자에 대한 보호 방안, 사람의 관리·감독, 향후에 문제를 추적하고 책임 소재를 파악할 수 있도록 하는 문서의 작성과 보관 등은 고위험 인공지능의 필수적인 요건이나 본 법안은 이러한 사항을 명시적으로 규정하고 있지 않음. 이는 국방인공지능이 야기할 수 있는 위험을 사전에 관리하거나, 문제 발생 시 원인을 규명하고 책임을 부과할 수 있는 기반이 부재함을 의미함.
- 아울러 국방인공지능은 무기체계뿐 아니라 감시, 분류, 평가 등의 목적으로 활용될 수 있으며, 이 과정에서 차별적 결과를 초래하거나 기본권을 침해할 위험이 존재함. 이에 따라 사전적 위험 평가가 필수적이며, 특히 인공지능기본법에서 ‘노력 의무’ 수준으로 규정되어 있는 인권영향평가는 국방인공지능의 경우 의무적으로 수행되도록 할 필요가 있음.
- 국방 분야에서의 인공지능 활용에서 국제사회가 지속적으로 지적해 온 핵심 문제 중 하나는 ‘책임 공백(accountability gap)’임. AI 기반 시스템이 오류를 범하거나 편향된 판단을 내렸을 경우, 그 결과에 대한 책임이 개발자, 지휘관, 제조사 중 누구에게도 명확히 귀속되지 않는 문제가 있음. 개발자는 모든 상황을 예측할 수 없었다고 주장하고, 지휘관은 기계의 판단에 따른 결과라고 책임을 부인하며, 제조사는 기술적 복잡성을 이유로 책임을 회피할 가능성이 있음. 이러한 상황은 인공지능 기술의 복잡성과 자동화된 의사결정 구조가 결합되며 책임 주체를 특정하기 어려운 구조적 문제를 초래함. 이러한 책임 공백은 피해자 권리 구제와 사법적 대응을 어렵게 만들 수 있음.

● 수정 요구 사항

- 안전성 확보 의무 조항은 현재와 같이 ‘할 수 있음’ 수준이 아니라 ‘해야 한다’로 법적 강제력이 있는 의무 사항으로 규정하고, 안전 인증이 없는 시스템 운용은 금지해야 함.

- 인공지능기본법 제34조에서 규정하고 있는 고영향 인공지능 사업자의 책무를 기본적으로 국방인공지능 사업자에게 적용해야 함.
- 국방부 및 산하기관은 인공지능 시스템을 도입할 경우 의무적으로 인권영향평가를 수행하도록 규정해야 함.

7. 특정 국방인공지능에 대한 금지 규정 필요

- 무엇보다 가장 큰 문제는 이 법안은 국방인공지능의 ‘기술 적용’, ‘도입 및 활용’, ‘활용 촉진’만을 규정하고 있는 것임. 한국 사회가 국방인공지능을 어디까지 허용하고 무엇을 금지·제한할지에 대한 기준을 마련하는 것은 매우 중요함. 미국과 이스라엘의 이란 공격에서 볼 수 있듯이 국방분야에서의 인공지능 기술 활용은 인명 피해에 직접적인 영향을 미치게 됨. 이에 국방인공지능법안에는 반드시 위험 수준에 따라 금지하거나 제한해야 할 인공지능에 대한 규정이 포함되어야 함. 금지 대상에 대한 규정이 부재할 경우 인간의 개입 없이 살상을 결정하는 자율무기체계(LAWS) 개발이 허용되는 것으로 해석될 여지가 있으며, AI기반 의사 결정 지원 시스템이나 AI 표적 선정 시스템 역시 무제한으로 개발·도입·운용될 가능성이 있음.
- 현재 전장에서는 무기체계 자체보다 AI 기반 의사 결정 지원 시스템(AI-DSS, Artificial Intelligence-based Decision Support System)이나 AI 표적 선정시스템(AI-based Targeting System)의 위험이 더 큼. 이스라엘이 팔레스타인 집단학살 전장에서 활용한 AI 기반 의사결정 지원 시스템은 표적 식별에서 승인까지만 20초가 걸렸다고 함. 개인의 행동 패턴을 분석해 ‘위험 점수’를 매겨 자동으로 표적으로 분류한 시스템은 잠재적 표적으로 3만 7천 명을 생성했는데, 전투원 1명을 제거하기 위해 민간인 15~20명 희생을 허용하는 등 심각한 오류를 일으켜 막대한 민간인 피해를 야기함. 제대로 된 규제와 통제 없이 이러한 시스템을 개발하여 배치할 경우 피해는 걷잡을 수 없이 커질 수밖에 없음.
- 2024년 유엔 사무총장 보고서에 따르면, 한국을 비롯해 129개 이상의 국가가 자율무기 시스템에 대해 법적 구속력이 있는 국제규범(legally binding instrument)의 필요성을 지지⁷했으며, 2024년 12월 유엔 총회는 자율무기체계 문제를 다루는 결의안(A/RES/79/62)을 166개국 찬성으로 채택한 바 있음⁸. 한국 정부는 REAIM과

⁷ UN, <Lethal autonomous weapons systems Report of the Secretary-General(A/79/88)>, 2024.7.1
<https://docs.un.org/en/A/79/88>

⁸ UN, <Resolution adopted by the General Assembly on 2 December 2024(A/RES/79/62)>, 2024.12.10,
<https://docs.un.org/en/a/res/79/62>

유엔 총회 결의안 채택을 주도하며 이러한 흐름의 선두에 서있음. 이처럼 금지하거나 제한하는 인공지능이 규정이 국내법에 포함되지 않는다면, 이는 국제규범 형성 과정에서 한국의 목소리와 반하는 것임.

