살상의 자동화, 책임의 공동화(空洞化)

: 공동화(空洞化)에서 공동화(共同化)로

문아영(피스모모 대표)

1. 들어가며

2026년 2월 28일, 미국과 이스라엘의 이란 공습이 시작된 날 이란 남부 미나브의 샤자레 타예베 초등학교가 폭격당했다. 유엔 이란 독립국제진상조사단은 어린이 123명을 포함해 157명이 사망한 것으로 파악했다. 조사단은 민간 시설에 대한 공격이 전쟁범죄에 해당한다고 볼 합리적 근거가 있다고 판단했다.[1)]

샤자레 테예베 초등학교 폭격은 누구의 책임인가? 이 책임에 대해 거슬러 가기 위해서는 갱신되지 않은 표적 정보를 관리한 담당자, 다른 데이터베이스에 남겨진 시설 변경 기록을 연결하지 못한 조직, 표적을 검토·승인한 군인, 민간인 피해 완화 기능을 축소한 지휘부, 소프트웨어 공급 기업, 군사작전과 전쟁 개시를 결정한 정치 지도자들의 책임을 각각 살펴보아야 한다.

블룸버그의 보도에 따르면 이 학교는 과거 군사시설로 분류된 뒤 민간시설로 전환되었으나 주표적 데이터베이스가 제때 수정되지 않았다. 2019년 분석관이 시설의 변화를 별도 체계에 기록했지만 관련 정보가 연결되지 않았다. 몇몇 관계자들은 표적 목록화에 활용된 AI 기반 메이븐 시스템이 이 오래된 정보를 필터링해줄 수 있을 것으로 기대했다고 한다.[2)]

수업 중이던 토요일 오전의 학교에 토마호크 미사일이 떨어질 때까지 누가 어떤 정보를 알고 검증할 시간이 있었는지, 학교와 학교에서 공부하던 학생과 교사들이 보호받지 못한 것은 무엇 때문인지 아직 밝혀져야 할 것들이 많다.

샤자레 테예베 초등학교 공습 사건의 핵심은 AI의 개입 여부가 아니다. 조사결과를 더 기다려봐야겠지만, 현재까지로는 AI가 아니라 사람이 했어야 하는 역할을 하지 못해 발생한 사건으로 보는 편이 합리적이다. 폴리티코의 보도에 따르면 미 중부사령부의 민간인 피해 완화 전담인력은 10명에서 1명으로 줄었고, 해당 전담조직은 공격 전 이 표적을 검토하지 않은 것으로 알려졌다. 그러나 아직까지 미 국방부 내부 조사의 세부 내용은 공개되지 않았다. 그렇다면 문제는 AI 이전에 사람과 법규범, 승인 절차가 존재하는데도 설명·검증·책임 추궁의 조건이 어째서 작동하지 않는가에 있다.

---

1) UN Human Rights Council. A/HRC/63/61, Independent Fact-Finding Mission on Iran (2026), 118–126.
2) Bloomberg. 'Inside US Military 'Kill Chain' That Destroyed an Iranian School' (2026).

2. 책임은 남아있지만, 책무성은 비워진다?

H. L. A. 하트는 책임을 역할, 인과, 법적 책임, 책임 능력으로 구분했다. 기술과 정보체계는 결과의 인과적 연쇄에 개입할 수 있지만, 각 개인에게 법적 책임이 성립하는지는 그의 역할·인식·통제 가능성과 구체적 법률요건을 따로 검토해야 한다. 시간이 부족했다는 이유만으로 운용자의 일반적 '책임 능력'이 사라지는 것도 아니다. 핵심은 합리적 판단을 수행할 정보와 기회가 제도적으로 보장되었는가에 있기 때문이다.[3)]

마르크 보벤스는 책무성(accountability)을 행위자와, 그에게 설명을 요구하고 판단할 수 있는 장(forum) 사이의 관계로 설명한다. 행위자는 자신의 행위를 설명하고 정당화할 의무를 지며, 그 설명을 요구하는 주체는 질문하고 판단할 수 있어야 한다. 행위자는 그 판단에 따르는 결과를 감당해야 한다. 여기서 책임을 묻는 장은 법원과 의회, 감사·조사기구 같은 공식 제도일 수도 있고, 공적 설명과 응답을 요구하는 언론과 시민사회일 수도 있다. 물론 이들이 행사하는 권한과 판단의 효과는 서로 다르다. 따라서 책임(responsibility)의 귀속을 선언하는 것과 책무성이 실제로 작동하는 것은 별개의 문제다. 핵심 기록에 접근할 수 없고 외부 검증과 후속 조치가 막혀 있다면, 책임의 주체가 분명하더라도 그 책임을 실질적으로 묻기는 어렵다.[4)]

자율살상무기체계를 둘러싼 '책임 공백(responsibility gap)' 논의는 학습 기계의 예측과 통제가 어려워질 때 누군가에게 책임을 정당하게 귀속할 수 있는지를 문제 삼아 왔다. 그러나 공적 책무성의 약화는 기계의 '블랙박스'만으로 설명되지 않는다. 미나브의 학교 폭격을 둘러싼 논의에서 AI의 관여 여부에만 시선이 집중될수록, 표적 데이터의 관리와 민간인 피해 예방, 공격 승인과 사후 조사에 관한 인간 조직의 선택은 오히려 시야에서 밀려날 수 있다. AI가 의사결정 과정에 개입했다는 사실이 왜 그 과정에 참여한 인간의 책임을 추적할 수 없다는 뜻이어야 하는가?

