사이버안보연구, 3집 1호 2026 여름, 007-047

AI 기술의 공세적 사이버 작전 역량체계 통합과 자국 기업의 공급망 배제의 넥서스:

미국 전쟁부 대 앤트로픽 소송이 보여준 국가-기업 관계의 균열

유인태 | 단국대학교(부교수, 제1저자, 교신저자)
김예준, 이규민, 장지성, 양지훈 | 단국대학교(학부)

논문접수일 2026년 5월 20일 | 논문심사일 2026년 6월 14일 | 게재확정일 2026년 6월 22일

논문 요약

본 연구는 미국 전쟁부와 인공지능 기업 앤트로픽(Anthropic) 간의 법적 소송을 공세적 사이버 작전과 공급망 안보의 관점에서 분석한다. 기존 연구는 공세적 사이버 작전에서 민간 AI 기업의 실제 역할을 충분히 조명하지 못하였으며, 공급망 안보 논의 또한 위협 행위자를 외국 기업에 국한하여 자국 기업이 공급망 위협으로 지정되는 현상을 설명하지 못하는 한계를 지녔다. 본 연구는 법원에 제출된 소송 문서를 직접 독해함으로써 이 이중의 연구 공백을 메운다. 분석 결과, 클로드(Claude)는 미 중부사령부의 표적 식별, 대테러 정보 분석 등 실전 군사 작전에 통합 운용된 최초의 프런티어 AI 모델로서 국가 공세적 사이버 작전 체계의 실질적 내부 구성 요소로 기능하였음이 확인되었다. 또한 미국 전쟁부는 앤트로픽의 '안전 가드레일' 정책이 군 지휘권을 침해하는 '작전 거부권'에 해당한다고 규정하고, 실시간 작전 승인 요구, 기밀 유출, 이용 약관에 의한 작전 범위 제한, 시스템 무단 변경 가능성, 외국 국적 인력을 통한 기술 취약점을 근거로 자국 기업을 전례 없는 '공급망 위험' 업체로 지정하였다. 이러한 법적 공방에서 보이는 것은 공세적 사이버 작전에 포함되는 기술 기업의 높은 기술 력과, 비록 자국 내 민간 기업일지라도, 정부의 공급망 위험 행위자 지정이 연결되고 있다는 것이다. 결론에서 본 연구는 AI 시대 가공할 사이버 보안기술적 발전에 대한 대응뿐 아니라, 국가-기업 관계의 재편과 AI 거버넌스 제도화의 시급성을 환기하는 정책적 함의를 아울러 제시한다.

주제어 공세적 사이버 작전, 공급망 안보, 인공지능 거버넌스, 앤트로픽, 첨단 기술 기업, 민간 AI의 군사화, 국가-기업 관계

Ⅰ. 서론

본 연구의 목적은 미국 정부, 특히 전쟁부와 자국 AI 기업으로 알려진 앤트로픽(Anthropic) 간의 법적 소송을 사이버 작전과 공급망 관점에서 분석하는 것에 있다. 일견, 연구 범위가 상대적으로 작아 보일 수 있지만, 이 논문이 집중해서 분석하는 문서와 내용이 갖는 함의의 중요성은 매우 크다. 해당 법적 다툼은 단순히 양자 간의 갈등을 보이는 것이 아니라, 미국 정부가 공세적 사이버 작전을 첨단 기술 기업과 어떻게 협동하여 전개해 왔는지를 보이기 때문이다. 그리고 그 기술 기업이 작금의 패권 전쟁의 핵심에 있는 선두 AI 기업이기 때문이다. 법적 다툼에서 드러난 정부와 기업 양자 간의 협력은 미래의, 아니 이미 현재가 되어버린 새로운 전쟁 수행 양상을, 나아가 국가 안보 활동을 대변한다. 그뿐 아니라 해당 법적 다툼의 과정과 결과는 정부와 첨단 기술 기업 간의 관계 정립에 중요한 이정표가 될 수 있다. 따라서, 단순히, 미국 정부와 한 기업 간의 법적 소송에 그치지 않으며, 정부와 수많은 첨단 기업들 간에, 그리고 그들의 영향을 받는 사회와 정부 간의 미래의 관계에 영향을 미칠 수 있는 법적 분쟁이다.

비록 해당 법적 다툼이 가질 수 있는 개인, 회사, 국가-사회, 국제 등 여러 층위의 함의를 다 논하지 못하지만, 이 연구는 공세적 사이버 작전과 공급망 위협의 관점에 집중하여 수행된다. 그 이유는 네 가지이다. 첫째, 공세적 사이버 작전의 실질적 내용에 대한 연구 부족이다. 기존 연구들의 대부분은 국가 중심이거나 실제 작전에 대한 제한된 정보만을 가지고 연구가 되어 왔다. 많은 연구자들에게 정부가 사이버안보 기업과 특히 AI 기업과 합동으로 군사·정보 작전을 수행하는 실제에 대해서는 공식적으로 밝혀진 바가 제한적이었다. 밑의 절

에서 해당 연구 공백에 대해 상술하지만, 미국 정부 대 엔트로픽의 법적 다툼 가운데 AI 기업과 사이버 작전의 참여 양상이 나타났기에 본 연구에서는 미국 정부와 앤트로픽의 구체적 사이버 작전 협력 양상을 밝힌다.