- 수정 요구 사항

- 금지대상, 고위험, 제한된 위험 등 위험도(risk-based)에 따른 분류체계를 도입하여야 함. 특히 의미있는 인간통제 없이 작동하면서 사람을 표적으로 삼는 자율무기의 연구, 개발 및 운용을 금지해야 함. 금지되지 않는 자율 무기에 대해서는 의미있는 인간통제가 어떻게 이루어질 수 있는지, 즉 그에 대한 의사결정 요건, 기술적 요건, 운영적 요건에 대해서 가능한 한 구체적으로 제한하는 규정을 두어야 함.

8. 영향받는 자의 권리구제와 책임자 처벌

- 본 법안은 국방인공지능으로 인해 영향받는 자에 대한 아무런 규정이나 권리를 포함하고 있지 않음. 그러나 국방인공지능 역시 다양한 ‘영향받는 자’를 발생시키며, 그 피해는 일반 인공지능보다 훨씬 중대할 수 있음. 여기서 영향받는 자는 인공지능 무기체제에 의한 대상이 되는 상대국가의 군인만이 아니라, 상대국가의 민간인, 자국의 군인, 민간인, 또는 제3국의 민간인이 될 수도 있음. 국방인공지능이라는 이유만으로 부당하게 권리를 침해당한 사람들의 권리가 배제되어서는 안될 것이며, 문제가 발생할 경우 국방인공지능 사업자, 지휘관, 국가 등 관련 주체 중 누군가는 책임을 부담할 수 있어야 함.

- 수정 요구 사항

- 피해 발생 시 권리구제 절차를 법률에 명확히 규정할 필요가 있음. 구체적으로는 피해 신고 절차, 조사 절차, 책임 주체 및 책임 구조, 시정조치 절차, 조사 결과의 통지, 피해 보상 방식 등이 포함되어야 함. 인공지능기본법 제3조에서 규정하고 있는 영향받는 자의 설명요구권(인공지능의 최종결과 도출에 활용된 주요 기준 및 원리 등에 대하여 기술적·합리적으로 가능한 범위에서 명확하고 의미 있는 설명을 제공받을 권리)도 보장받아야 함.
- 인공지능기본법에서 과기정통부 장관은 인공지능사업자에 대한 사실조사를 하고 시정명령을 내릴 수 있음. 그런데 국방인공지능에 대해 국방부에서 이러한 역할을 담당할 경우 자칫 이해충돌이 발생할 수 있음. 따라서 국방인공지능에 대해 위험성을

평가하고, 사고를 조사하고, 시정조치를 권고할 수 있는 독립적인 감독기구를 두어야 함.

- 인공지능 오작동 또는 운용 과정에서 발생한 피해에 대해 국가, 기업, 개인의 책임 범위와 기준을 명확하게 규정해야 하며, 책임 규명과 사후 조사가 가능하도록 알고리즘 설계, 데이터 사용 및 운용 과정에 대한 기록 보존과 문서화 의무를 명시해야 함.

- 현재 국회에 발의된 <국방인공지능법안>은 국방인공지능에 대한 위험성이나 법적·윤리적 쟁점 등에 대한 충분한 고려 없이 국방인공지능 개발, 도입, 운용에 초점을 맞춘 국방인공지능산업 육성법임.
- 그러나 아무런 규제나 통제 없이 실제 전장에서 사용되고 있는 국방인공지능은 단순한 보조 도구가 아님. 정보 수집과 분석, 표적 식별, 판단까지 전 과정을 아우르며 전쟁을 주도하고 있음. AI 시스템의 정밀성과 빠른 속도는 흔히 ‘효율성’으로 표현되지만 데이터 오류로 인한 피해 발생 시 그 책임은 누구에게 물어야 하는지 명확하지 않음.
- 이에 국방인공지능 개발, 도입, 운용에 있어 무엇보다 중요한 것은 1) 국제인도법, 국제인권법 등 국제법 준수, 2) 민간인 보호, 인간에 대한 살상 결정은 기계에 위임하지 않는다는 것을 국방인공지능 기술의 도입과 활용에 대한 기본 원칙을 수립하는 것임. 우리 사회가 국방인공지능을 어디까지 허용하고 금지·제한할지에 대한 기준을 마련하고 법적 구속력이 있는 규제와 통제 방안을 마련하는 것도 반드시 수반되어야 하는 일임.
- 규제와 통제는 국방인공지능 개발과 활용에서 ‘걸림돌’이 아님. 이는 우리가 반드시 지켜야 할 최소선이자 넘지 말아야 할 레드라인임. 국방인공지능에 대한 법적 구속력이 있는 규제와 통제 방안을 마련하지 않는다면 멀지 않은 시점에 인류는 인간이 통제할 수 없는 인공지능에 위협당하는 일이 발생할 것임.
- 국회는 입법 공청회를 통해 국방인공지능에 대한 우려와 법적·윤리적 쟁점을 충분히 숙의 토론하여 법안을 전면 수정해야 할 것임.