필자는 그간 '살상의 자동화, 책임의 익명화'라는 제목으로 글을 쓰고 강의를 진행한 적이 있다. 그러나 최근에는 누가 결정했는지를 확인하기 어려워지는 '익명화'만으로 이 문제를 설명하기 어렵다고 느낀다. 실제로 지휘관과 계약 기업, 개발 부서와 승인권자의 이름은 문서에 남아 있을 수 있다. 문제는 그 이름을 확인한 뒤에도 무엇을 물을 수 있으며 그 답변을 누가 검증하고 어떤 결과로 연결할 수 있는가이다. 책임 공백이 언제나 책무성 약화에서 생겨나는 것은 아니다. 하지만 정치적·조직적 선택이 설명과 검증의 조건을 계속 비워 낸다면, 본래 물을 수 있었던 책임마저 추궁하기 어려운 상태로 만들어질 수 있다.

산토니 데 시오와 메카치는 인공지능이 초래하는 책임의 문제를 단순히 '누가 잘못했는가'라는 질문만으로 다룰 수 없다고 보았다. 그들이 구분한 여러 유형의 책임 공백에는 '공적 책무성의 공백(public accountability gap)'도 포함된다.[5)] 가령 어느 행정기관이 AI 시스템을 활용해 복지 지원 대상자를 선정했다고 가정해 보자. 지원에서 탈락한 시민이 이유를 묻자 담당 공무원은 시스템이 산출한 결과라고 답한다. 개발 기업은 행정기관이 제공한 자료를 바탕으로 작동했을 뿐이라고 하고, 행정기관은 알고리즘의 구체적인 작동 방식은 기업의 기술 영역이라고 맞선다. 시민은 불리한 결정이 내려졌다는 사실은 알지만, 어떤 정보와 기준이 사용되었는지는 확인할

3) Hart, H. L. A. Punishment and Responsibility. Oxford University Press (1968).
4) Bovens, M. 'Analysing and Assessing Accountability.' European Law Journal 13(4) (2007).
5) Santoni de Sio, F., ' Mecacci, G. 'Four Responsibility Gaps with Artificial Intelligence.' Philosophy ' Technology 34 (2021).

길이 없다. 문제는 잘못된 결정을 내린 사람을 찾기 어렵다는 데 그치지 않는다. 시민이 공적 결정의 근거를 요구하고 정당성을 다툴 권리 자체가 실질적으로 제약된다는 데 있다.

그런데 이러한 문제는 인공지능이 등장하기 훨씬 전부터 존재했다. 정치철학자 데니스 톰슨은 이를 '많은 손의 문제(problem of many hands)'라고 불렀다. 하나의 정책이 여러 부서와 담당자의 손을 거쳐 실행될 때, 각자가 맡은 역할은 분명하지만 결과 전체에 대한 책임 소재는 불분명해지는 현상이다. 예컨대 어떤 공공사업에서 심각한 인명 피해가 발생했다고 해 보자. 예산 담당자는 비용을 승인했을 뿐이라 하고, 설계 담당자는 주어진 기준에 맞춰 설계했다고 하며, 현장 책임자는 승인된 계획대로 집행했다고 답한다. 각자의 설명은 사실일지 모른다. 그러나 이 설명들을 모두 모아 놓아도 왜 그러한 결정이 내려졌고, 누가 위험을 방치했으며, 피해에 대해 누가 책임져야 하는지는 여전히 밝혀지지 않는다.

지그문트 바우만은 이러한 현상을 근대 관료제의 작동 방식과 연결해 분석하기도 했다. 거대한 조직에서 업무가 세분화될수록 사람들은 자신이 수행하는 일의 최종 결과로부터 멀어지기 쉽다. 누군가는 정보를 수집하고, 누군가는 분류하며, 누군가는 결재하고, 또 다른 누군가는 명령을 집행한다. 이 과정에서 사람들은 자신의 업무를 정확하게 처리했는지는 확인하지만, 그 업무가 타인의 삶에 어떤 결과를 가져오는지는 깊이 성찰하지 않게 된다. '맡은 일을 제대로 했는가'라는 질문이 '그 일을 하는 것이 과연 옳았는가'라는 질문을 밀어내는 셈이다. 한나 아렌트가 관료제를 '아무도 아닌 자의 지배(rule by Nobody)'라고 비판한 맥락도 이와 맞닿아 있다. 수많은 사람이 조직을 움직이고 명령과 규칙도 존재하지만, 정작 그 결정의 정당성을 따져 물으려 하면 책임 있게 응답하는 정치적 주체를 만나기 어렵다.

AI 군사화는 이 오래된 문제에 새로운 조건, '속도'를 더한다. 과거에는 여러 사람과 부서를 거치던 정보 분석과 표적 선정의 일부가 이제는 알고리즘을 통해 훨씬 빠른 속도로 처리된다. 인간이 여전히 최종 승인권자로 남아 있다 하더라도, 그가 검토하는 정보는 다른 조직이 수집하고 기업이 개발한 시스템이 가공한 결과다. 표적 정보의 오류로 민간인 피해가 발생했을 때 데이터 관리자는 입력 정보의 문제를, 기업은 소프트웨어의 한계를, 현장 지휘관은 승인 절차의 준수를 내세우기 쉽다. 각자의 역할과 책임이 존재함에도 전체 과정에 대해 설명하고 판단의 정당성을 검증하기는 더욱 어려워진다. 요컨대 AI가 책임의 문제를 처음 만들어 낸 것은 아니다. 다만 이미 여러 손에 나뉘어 있던 판단을 더 빠르고 복잡한 기술적 절차 속에 편입시킴으로써, 책임을 물을 수 있는 조건을 더욱 약화시킬 위험이 있다.