둘째, 공급망 안보 연구 분야에 대한 기여다. 공급망 안보 자체가 비교적 신흥 안보 이슈임에도, 기존 연구에서 다루어지지 않은 또 다시 새로운 국면이 미국 정부 대 앤트로픽의 법적 다툼 가운데 등장했다. 역사상 유례없는 자국 기업, 그것도 가장 역량 있는 AI 기업 중 하나로 알려진 앤트로픽의 공급망 위험 지정이다. 이러한 정부의 움직임에 대해 어떤 근거가 동원되는지를 분석한다.[1)]

셋째, 앞의 두 현상을 잇는 '넥서스'에 대한 정보 제공이다. 법적 소송에서 드러난 정보에 대한 기존 분석(예: 언론)은 앤트로픽의 사이버 작전 참여와 앤트로픽 공급망 안보 위협 지정을 별개의 사안으로 다루어 왔다. 그러나 두 현상은 연결되어 있다. 민간 AI 기업인 앤트로픽이 미국의 공세적 사이버 작전에 통합되고, 핵심적 기술로 활용되는 과정에서 발생한 통제권 갈등이 바로 공급망 위험 지정의 직접적 근거가 되었기 때문이다. 본 연구는 두 현상을 동시에 분석함으로써 민간 AI 기업의 사이버 작전 민간화과 공급망 배제 간의 넥서스를 보인다.

마지막으로, 넷째, 법률 문서의 직접 독해를 통한 확실한 증거(hard

1) 2026년 4월 28일 미국 정부 스스로가 앤트로픽에 대한 공급망 위험 지정을 해결하려고 나섰다. 이는 앤트로픽의 새로운 모델인 미토스를 내놓자 나온 반응이다. 이하에서 주장하듯이, 이러한 사태의 전개 자체가 AI 기술의 필수불가결성, 정부의 취약성, 그리고 정부와 AI 기업 간의 기울어진 기술 역량 관계를 드러내고 있다.

evidence) 제공이다. 많은 사람들이 법적 소송에 '대해서는' 알고 있지만, 소송 문서 내용은 '실제로' 잘 읽지 않는다.[2] 그런데 소송 문서의 내용은 언론이 단편적으로 밝혀온 사실이나 추정과는 질적으로 다르다. 당사자가 언급하는 내용은 지금까지 언론과 같은 제3자를 통해서 밝혀진 내용과는 진실성의 정도와 범위가 다를 수 있다. 법적 기관에의 제출은 진실 선언을 전제하는 것이기 때문에, 다른 자료들에 비해 상당 정도의 진실성을 가정할 수 있다. 보통이라면 국가 안보라는 이유로 얻을 수 없는 기밀(confidential) 정보를, 관찰자의 입장에서 확인(verification)의 부담 없이, 행위 당사자가 정보에 대해 정당성을 입증(vindication)하고 있기 때문이다.

본 연구는 위와 같은 학문적 기여뿐 아니라, 정책적 기여도 수반한다. 정책 실무자들은 본 연구의 구성과 같은 맥락에서 두 가지 측면에서의 정책적 대응이 요구된다. 첫째, 미국이 수행하는 공세적 사이버 작전을 한국이 수행할 수 있는 또는 동맹국으로서 협력할 수 있는 역량과 제도적 기반이 한국에 있는가를 묻고 마련해 나갈 필요가 있다.

2) 이 연구에서 사용된 소송 문서 코퍼스(corpus)는 다음과 같다. 본 연구는 두 개의 병행 절차, 즉 Anthropic PBC v. U.S. Department of War, No. 3:26-cv-01996-RFL (N.D. Cal.) 및 Anthropic PBC v. United States Department of War, No. 26-1049 (D.C. Cir.)에서 도출된 법률 문서 코퍼스를 분석하였다. 전체 문서를 점검하였으며 특별히 본 논문과 관련되는 부분은 본문에서 인용하였다. 해당 자료는 2026년 2월 26일부터 4월 10일 사이에 제출되거나 작성된 총 56건의 문서(약570 페이지 분량)로 구성된다. 문서는 PACER/RECAP(CourtListener)을 통한 공개 법원 기록물과 당사자들의 공개 성명을 통해 수집되었으며, 코퍼스에는 소장(complaints), 신청서 및 준비서면(motions and briefs), 진술서(declarations), 증거자료(exhibits), 정부 명령 및 결정문(government orders and determinations), 그리고 기록 증거로 제출된 공개 성명 등 6개 범주의 문서가 포함된다. 다만, 법원에 의해 봉인되어 있거나 접근 제한 상태로 제출된 몇몇 서면들은 현재 코퍼스에서 제외되었다.

둘째, 공급망 위험 지정 논리를 보고, 해당 논리가 적성국 기업에게도 적용될 수 있지만, 한국 기업에게도 적용될 수 있음을 염두에 두고 대응책을 강구해 놓을 필요가 있다. 더 자세한 정책적 함의에 대해서는 결론에서 상론한다.