여기서 필자가 주목하는 것은 책임을 귀속하기 어려운 '상태'뿐 아니라 그러한 상태를 만들어 내는 '과정'이다. 책임은 특정 행위와 결과에 대해 개인과 조직, 국가가 부담하는 역할과 의무, 법적·도덕적 귀속의 문제를 포괄한다. 책무성은 그들에게 설명을 요구하고, 답변을 검증하며, 판단과 후속 조치로 이어 가는 관계적·제도적 조건이다. 책임 주체의 이름과 관련 법규정이 남아 있다는 사실만으로 책무성이 작동한다고 볼 수 없는 이유다.

책임을 져야 할 사람과 조직, 법과 절차가 존재하는데도 판단할 시간과 검증할 정보가 부족하고, 독립 조사기구의 접근권이 제한되며, 위법성 판단이 후속 조치로 이어지지 않을 수 있다. 필자는 이처럼 책임을 묻고 응답하는 조건들이 약화되는 동태적 과정을 '책무성의 공동화(空洞化)'라고 부르고자 한다. 그것이 진행된다고 해서 법적 책임 자체가 소멸하는 것은 아니다. 다만 책임의 형식적 주체와 의무는 남아 있는데 책임을 확인하고 이행하도록 하는 기능은 점점 비워질 수 있다. 이 글의 제목인 '책임의 공동화'는 바로 이러한 더 넓은 현상을 가리킨다.

따라서 '책임의 공백'은 책임을 정당하게 귀속하기 어려운 문제를, '책임의 익명화'는 행위자와 결정 경로를 식별하기 어려운 문제를, '책무성의 공동화'는 설명·질문·판단·후속 조치를 연결하는 제도적 기능이 약화되는 과정을 각각 가리킨다. 세 현상은 연결될 수 있으나 동일하지도, 언제나 단일한 인과관계로 이어지는 것도 아니다. 관건은 책무성의 조건을 비워 낸 것이 무엇인지, 누구의 선택이었는지를 추적하는 일이다. 예산과 인력의 배분, 조달 계약, 시스템 설계, 기록 보존과 비공개 결정은 모두 이 질문의 대상이 된다

3. 책무성의 공동화: 네 층위에서 약화되는 책임 추궁의 조건

책무성의 공동화는 적어도 네 층위에서 관찰된다. 첫째, 작전 현장에서는 인간의 승인이 실질적 검증 없는 '고무도장(rubber stamp)'으로 변한다. 둘째, 정부와 기업 사이에서는 데이터·기술·작전의 책임이 서로 다른 계약과 기관으로 분산되면서 설명의 연결이 끊어진다. 셋째, 국제규범의 협상에서는 책임 원칙이 남아 있어도 이를 이행할 조사·보고·검증의 구체적인 요건이 완화될 수 있다. 넷째, 민주주의에서는 의회와 조사기구, 언론과 시민사회의 질문이 계속되더라도 필요한 자료와 권한이 주어지지 않아 검증의 장이 무력화된다. 이 네 가지는 판단의 실질, 설명의 연결, 이행의 조건, 공적 검증의 권한이 각각 어떻게 비워지는지를 보여 준다.

먼저 현장에서는 승인 절차의 외형이 유지되면서 판단의 실질이 비워질 수 있다. 엘리시(Elish)의 '도덕적 완충지대(moral crumple zone)'는 복잡한 시스템의 실패가 실질적 통제권이 작은 최전선 운용자에게 집중적으로 전가되는 현상을 설명한다. 미나브 사건에서 '누가 마지막으로 승인했는가'만 묻는다면 표적 정보의 갱신 실패, 민간인 피해 예방 인력의 감축, 작전 속도를 결정한 상층부의 책임이 시야에서 사라질 수 있다. 인간이 최종 승인했다는 기록은 남더라도 그가 실제로 무엇을 검토할 수 있었는지, 원자료에 접근하거나 결정을 거부할 시간과 권한이 있었는지는 별도로 확인해야 한다. 사전 판단의 형식화가 사후 책무성의 약화로 이어지는 지점이 바로 여기에 있다.[6)]

기업과 정부 사이에서는 데이터의 정확성은 정부의 몫이고 소프트웨어의 기능은 업체의 소관이라는 식의 해명이 이어질 수 있다. 계약상 업무 분담은 필요하지만 전체 위험을 누가 통합적으로 점검하는지 불분명하다면 책무성은 계약의 경계에서 멈춘다. 소프트웨어 업데이트는 재발 방지의 일부이지, 왜 그런 결정이 내려졌고 피해자에게 누가 설명하고 배상할 것인가에 대한 답변을 대신할 수 없다.

2026년 10월 우크라이나군의 스타링크·스타쉴드 통신 이용 제한 논란에서도 스페이스X와 미국 정부의 통제·책임에 관한 공개 설명은 엇갈렸다. 관련 보도와 의회 질의가 이어졌지만, 우크라이나 측은 제한의 경위에 대해 충분한 설명을 받지 못했다고 밝혔다. 이는 AI 표적 선정 장치 자체의 사례는 아니다. 그러나 전쟁에 필수적인 민간 기술 인프라를 둘러싸고 기업과 정부의 권한이 겹칠 때, 피해 당사자가 누구에게 답변을 요구하고 어떤 기관이 기록을 검증할 수 있는지를 묻게 한다.

---

6) Elish, M. C. 'Moral Crumple Zones.' Engaging Science, Technology, and Society 5 (2019).

국제규범의 층위에서는 책임 원칙이 남아 있어도 그 원칙을 실현할 제도적 장치가 약화될 수 있다. 2026년 9월 채택된 특정재래식무기금지협약(CCW) 정부전문가그룹(GGE) 보고서 제25항은 인간의 개입 없이 표적을 식별·선택·교전할 수 있는 체계를 자율살상무기체계(LAWS)로 성격 규정한다. 사람이 추천 표적을 검토하고 승인하는 메이븐(Maven) 같은 AI 의사결정지원체계(AI-DSS)는 통상 이 성격 규정의 바깥에 놓인다. 그러나 LAWS의 범위에서 제외된다는 사실이 국제인도법의 적용까지 면제한다는 뜻은 아니다. 구별·비례성·예방조치 의무는 체계의 자율성 수준과 무관하게 적용된다. 문제는 AI-DSS가 인간의 표적 판단에 미치는 영향을 실효성 있게 규율할 구체적인 절차가 충분하지 않다는 데 있다.

같은 GGE 보고서 제34~35항은 '통제와 인간 판단' 및 작전 범위 제한 등의 요소를 담고 있지만, 인간의 검증과 거부권이 어느 정도로 보장되어야 실질적 통제라고 볼 수 있는지는 충분히 명확하지 않다. 제40~43항에는 기존 국제법상 책임과 개인 책무성에 관한 원칙이 남아 있다. 그러나 제44항의 내부 사건 대응 메커니즘에 관한 문구는 협상 과정에서 강한 의무 표현에서 '권장' 수준으로 완화되었다. 이는 현행 국제인도법상 의무가 소멸했다는 뜻이 아니다. 더욱이 이 보고서는 새로운 구속력 있는 조약이 아니라 향후 규범 형성을 위한 요소들의 합의문이다. 따라서 이 변화는 기존 책임의 소멸이 아니라 향후 규범에서 책임의 이행·조사 조건을 얼마나 강하게 보장할 것인가에 관한 합의가 약해졌다는 점에서 중요하다.[7)]

민주주의의 층위에서 공동화는 책임을 묻는 장(forum)의 부재보다 그 무력화로 나타난다. 유엔 조사단과 언론이 의문을 제기하고 의회가 질의해도 기밀 자료나 디지털 로그에 접근할 수 없다면 답변의 진위를 가리기 어렵다. 국가안보상 정보의 공개 범위를 제한해야 할 경우는 있다. 그러나 일반 대중에 대한 비공개가 곧 독립 감독기구와 국회의 비공개 검증까지 봉쇄하는 근거가 되어서는 안 된다. 공론장에서 질문이 제기되는 것과 법적 권한을 가진 기관이 답변을 강제하고 판단의 결과를 실현하는 것은 서로 다르다. 두 경로가 끊어질 때 책무성을 이루는 관계는 제도적 외형만 남게 된다.

4. 살상의 자동화는 책무성의 조건을 어떻게 약화시키는가.

앞의 네 층위는 책무성이 어디에서 어떻게 약화되는지를 보여 주었다. 그렇다면 AI 군사화는 이 과정에 어떤 변화를 더하는가. AI가 책임을 자동으로 지워 버리는 것은 아니다. 다만 정보가 처리되는 속도, 운용자가 전장을 인식하는 방식, 표적이 표현되는 양상을 바꿈으로써 검증과 숙고의 조건을 압박할 수 있다. 아래의 시간·인식·대상의 세 가지 압축은 책무성의 공동화와 동일한 현상이 아니라, 그 공동화를 촉진할 수 있는 기술적·조직적 조건이다.

첫째, 시간의 압축이다. 2024년 독립 언론 〈+972 매거진〉은 이스라엘군이 가자지구에서 운용한 AI 표적 생성 시스템 '라벤더(Lavender)'가 제시한 일부 표적을 정보요원이 단 20초가량만 검토한 뒤 공습을 승인했다는 내부 증언을 보도했다. 물론 이 수치는 모든 작전에 일괄 적용된 공식 기준이 아니라 취재원의 증언에 기반한 것이지만, 알고리즘이 표적 산출 속도를 끌어올릴 때 인간의 검토 시간이 어떻게 내몰리는지를 단적으로 보여준다. 앞서 살펴본 미나브 사건에서도 과거 수 시간이 걸리던 표적 목록화 과정이 보도상 수 분 단위로 압축되었다. 여기서 핵심

7) CCW GGE LAWS. Final Report, CCW/GGE.1/2026/3, pp.25–45 (2026): 피스모모 CCW GGE LAWS 2차 회의 모니터링 보고서

은 인간이 누르는 '승인 버튼'이 형식적으로 존재하는가 여부가 아니다. 인간 검토자가 데이터의 출처와 시점을 확인하고, 알고리즘의 추천과 다른 해석의 가능성을 따져보며, 필요할 경우 공격을 멈추도록 요구할 수 있는 실질적인 시간과 절차가 보장되었는가에 있다.