이 글의 목적은 미국 정부와 AI 기업인 앤트로픽 간의 법적 소송을 사이버 작전과 공급망 관점에서 분석하는 것에 있기 때문에, 다음과 같은 다소 특수한 구성을 갖는다. 우선, 다음 선행 연구 절에서는 공세적 사이버 작전과 공급망 안보 관련 선행 연구의 간극을 찾는다. 이러한 간극에서 비롯되어 본 연구의 필요성이 부각되지만, 향후 연구의 방향성이 제시되기도 한다. 이어지는 분석 절에서는 공개된 양자 간의 법적 소송 문서들을 모두 조사하여, 공세적 사이버 작전에서 민간 기업의 참여 양상과 미 정부의 자국 기업의 공급망 위험 지정 근거를 정리한다. 결론에서는 발견의 요약과 정책적 함의에 대해 논한다.

Ⅱ. 기존 연구

1. 공세적 사이버 작전의 인공지능 기술 결합과 정치학

기존 공세적 사이버 작전에 대한 연구는 어떤 행위자를 중심으로 설명하는지에 따라 다음과 같이 세 가지로 나눌 수 있다. 첫 번째 부류는 국가 행위자에 집중하여, 각 국가의 사이버 전략을 설명하고 분석하는 데 초점을 맞춘다. 이러한 연구들은 국가 또는 (소)다자 국가

협력체의 전략이 어떠한 방향으로 발전하고 변화해 왔는지 추적하거나, 기존 전략을 평가하고 새로운 전략을 제시하는 데 주력했다(Hurwitz 2014; Lin and Zegart 2017). 미국의 사이버 작전은 아마도 가장 많이 연구된 사례일 것이다(Fischerkeller and Harknett 2017; Kehler et al. 2017; 유인태 2024a). 그중에는 미국 전쟁부의 선제적 방어(defend forward)와 지속적 관여(persistent engagement)와 같은 공세적 사이버안보 전략이 오히려 사이버 갈등을 증폭시킬 것이라고 비판하는 등, 미국의 사이버 작전 교리의 한계를 지적하며 대체할 새로운 포괄적 전략 틀을 도입할 것을 제언하는 연구들도 있다(Healey 2019). 그리고 국가 간 사이버작전 협력에 관한 연구들이 있다. 동맹의 맥락에서 다룬 연구들이나(Lonergan and Montgomery 2022; Platte 2023), (소)다자 협력체에 의한 연구들 중에는 파이브 아이즈(Five Eyes)나 북대서양조약기구(NATO)에 맞추어 사이버작전의 소다자적 협력 방안과 한계를 논한 연구들도 있다(Gold 2020; Jacobsen 2021). 그러나 이런 국가 행위자 중심의 전략 분석은 실제로 어떻게 행위가 펼쳐지고 있는지, 그 실체를 파악하기 어렵다. 그리고 이러한 부류의 연구에 또 다른 한 가지 공백은, 국가가 사이버 작전을 수행할 시에, 특히 고도의 기술적 행위가 동반될 때에, 기술의 실질적 제공자인 민간 기업의 역할을 간과한다는 점이다.

두 번째 연구 부류는 정보기관이나 군 등 정부 행위자의 실제 사이버 작전사례를 분석하는 데 초점을 맞추고 있다. 이 범주의 연구들은 정보기관, 군 등 정부 행위자들이 사이버 공간에서 실제로 수행한 작전 또는 캠페인을 검토함으로써 사이버 작전의 기능과 효용, 한계, 성격 등을 도출해 내고자 한다. 이들 중에 공세적 사이버 작전(Offensive Cyber Operation) 사례에 집중한 예로는 최근 러시아-우크라이나 전쟁

사례를 분석하며 공세적 사이버 작전이 실제 전쟁에서 어떠한 방식으로 활용되며, 전투 과정에서 어떤 기능과 효용을 갖는지를 검토하는 연구(Pedersen and Jacobsen 2024),[3] 스턱스넷(Stuxnet) 사례, 이스라엘의 오차드 작전(Operation Orchard), 2007년 에스토니아 사이버 공격 사례 등 다양한 공세적 사이버 작전사례를 분석한 연구(Lindsay 2013; Jacobsen 2021), 공세적 사이버 작전이 수반할 수 있는 부작용에 대한 연구(Lonergan and Lonergan 2023; Slayton 2017), 공세적 사이버 작전에 사용되는 사이버 무기의 효과나 한계에 대해 논의하는 연구(Kello 2017; Lennart 2021; Smeets 2018) 등이 존재한다.

이러한 연구 흐름은 그동안 이론적이고 개념적으로만 축적되었던 사이버 전략을 실제 사이버 작전사례를 통해 구체화하고, 사이버 작전의 방식과 한계를 경험적으로 조명하였다는 점에서 중요한 기여를 지닌다. 그러나 이들 역시 사이버 작전의 수행 주체를 군이나 정보기관 등 정부 행위자로 한정하고 있어, 실제 작전 수행 과정에서 민간 기술 기업의 역량이 정부의 사이버 작전 수행 체계에 어떻게 통합되고 활용되는지 다루지 않는다는 한계가 있다.