둘째, 인식의 압축이다. 수집 시점과 출처, 신뢰도가 제각기 다른 파편적인 정보들이 알고리즘을 거쳐 하나의 정돈된 화면과 우선순위 추천 목록으로 결합되면, 그 매끄러운 시각적 외양이 마치 정보 자체의 정확성과 완결성을 보증하는 것처럼 보이기 쉽다. 이 과정에서 원본 데이터가 지녔던 불확실성과 맥락은 화면 뒤로 사라진다. 이른바 '자동화 편향(automation bias)' 연구들이 일관되게 지적하듯, 시간적 압박을 받는 인간 운용자는 기계가 깔끔하게 정리해 제시한 결과를 객관적 사실로 과신하고 번거로운 교차 검증을 생략할 위험이 크다. 따라서 메이븐(Maven)과 같은 AI 의사결정지원체계(AI-DSS)를 둘러싼 핵심 문제 역시 단순히 알고리즘의 연산 속도에만 있지 않다. 데이터가 얼마나 최신 상태로 갱신되는지, 인터페이스가 정보의 불확실성과 오류 가능성을 어떻게 경고하는지, 그리고 운용자가 추천 결과 너머의 원자료(raw data)에 얼마나 쉽게 접근해 의문을 제기할 수 있는지가 판단의 질을 좌우한다.

셋째, 대상의 압축이다. 구체적인 삶의 맥락을 지닌 사람과 학교, 주거지, 그리고 그 안의 일상적 관계들이 군사 데이터베이스 안에서는 추상적인 표적 코드와 좌표, 위험 점수로 환원된다. 과학기술학자 루시 서치먼(Lucy Suchman)이 비판한 '전지(全知)의 상상'은 이처럼 방대한 데이터를 통합해 화면에 띄우면 전장의 현실을 빈틈없이 파악할 수 있다는 믿음을 가리킨다.[8] 그러나 데이터베이스에 포착된 세계는 현실의 일부일 뿐이며, 때로는 이미 지나간 과거의 기록에 머물러 있다. 미나브의 학교는 군사 데이터베이스상 '과거의 군사시설'이라는 코드로 남아 있었지만, 공격 당시 그 건물 안에는 그 분류와 무관하게 일상을 살아가던 아이들이 실제로 존재했다. 대상을 데이터로 압축하는 순간, 숫자로 번역되지 않은 현장의 삶과 변화는 시야에서 지워진다.

물론 AI 기술이 민간인 밀집 지역에 관한 정보를 실시간으로 종합하고 부수적 피해 위험을 조기에 탐지하는 데 기여할 가능성마저 부정할 필요는 없다. 국제적십자위원회(ICRC) 역시 그러한 긍정적 활용 가능성을 배제하지 않으며, 따라서 자동화 자체가 언제나 인간의 판단을 제거한다고 단정할 수는 없다. 그러나 실제 전장 환경에서는 세 가지 조건이 동시에 맞물리며 인간의 숙고를 압박한다. 인간이 감당하기 어려운 수준으로 폭증하는 표적 처리량, 데이터의 불확실성을 가리는 단일한 시각화 화면, 그리고 알고리즘의 추천을 따를 때보다 거부할 때 더 큰 설명 책임을 지게 만드는 '입증 부담의 전도'가 결합할 때, 인간 판단의 독립성은 구조적으로 약화될 수밖에 없다.

같은 맥락에서 '정밀타격'과 '적법한 표적 선정' 역시 엄격히 구분되어야 한다. 유도무기가 지정된 좌표를 오차 없이 정확히 타격했다는 사실과, 애초에 그 좌표가 국제인도법상 공격이 허용되는 적법한 군사목표물이었는지는 전혀 다른 차원의 질문이다. 시간과 인식, 대상이 압축된 상태에서 잘못된 표적이 선정되었다면, 타격의 물리적 정밀성은 오히려 잘못된 결정을 더 확실하고 치명적으로 집행하는 수단이 될 뿐이다. 요컨대 무기의 기술적 정확성은 표적 선정 과정의 법적·윤리적 정당성을 결코 대신할 수 없다.

결국 쟁점은 속도 자체가 아니라 빨라진 의사결정과 동시에 독립적인 검증과 거부, 기록 보존이 강화되는가이다. 인간이 짧은 시간에 추천을 승인하도록 설계된 체계가 사후에는 승인자의

8) Lucy Suchman, &Imaginaries of Omniscience,& Social Studies of Science 53(5) (2023).

이름과 클릭 기록만을 남긴다면, 절차가 있었음을 증명하는 자료와 실질적 판단이 있었음을 증명하는 자료 사이에 간극이 생긴다. 이 간극을 조사할 독립적인 장치가 마련되지 않을 때 기술적 압축은 책무성의 공동화로 이어질 수 있다.

5. 책임의 이행을 어렵게 만드는 규범적·제도적 선택들

국제법상 전쟁 개시 결정의 적법성(jus ad bellum)과 전쟁 수행 중 개별 공격의 적법성(jus in bello)은 구별되는 개념이다. 또한 개인에 대한 국제형사책임, 지휘관의 감독 책임, 국가의 국제법상 책임, 피해자의 배상과 구제는 각각 다른 법적 경로를 따른다. 국제형사재판소(ICC) 로마 규정 제25조는 재판소가 직접 형사소추하는 대상을 자연인으로 한정하지만, 그것만으로 기업의 국내법상 민·형사책임이나 임직원의 책임 가능성이 사라지지는 않는다.[9)]

민간시설 공격에 대한 개인의 형사책임을 입증하려면 해당 행위자의 인식과 고의, 지휘관의 정보 접근성 및 주의 의무 등 개별 요건을 검토해야 한다. 잘못된 데이터는 공격의 원인을 설명하는 자료이면서, 동시에 누가 무엇을 알고 있었는지를 가리는 단서가 된다. 그러나 '표적 정보가 틀렸다'는 한 문장만으로 모든 책임이 면제되지는 않는다. 무엇보다 개별 형사책임의 입증이 어렵다는 이유로 국가의 예방·조사·구제 의무까지 사라지는 것도 아니다. 따라서 데이터, 추천 내역, 승인 근거, 지휘 명령과 변경 이력의 기록·보존은 정의 실현을 위한 필수 조건이다.[10)]

그러나 한국에서는 책임을 실질적으로 추궁할 조건을 충분히 구체화하지 못한 채 군사 AI의 개발과 획득을 촉진하는 법률과 법안들이 등장하고 있다. 현행 「인공지능기본법」 제4조 제2항은 '국방 또는 국가안보 목적으로만 개발·이용'되며 대통령령으로 정하는 AI를 적용 대상에서 제외한다. 모든 군사 관련 AI가 일률적으로 적용 제외되는 것은 아니지만, 시행령에 따른 예외 범위가 넓게 설정될수록 이중용도 기술에 대한 외부 통제는 취약해질 수 있다. 국가인권위원회도 시행령 입안 과정에서 예외의 범위와 인권 보호체계에 관해 우려를 제기한 바 있다.[11)]

2026년 발의된 「국방인공지능법안」(의안번호 2216355)은 AI 산출물에 '예측·추천·결정'을 포함하면서도, 인적 개입 보장 조항은 생명·신체에 중대한 위해 우려가 있는 '자동화된 결정(Automated decision-making)'을 중심으로 규정한다. 국방부장관 중심의 위원회 구성에서도 독립적인 인권·국제인도법 전문가의 참여가 명시적으로 보장되지 않으며, 제네바협약 제1추가의정서 제36조에 따른 신무기·전쟁 수단 및 방법의 법률검토와 정기적 국회 보고 역시 법안에 충분히 명문화되지 않았다. 반면 기술 획득과 신속한 전력화에 관한 특례는 구체적이다.

여기에 2026년 8월 31일 여야 의원들이 공동 발의한 「국방첨단전력사업에 관한 법률안」(의안번호 2220937)이 더해진다. 이 법안은 인공지능·드론·로봇 등 첨단기술의 빠른 군사적 도입을 위해 기존의 무기 획득체계와 별도로 신속한 개발·배치 절차를 마련하려는 취지다. 구체적으로 제3조는 국방첨단전력사업에 관하여 다른 법률보다 이 법을 우선 적용하도록 하고, 제16조는 운용군 중심의 최종사용자집단이 실시한 실사용평가 결과가 확정되면 「방위사업법」에 따른 시험평가를 거친 것으로 간주한다. 나아가 제17조와 제18조는 최소기능품을 우선 배치하고 실제 운용 과정에서 반복적으로 성능을 개량하는 방식을 허용한다. 기술 변화가 빠른 첨단체계의 특

9) Rome Statute of the International Criminal Court (1998), arts. 8, 25, 28, 30, 32.
10) 1977 Additional Protocol I, arts. 36 ' 57; ICRC, A Guide to the Legal Review of New Weapons (2006).
11) 국가인권위원회, 「인공지능기본법 시행령 제정안에 대한 의견」(2025.12).

성을 고려한 제도라는 취지는 이해할 수 있다. 그러나 무기의 기술적 성능을 확인하는 시험평가와 민간인 보호, 국제인도법 준수, 인간 통제의 실효성을 검증하는 법적·윤리적 심사는 서로 대체될 수 없다. 특히 성능과 운용 방식이 계속 변화하는 AI 무기체계에 대해서는 도입 이전의 법률검토뿐 아니라 중대한 변경이 이루어질 때마다 적법성을 다시 확인하는 절차가 필요하다.

참여연대 평화군축센터와 성공회대학교 평화월딩연구소는 2026년 9월 발표한 입법의견서에서 이 법안이 기존 무기체계 검증절차를 우회하고, 책임 경감·면제 및 군사기술의 공공부문 확장을 통해 안전과 인권 보호체계를 약화시킬 수 있다고 지적했다.[12] 이러한 선행 분석을 '책무성의 공동화'라는 관점에서 다시 살펴보면, 검증절차가 간소화된다는 사실을 넘어 더 근본적인 문제가 드러난다. 개발·획득·운용의 각 단계에서 담당자와 법정 절차가 남아 있더라도, 그 절차의 적정성을 누가 어떤 기준으로 검증할 수 있는지 보장되지 않는다면 책임 추궁의 제도적 연결이 약화되기 때문이다.

이러한 관점에서 더욱 주목할 조항은 제24조의 책임 경감·면제 규정이다. 이 조항은 법에서 정한 절차에 따라 업무를 성실하게 수행하고 고의 또는 중대한 과실이 없는 경우 징계나 손해배상 책임을 경감하거나 면제할 수 있도록 한다. 물론 이것이 모든 법적 책임을 면제하거나 국제인도법상 의무를 배제한다는 뜻은 아니다. 그러나 정해진 절차의 준수가 책임 경감의 중요한 근거가 되는 반면, 그 절차 자체가 인간의 실질적 판단과 충분한 안전 검증을 보장하는지는 별개의 문제로 남는다.

바로 이 지점에서 책무성의 공동화가 제도적으로 발생할 위험이 커진다. 무기 도입과 운용에 관여한 사람들이 각자 법정 절차를 충실히 따랐다고 설명하더라도, 그 절차 자체가 안전성과 적법성을 충분히 검증하도록 설계되었는지 별도로 확인해야 한다. 어떤 기구가 위험을 사전에 심사하고, 누가 실사용평가의 근거를 검증하며, 사고가 발생하면 어떤 기록을 제출받아 판단할 수 있는가. 정해진 절차의 준수가 책임 경감의 근거로 작동하면서 절차의 적정성을 독립적으로 점검할 통로가 취약하다면, 책임을 묻는 제도는 외형만 남기 쉽다. 다만 법안의 시험평가 갈음이나 책임 경감 규정이 제네바협약상 법률검토 의무 또는 모든 국내외 법적 책임을 자동으로 면제한다는 뜻은 아니다.