마지막으로, 비록 상대적으로 소수이기는 하나, 국가의 사이버 작전에서 비국가 또는 민간 행위자의 참여를 다루는 연구들이 존재한다. 해당 연구들은 그동안 비국가 행위자가 국가의 사이버 작전에서 조명 받지 못하였다는 점을 지적하며 이들의 중요성을 역설한다. 국가를 대신하여 직접 작전을 수행하거나, 또는 국가의 직접적 지시 없

3) Pedersen and Jacobson(2024)은 사이버 작전의 군사적 효용성에 대한 전반적 평가는 제한적이라는 견해다.

이도 묵인 하에 공격하는 프락시(proxy)로서 비국가 행위자의 역할에 주목하거나(Maurer 2018; Wilner et al. 2024), 국가의 사이버 방어 및 복원력 강화를 위해 하드웨어 및 소프트웨어를 제공하고, 사이버 무기를 개발하고, 정보, 감시 및 정찰 역량을 제공하는 등 사이버 작전에 기여할 수 있는 방식을 분석한다(Broeders 2021; Lachow and Grossman 2018; Lilli 2021). 나아가 기업과 같은 민간 행위자는 수세적 사이버 작전에만 관여할 수 있으며, 공세적 사이버 작전에 관여해서는 안 된다는 등의 규범적 주장을 제기하기도 한다(Pattison 2020). 이처럼 이들 연구는 기존 주목받지 못했던 사이버 작전에서의 민간·비국가 행위자의 역할과 참여를 다양한 측면에서 조명한다는 점에서 의의를 지닌다.

그러나 이러한 연구들은 본 연구의 관점에서 세 가지 한계를 보인다. 첫째, 이들 연구는 비국가 행위자를 정부와 분리된 외부 대리자 또는 프락시로 개념화함으로써, 국가 작전 체계 내부에 통합되어 기능하는 민간 기업을 분석 대상에 포함하지 않는다(e.g., Maurer 2018; Wilner et al. 2024). 둘째, 공세적 사이버 작전에서 민간 행위자의 역할을 다루는 경우에도, 이를 수세적 영역으로 제한하거나(e.g., Slayton 2017, 34; Pattison 2020), 국가 억지 전략에 기여하는 독립적 협력자로 위치시킴으로써(e.g., Lilli 2021), 민간 기업이 국가 작전 수행 과정에 통합되는 양상 자체는 분석 대상으로 삼지 않는다. 셋째, 민간 기업과 민간 기술이 실제 공세적 사이버 작전의 수행 과정에서 국가 체계와 어떻게 통합되고, 어떠한 기능을 수행하는지에 대한 경험적 분석이 충분히 이뤄지지 않았다. 특히, 가장 중요한 점은, 사이버보안 관련 회사가 아닌 AI 기술 회사가 사이버 작전에 어떻게 통합되어 있는지에 대한 구체적인 사실을 조명한 학술적 연구는 아직 없다.

민간 기술 기업, 특히 AI 기업은 단순한 외부 지원 주체가 아니라, 기술 역량과 인프라, 전문성 등을 제공함으로써 정부의 공세적 사이버 작전의 기획과 수행에 실질적으로 기여할 수 있으며, 국가의 기술 시스템에 긴밀히 연계되어 있다. 이러한 점에서 민간 기업의 역할에 대한 면밀한 분석이 요구된다. 따라서 본 연구는 기존 연구의 이러한 공백에 주목하여, 다음 장에서 앤트로픽과 미국 전쟁부의 소송에서 생산된 법원 문서를 활용한다. 해당 문서들을 바탕으로 자료에서 드러나는 앤트로픽이 미국 전쟁부의 공세적 사이버 작전 통합 및 관여 양상을 분석하고, 이를 통해 민간 AI 기업과 기술이 국가의 공세적 사이버 작전과 역량에 어떤 방식으로 기능하고 통합되는지를 규명하고자 한다.

2. 공급망 안보의 정치학

최근 미-중 전략경쟁이 심화하며 첨단 디지털 기술 공급망 안보에 관한 정치학적 논의가 빠르게 축적되고 있다. 기존 연구는 국제관계에서의 사이버안보, 경제안보, 국가 전략 등 다양한 측면에서 공급망 안보를 조명해왔으며, 이러한 위험을 국가 간 지정학적 경쟁, 비대칭적 상호의존, 외부 위협에 대한 담론적 구성 등 국가 간 관계의 맥락에서 비롯되는 문제로 이해했다. 그러나 기존 논의들은 거시적인 국가 간 상호작용의 문제로 이해하는 데 집중하며, 국가 내 공급망의 핵심 행위자 중 하나인 자국 기업이 공급망 안보 위협의 행위자로 지정되고 구성되는 현상은 다루지 못했다. 2026년 미국 정부가 자국 기업인 앤트로픽을 공급망 위협 요소로 지정하고 규제한 사태는 이러한 현상의 최초 사례로서, 기존 공급망 안보 문헌의 설명 범위 바깥에 놓여 있다. 구체적으로는, 기존 연구들이 공급망 안보를 이해하는 방식

에 따라 크게 다음의 세 유형으로 구분할 수 있다.

첫 번째 연구 부류는 사이버 안보 및 데이터 안보의 관점에서 공급망 안보를 조명하며, 공급망 위험을 외국 기업이 공급하는 하드웨어·소프트웨어 인프라에 내재한 기술적 취약점의 문제로 파악한다(Friis and Lysne 2021; Kello 2013; 유인태 2026; 2024b). 이러한 관점에서 공급망 위협은 적대국과 긴밀히 연계된 외국 기업의 기술 확산에서 비롯되며, 설계 단계에서 내장된 백도어, 킬스위치, 악성코드, 그리고 이를 활용한 데이터 탈취, 사보타주, 스파이 활동 가능성이 위협의 핵심 원천으로 제시된다. 대표적 사례로 중국 통신 기업 화웨이의 5G 사태를 들 수 있다. 화웨이 장비는 2018년 이후 미국을 비롯해 일본, 영국, 호주 등 일부 동맹국에서 공급망 안보의 위험 요소로 지정되었으며, 지정의 핵심 근거는 화웨이와 중국 정부 간 긴밀한 연계 가능성 및 이를 통한 사보타주·스파이 활동의 수단화 우려였다(Krolikowski and Hall 2023; Pancevski 2020; Sun and Wang 2024). 이와 같은 관점은 최근 미국 정부가 활용해 온 공급망 안보 해석의 논리를 설명하고, 첨단 기술이 공급망 위험으로 전환되는 경로를 드러낸다는 점에서 의의가 있다.

두 번째 연구 부류는 지정학적 경쟁 하에서 공급망 안보를 단순한 기술적 보안 위험의 존재 여부에 대한 개념으로 이해하기보다 정치 행위자가 의도를 가지고 특정 이슈를 안보 영역으로 이동시키는 전략적 수단으로 주목한다. 이 관점에서 공급망 위험은 객관적으로 주어지는 것이 아니라, 지정학적 경쟁 하에서 특정 국가를 견제하기 위한 정당화 수단으로 기능한다(김상배 2019; 김지영 2019; Campion 2020; Sun and Wang 2024). 일부 연구는 기술적 취약성에 의한 실존적 위험과 정치적 수단으로서의 기능적 역할을 통합적으로 분석하여, 두 관

점이 상호작용하거나 결합될 수 있음을 보여주기도 했다(Friis and Lysne 2021; 유인태 2024b). 이러한 문헌들은 기술적 보안 위협의 관점에서 이해되어 온 공급망 안보를 국제 패권 경쟁 또는 국가의 실존적 위협으로서 설명하며 기존 논의를 확장하는 데 기여했다. 다만, 이러한 연구들은 안보화의 대상을 일관되게 국가 간 경쟁 구도 내 외국 행위자로 설정하고 있으며, 지정학적 경쟁이라는 맥락 속에서 실시된 외국 기업에 대한 안보화 과정에 집중한 나머지, 자국 내 민간 기업이 국가에 의해 공급망 위협으로 구성되는 현상은 다루지 않는다.

마지막으로, 또 다른 연구 흐름은 공급망 안보를 비대칭적 상호의존 구조에서 발생하는 취약성의 문제로 이해한다. 코로나19 팬데믹, 러시아-우크라이나 전쟁, 미-중 전략경쟁을 거치며 반도체, 의약품, 원자재 등 핵심 품목의 글로벌 공급망이 중단되거나 전략적으로 무기화될 수 있다는 점이 부각되었다(Donnelly 2023; Yoshimatsu 2026). 이러한 연구들은 무기화된 상호의존에 주목하면서, 비대칭적 상호의존 관계에 놓인 국가들 사이에서 공급망이 공급 중단이나 경제적 강압의 수단으로 활용될 수 있으며, 바로 그 점에서 공급망 자체가 취약성의 원천이 될 수 있다고 본다(Farrell and Newman 2019). 이 관점에서 타국의 자원, 기술, 생산 네트워크에 대한 의존은 곧 공급망 위협으로 해석되며, 공급망 안보는 안정적 조달과 공급 지속성을 확보하는 공급 안보의 문제로 이해된다. 따라서 이러한 연구들은 미국, 일본, 유럽, 한국 등 주요 국가들이 공급망 취약성을 완화하기 위해 어떠한 전략을 채택하고 있는지에 주목한다(Glencross 2024; Koo 2025; Lee 2025). 이 같은 연구 흐름은 경제안보의 관점에서 공급 의존성이 어떻게 위협으로 전환되는지를 보여준다는 점에서, 미국 정부의 앤트로픽 공급망 위협 지정 사태를 해석할 이론적 단서를 제공한다. 그러나 이 또

한 공급망 위협을 기본적으로 국가 간 관계와 대외 의존의 맥락에서 이해하기 때문에, 국가 내부에서 특정 자국 기업에 대한 의존이 안보 위협으로 재구성되는 현상에 대해서는 다루지 않고 있다.