또한 법안 제14조는 군이 사용하는 첨단전력체계를 다른 정부기관이나 공공 분야에서도 활용할 수 있도록 협력의 근거를 두고 있다. 이는 군사적 목적에 따라 개발된 감시·식별·추적 기술이 경찰 등 다른 공공기관의 업무로 확장될 가능성과 직결된다. 군사 AI의 위험을 전장에서 발생하는 살상에만 국한할 수 없는 이유다. 전장을 전제로 개발된 기술이 시민의 일상에 도입될 때에는 개인정보 보호, 차별과 오분류, 감시 권한의 확대에 대한 별도의 인권적 검토가 필수적이다. 그러나 이 법안은 군과 공공 영역 사이의 기술 공유를 촉진하면서도, 그러한 전환 과정에 적용할 독립적인 인권영향평가와 민주적 감독 장치는 충분히 구체화하지 않았다.

이처럼 한국의 AI 군사화 관련 입법은 일반법의 안보예외와 군사 특별법의 불충분한 통제가 맞물리는 '이중 공백'을 드러낸다. 첨단전력 획득 절차의 특례까지 더해지면서 개발과 배치를 빠르게 하는 제도적 장치는 구체화되지만, 위험 평가의 독립성, 사용 중단의 조건, 기록 접근권, 피해 발생 후 조사와 구제에 관한 규정은 그만큼 강화되지 않는다. 여기서 '공동화'는 국가나 개인의 법적 책임 자체가 없어지는 현상을 뜻하지 않는다. 실질적 검증과 설명, 판단과 구제가 하나의 연쇄로 작동하도록 보장할 제도적 조건이 약해지면서 책무성이 비워지고, 그 결과 책임

12) 참여연대 평화군축센터・성공회대학교 평화월딩연구소(2026. 9. 16.), 「국방첨단전력사업법 제정안에 대한 입법의견서」.

의 이행이 어려워지는 위험을 가리킨다.

6. 空洞에서 共同으로: 책임을 묻는 관계의 복원

이 글에서 책임의 몫을 기계, 즉 AI에게 조금도 나누어주지 않는 이유는, 기계가 결과에 아무런 영향을 미치지 않아서가 아니다. 책임을 일정한 총량처럼 여겨 기계에 일부를 배분하는 순간, 정작 그 기술을 설계·조달·운용하고 감독 규칙을 만든 인간과 국가, 기업이 자신의 몫을 기계 뒤로 숨길 퇴로가 열리기 때문이다.

정치철학자 아이리스 매리언 영(Iris Marion Young)은 과거의 위법행위에 대해 행위자에게 묻는 법적 귀책의 모델과, 여러 행위자가 얽혀 만들어 낸 구조적 부정의를 바로잡기 위해 함께 져야 할 미래지향적인 '사회적 연결 모델'의 책임을 구분했다. 구조를 바꾸기 위한 공동의 책임은 개별 행위자의 법적 책임을 대체하거나 희석하는 면죄부가 아니다. 오히려 법적 귀책과 구조적 변화를 위한 정치적 책임은 각각의 방식으로 동시에 추구되어야 한다.[13)]

이 글에서 '책임의 공동화(空洞化)'는 책임져야 할 사람이나 법적 의무가 사라졌다는 의미로 사용되지 않았다. 책임의 주체와 규범이 유지되는 가운데, 설명을 요구할 권한과 검증할 정보, 독립적으로 판단할 제도, 그 결과를 실현할 수단이 비워지면서 책임의 이행이 약화되는 현상을 짚고자 했다. 그 내부에서 진행되는 핵심 과정이 '책무성의 공동화(空洞化)'다. 그렇다면 이에 맞서는 '공동화(共同化)'는 모두가 똑같이 잘못했다는 뜻이 아니라, 각자의 책임을 끝까지 물을 수 있도록 공적 관계와 제도적 조건을 함께 만들어 가자는 요청이다. 적어도 다음 다섯 가지 조건이 필요하다.

첫째, 실질적으로 판단할 시간과 거부권이 보장되어야 한다. 인간이 최종 승인 버튼을 누른다는 형식만으로 실질적 통제가 이루어졌다고 인정해서는 안 된다. '시간의 압축'과 '인식의 압축'을 넘어설 수 있도록, 운용자에게는 데이터의 출처와 갱신 시점, 불확실성을 교차 확인할 수 있는 원자료와 물리적 시간이 제공되어야 한다. 나아가 알고리즘의 추천을 거부하거나 작전 중단을 요청하더라도 불이익을 받지 않을 실질적인 권한이 제도적으로 보장되어야 한다.

둘째, 추적 가능한 기록(log) 보존과 독립적인 감사 접근권이 결합되어야 한다. 표적 데이터의 원천과 변경 이력, 알고리즘이 제시한 추천 목록, 인간 운용자의 판단 근거, 그리고 승인·거부·중단에 이르는 전 과정의 결정 이력이 빠짐없이 보존되어야 한다. 다만 디지털 로그가 서버에 존재하는 것과 외부의 독립 조사기구가 이를 실제로 열람하거나 검증할 수 있는 것은 전혀 다른 문제다. 따라서 기록의 생성, 보존 의무와 독립 기구의 접근권은 반드시 하나의 세트로 의무화되어야 한다.