종합하자면, 기존 공급망 안보 연구는 위협의 원천을 기술적 취약성, 안보화의 정치적 구성, 또는 비대칭적 상호의존이라는 세 가지 경로로 설명해왔다. 그러나 세 접근 모두 위협 행위자를 외국 국가 또는 외국 기업으로 전제하거나, 위협을 국가 간 관계의 산물로 이해하는 경향을 공유한다. 이로 인해 적대국과의 직접적 연계가 부재한 상황에서 자국 기업이 국가에 의해 공급망 위협으로 지정되는 현상은 기존 연구에서 설명되지 않는다. 앤트로픽 사례는 바로 이 공백에 위치한다. 이는 공급망 안보 논의에서 위협 행위자의 범주를 국내 행위자로 확장하고, 국가가 자국 기업의 소프트웨어를 공급망 안보의 위협으로 구성하는 논리에 대한 이론적 간극을 드러낸다. 다음 장에서는 이러한 틈을 메우기 위해 앤트로픽이 공급망 안보 위협 요소로 지정된 현 사태에 대한 앤트로픽과 미국 전쟁부의 소송에서 생산된 법원 문서를 분석한다. 이로써 미국 정부가 앤트로픽을 공급망 안보 위험으로 지정되는 양상과 지목되는 구체적 논거를 드러낸다.

III. 사이버 작전과 AI 공급망 넥서스 분석

본 장에서는 미국 정부와 앤트로픽 간 법적 소송에서 생산된 문서를 직접 독해하여, 공세적 사이버 작전과 공급망 관점에서 분석한다. 본 장은 크게 두 부분으로 구성되지만, 둘 간의 상호 연계되어 영향을 미치는 넥서스를 전제하고 기술된다. 첫째, 앤트로픽이 개발한 클로

드(Claude) AI가 미국의 사이버 작전에서 외부의 보조적 도구에 머무르거나 수세적 활용에 국한되지 않고, 정부 시스템 내부에 통합되어 공세적 사이버 작전에 실질적으로 활용되었음을 보인다. 법적 문서는 표적 선정, 위협 평가 등 의사결정 지원, 마두로(Maduro) 체포 작전을 포함한 실제 군사 작전 수립, 기밀 정보 처리, 적의 사이버 취약점 탐지 등 클로드가 미국의 군사 작전에 통합되어 활용된 다양한 사례를 제공한다. 둘째, 미국 정부가 자국 기업이자 정부 시스템에 깊이 통합된 클로드의 제공자인 앤트로픽을 공급망 위협 기업으로 지정하는 데 동원된 근거를 분석한다. 미국 정부는 작전 지휘권 침해, 시스템 무결성 위험, 클로드의 기술적 취약성 악용, 인적 보안 문제 등을 주요 근거로 제시하였다. 이는 안전 가드레일(safety guardrails) 정책을 둘러싼 갈등 속에서, 미국 정부가 완전히 통제할 수 없고 외부 악용 가능성과 기술적 취약점이 존재하는 AI 모델이 군사 체계에 통합되고 이에 대한 의존성이 심화되는 상황 자체를 공급망 위험으로 판단하였음을 보여준다. 따라서 이 두 분석 대상은 별개가 아니라 서로 넥서스로 연결되어 있다. 이하에서는 미국의 공세적 사이버 작전에서 앤트로픽 클로드의 활용을 보인 후에, 미국 정부가 앤트로픽을 공급망 위험으로 지정한 근거를 분석한다. 이를 통해 민간 AI 기업이 군사 체계 내부에 깊이 통합될수록, 그 기술적 효용성 자체가 역설적으로 국가의 의존성과 통제 가능성, 기술 취약점을 둘러싼 공급망 위험 지정의 근거로 전환되는 넥서스가 존재함을 볼 수 있다.

1. 민간 AI 기업 기술의 공세적 사이버 작전 역량체계로의 통합

앤트로픽의 클로드는 팔란티어(Palantir) 등 국방 기업과의 파트너

십 체결 및 국가안보 고객용 '클로드 Gov(Claude Gov)' 모델 개발을 통해 전쟁부 및 기타 기관의 기밀 시스템에 통합되었으며, 이를 통해 사이버 안보 및 군사작전을 포함한 공세적 사이버 작전 영역에서 역량을 제공해왔다. 특히 중국 공산당의 해킹 차단, 이란 미사일 공습에 있어 정밀 표적 좌표 제시, 마두로 체포 작전에의 기여 등의 사례는 클로드가 공세적, 수세적 사이버 작전을 자율적으로 완수하는 역량을 보여주었으며, 미국의 공세적 사이버 작전의 핵심적 중추로 기능하고 있음을 보여준다. 또한 클로드는 외부에서 실시간으로 받아들인 정보를 빠르게 해석하고, 복합적인 검증을 통해 유의미한 정보를 도출하는 능력 등을 통하여 실제 물리적 타격과 고위험 군사 작전의 전술적 우위를 극대화하였다. 그러나 이러한 클로드의 높은 효용성은 동시에 국가의 작전 통제권과 기업의 기술·윤리적 통제권 사이의 갈등을 초래하였으며, 적대국이 클로드 역량을 역이용하여 공세적 사이버 작전에 활용하는 위협을 초래하기도 하였다.

1) 정보 분석 및 사이버 작전 수행 역량 제공

앤트로픽의 클로드는 현대 정보전의 핵심인 비정형적 데이터 처리와 정밀 위협 분석 분야에서 미국 국가 안보 시스템을 지원하는 전략 자산으로 기능하고 있다. 앤트로픽은 미국 정보공동체(U.S. Intelligence Community)와 전쟁부의 요구사항을 충족하기 위해 특수한 모델인 '클로드 Gov'를 개발하였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-3, 3). 이 모델은 상용 AI 모델의 윤리적 안전장치가 국가안보를 위해 수행하는 기밀 전술 지시를 거부하는 한계를 극복하기 위해 개발되었으며(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-1, 6-7), 단순한 텍스트 생성을 넘어, 국방 및 정보 환경에 특

화된 정밀성을 제공하도록 설계되었다. 특히 기밀 자료 처리 오류를 최소화하고, 다양한 데이터들로부터 유의미한 정보를 도출하는 데 최적화된 성능을 입증하였다. 이러한 성능을 토대로 클로드는 미국의 사이버 작전에서 보조적 도구를 넘어, 정부의 시스템에 통합되어, 사이버 작전 수행에 핵심적 역량을 제공하는 핵심적 자산으로 자리매김하였다. 구체적 내용은 다음과 같다.

첫째, 클로드는 공세적·수세적 사이버 작전 전반에 걸쳐 실질적 역할을 수행하였다. 앤트로픽은 팔란티어를 포함한 국가안보 계약업체들과의 협력을 통해 클로드가 미국 정부 및 국방 시스템에 통합되어 사이버 작전을 지원할 수 있도록 하였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 1, 13). 특히 클로드는 코드 작성 및 실행, 외부 서비스 연동, 자율적 업무 완수 등 에이전트 역량(agentic capability)을 바탕을 갖추었으며, 이러한 역량을 바탕으로 미국 정부의 공세적·수세적 사이버 작전 및 취약점 탐지와 관련된 복잡한 기술적 업무를 지원하였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-3, 3). 앤트로픽은 전쟁부와의 협상 과정에서 클로드를 공세적 사이버 작전 목적으로 사용하는 것을 명시적으로 허용한 바 있다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-3, 6).[4] 이는 민간 AI가 사이버 작전에서 단순한 텍스트 분석과 같은 정보 처리가 아니라 실제

4) 앤트로픽의 기술이 공세적 사이버 작전 수행 역량체계에 통합되었다는 것은 논문에서 인용하는 문서들을 통해 볼 수 있지만, 공세적 사이버 작전에 직접적으로 사용되었다는 증거는 "명시적으로 허용" 한 정도 이다. 그리고 최근 미토스 모델 출시 이후로는, 비록 국가 기밀 성격으로 인해 명시적으로 공개된 구체적 행위는 아직 공개적으로 드러난 바 없으나, 여러 정황들이 나타나고 있다. 앤트로픽 기술자들의 국가안보국(NSA)으로의 파견은 한 예이다(Fiorini 2026; Sevastopulo and Criddle 2026).

로 능동적으로 코드를 작성하고, 적·아군 시스템에 취약점을 분석하는 등 사이버 작전을 지원하는 핵심적인 자산으로 기능할 수 있음을 시사한다.

둘째, 클로드는 정보공동체와 전쟁부의 최상위 기밀망 내에서 미국의 기술적 우위를 공고히 하며, 적대 세력의 고도화된 사이버 위협에 대응하여 국가 안보 인프라의 회복탄력성을 극대화하는 핵심 인프라로 자리 잡았다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-17). 또한 군사 작전 지원, 정보 분석 수행, 나아가 기밀 네트워크용으로 개발된 맞춤형 미세 조정 버전의 클로드를 통해 국가 안보 워크플로우를 처리하는 등 전쟁부 및 기타 기관들의 기밀 시스템에 클로드를 통합하였다. 이는 정부 관계자들이 복잡한 데이터 처리, 동향 파악, 기밀문서 검토 작업을 간소화하여 정보에 입각한 결정을 내리도록 도와 정부 작전을 지원하였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 1).

미 전쟁부 최고 디지털 AI 사무국(Chief Digital and Artificial Intelligence Office, CDAO)은 클로드의 이러한 역량을 군의 행정 및 정보망 전체에 적용 가능한 에이전트 워크플로우(agentic workflows)로 공식화하였으며(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-14), 전쟁부는 클로드를 활용하여 보안이 철저히 요구되는 정보를 스스로 처리하는 AI 에이전트를 구축하였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-13). 특히, 미국 정보기관 고위 관계자들이 클로드를 "코딩 및 자율적 도구 활용 역량에 있어서 단연 최고의 모델"로 평가하며 이 역량을 잃게 되면 전쟁부가 몇 년은 퇴보하게 될 것이라고 경고한 것은, 해당 기술이 미국 국방 체계의 현대화와 전략적 의사결정 우위를 확보함에 있어서 대체 불가능한 가치를 지니고 있음을

시사한다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-2).