셋째, 무기와 전쟁 방법의 생애주기 전반에 걸친 법률검토가 이루어져야 한다. 제네바협약 제1추가의정서 제36조의 신무기 및 전쟁 수단·방법 법률검토를 형식적 절차에 그치지 않고 실효성 있게 이행해야 한다. 특히 자율살상무기체계(LAWS)뿐 아니라 새로운 전쟁 수단·방법에 해당하거나 그 운용 방식을 실질적으로 바꾸는 AI 의사결정지원체계(AI-DSS)까지 법률검토의 범위에 포함하도록 국내 절차를 명확히 해야 한다. 배치 이후에도 알고리즘 재학습, 주요 데이터·모델 변경, 운용 환경의 중대한 변화로 최초 검토의 전제가 달라지면 적법성을 다시 확인하는 상시

13) Young, I. M. Responsibility for Justice. Oxford University Press (2011)

적 재검토 체계를 마련해야 한다.[14]

넷째, 정부와 기업 사이에 희미해진 책무성의 고리를 복원해야 한다. 무기체계의 조달 및 계약 단계에서부터 데이터의 품질 검증 책임, 위험 및 오류 통지 의무, 로그 기록 보존, 사고 발생 시 기업의 조사 협조 의무, 그리고 심각한 위험이 확인될 경우 즉각 수정을 요구하거나 사용을 중단할 수 있는 조건을 계약서에 명문화해야 한다. 국가는 계약상의 업무 분장이나 기업의 영업비밀을 핑계로 시민의 생명과 직결된 공적 의무를 외주화할 수 없다.

다섯째, 법률과 국회를 통한 민주적 감독을 복원해야 한다. 「국방인공지능법안」의 인적 개입 보장을 '자동화된 최종 결정'만이 아니라 AI의 '예측·추천·결정' 전반으로 확장하고, 「국방첨단전력사업법안」의 신속 획득과 실사용평가 특례도 독립적인 적법성·인권 심사에서 면제되지 않도록 해야 한다. 국제인도법과 국제인권법 전문가, 독립적인 민간 전문가가 참여하는 검증체계와 정기적인 국회 보고를 법제화할 필요가 있다. 군사 보안상 대중에게 공개하기 어려운 자료라도 법적 권한을 가진 감독기구와 국회의 비공개 심사 대상에서까지 제외되어서는 안 된다.

나아가 국제사회 차원에서는 사람을 직접 표적으로 삼는 대인 자율살상무기와 결과를 예측·통제할 수 없는 무기체계를 전면 금지하고, 그 밖의 자율무기체계 역시 엄격히 제한하는 법적 구속력 있는 국제협약을 조속히 마련해야 한다. 물론 현 시점의 전망은 좋지 않다. 또한 좁은 의미의 LAWS 규율과는 별개로, AI 의사결정지원체계(AI-DSS)가 표적 추천 과정에서 만들어 내는 판단의 위험 역시 국제인도법의 엄격한 해석과 각국의 국내 입법을 통해 촘촘히 규율하되, 국제인도법은 출발선이지 도착지점이 아니라는 것을 환기하며, 국제인권법의 영역으로 확장되어야 한다.

7. 나가며: 공동화(空洞化)에서 공동화(共同化)로

책임은 결국 인간에게 있다는 말이 공허한 수사에 그치지 않으려면 그 인간에게 실제로 책임을 추궁할 수 있는 제도가 마련되고 작동해야 한다. 자원과 생명을 끊임없이 채굴하고 수탈해 온 추출주의가 지구의 삶의 토대를 갉아먹어 온 것처럼, 책임 회피와 기술 만능주의는 책임을 물을 시간과 정보, 권한이라는 공적 토대를 소진시킬 수 있다. 살상을 더 빠르고 효율적으로 수행한다는 명분 아래 숙고할 시간은 줄어들고 오류를 검증할 기회는 축소되며, 설명을 요구할 기록과 권한은 접근하기 어려운 곳으로 옮겨 간다. 사람과 조직, 법과 제도의 외형은 남아 있지만 이를 실제로 작동시키는 책무성의 조건은 비워지는 것이다.

그러나 이 공동화(空洞化)를 그저 비어 있는 황무지로 남겨 둘 이유는 없다. 땅을 파헤치고 관계를 해체하는 추출주의적 폭력에 맞서 훼손된 삶의 토대를 회복하듯, 책임의 공동화에 맞서는 일은 서로의 질문과 판단이 연결되는 공적 관계를 다시 일구는 일에서 시작할 수 있다. 공동의 책임을 만들어 가는 공동화(共同化)는 개별 행위자의 위법행위와 국가의 책임을 희석하는 말이 아니라, 그 책임을 끝까지 묻기 위한 사회적·제도적 역량을 함께 복원하자는 요청이다. 평화와 안전이 특정 권력이나 기술의 전유물이 아니라 모두의 삶이 의존하는 토대라면, 책임을 묻는 자리 또한 모두의 것이어야 한다. 누가, 어디에서, 무엇을 근거로 책임을 물을 것인가. 그 질문에 응답하도록 요구할 권리와 조건을 서로의 손으로 다시 짓는 일, 그것이 비어 버린 정의의 자리를 살아 있는 실천으로 채워 가는 길일 것이다.

---

14) 1977 Additional Protocol I, arts. 36 ' 57; ICRC, A Guide to the Legal Review of New Weapons (2006).