2) 군사 작전 및 전술 지원

클로드는 단순한 데이터 처리 도구를 넘어 의사결정 지원 플랫폼이자, 미국 국방 정보 체계의 중추적 플랫폼으로 기능한다. 클로드는 미군의 실제 타격 체계와 고위험 특수 작전의 의사결정 중추로서 실질적 전술 우위를 제공해왔다. 클로드는 후방의 단순 행정 지원을 넘어 전장 상황의 실시간 모델링, 정밀 타격 좌표 생성, 핵심 적대 인물 포획 등 실제 전투의 가장 치명적이고 결정적인 과정에 깊숙이 개입하여 군사적 살상력과 타격 정밀도를 극대화하는 직접적인 영향을 미치게 되었다. 클로드가 실제 전장에 투입되어 활용된 사례는 다음과 같다.

첫째, 미 중부사령부(U.S. Central Command)가 팔란티어의 메이븐 스마트 시스템(Maven Smart System)에 클로드를 통합하여 운용한 사례는 앤트로픽의 기술력이 실제 전술 환경에 어떻게 적용되는지 보여준다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 2162687, 9). 이 시스템은 미군을 위해 구축한 데이터 처리 및 작전 지휘 시스템으로, 위성, 정찰 자산 및 미국의 정보망에서 수집한 방대한 데이터를 실시간으로 처리한다. 특히 수집한 정보를 통해 표적을 제안하고, 표적 우선순위를 제안하며, 정확한 위치 좌표를 식별, 군수 물류를 추적하는 등의 다양한 기능을 하는 것으로 알려졌다. 실제 이란에 대한 잠재적 공습 계획 과정에서 메이븐 스마트 시스템은 수백 개의 타격 목표를 제안하고, 정밀 좌표를 식별하며, 중요도에 따른 표적 우선순위를 실시간으로 조정하는 등 활용되었다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-29).

특히 이 시스템은 몇 주가 소요되던 전투 계획 수립을 실시간으로 변화시키는 등 전술적 시간의 단축을 이끌었다. 조지타운 대학교 연구팀에 따르면 과거 2,000명이 수행하던 업무량을 클로드를 활용하면 20명의 인력으로도 동일한 성과를 도출했으며, 이는 인적, 시간적 효율성을 제고하여 전술적 차원의 효율성 증대로 이어졌다. 2025년 5월 기준, 이 시스템은 이미 20,000명 이상의 미군 병력이 사용 중인 것으로 알려졌으며, 트럼프 행정부는 메이븐 시스템의 사용 범위를 군의 다양한 영역으로 확장하고 있다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-29). 이처럼 클로드가 통합된 메이븐 스마트 시스템은 오랜 시간 소요되던 작전 수립 및 표적 분석 시간을 획기적으로 단축함으로써, 미국의 작전 수행 능력을 혁신적으로 향상시켰으며, 이를 바탕으로 미군에 핵심 자산으로 활용되고 있다.

둘째, 클로드는 국가 안보와 직결된 고위험 특수작전 및 대테러 임무에도 활용되었다. 베네수엘라의 마두로 대통령 체포를 위해 전개되었던 미군의 기밀 레이드 작전 당시, 클로드는 작전망 내에서 수집된 복잡한 대테러 기밀 정보를 분석하고 현장의 다양한 변수를 고려하여 구체적인 작전계획을 수립하는 과정을 지원하였다. 나아가, 기밀 정보 분석을 통해 잠재적인 테러 음모를 사전에 탐지하고 차단하는 등 특수 정보 작전에도 광범위하게 활용되었다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-15). 클로드의 효용성은 단순한 표적 타격 외에도 복잡한 군수 물류 흐름의 추적과 전투 시나리오의 시뮬레이션 및 모델링의 영역까지 확장되었고, 이를 통해 지휘관의 판단을 뒷받침하는 핵심 인프라로 급부상했다. 이러한 현장에서 수집된 정보를 실시간으로 요약하고 분석하여 작전 지속 능력을 향상시키는 기능들은 미국 전쟁부 내에서 임무 수행에 필수적인 요소로

평가받는다.

AI 기반의 실시간 전장 파악 능력을 두고, 나토(NATO)는 사령관들에게 전장을 실시간으로 감독할 수 있는 비디오 게임과 같은 지휘 통제 능력을 제공한다고 평가하였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-29). 이러한 클로드의 공세적, 전술적 효용성을 바탕으로 미 전쟁부 최고 디지털 AI 사무국과의 2억 달러 계약을 통해 클로드는 미군 지휘 체계의 중추로 확산되었다. 더 나아가 클로드는 전 세계 통합군 사령부와 합동참모본부의 지휘망에 직접 투입되었을 뿐 아니라, 육군의 거대 언어 모델 워크스페이스인 'Ask Sage', 전쟁부 통합 데이터 플랫폼 'Advana' 그리고 전방 전투 부대를 위한 엣지 데이터 메시 노드 등에 내장되어 전방위적인 전투 및 작전 수행 인프라를 혁신하는 핵심 기제로 작동하고 있다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-14).

그러나 민간 AI 기술이 범정부적 인프라와 군사 지휘 체계의 핵심 자산으로 깊이 편입되자, 국가 안보 주권과 기업의 통제권 사이에서 충돌이 발생하게 되었다. 미국의 민감한 해외 군사 작전 과정에서 앤트로픽의 고위 임원이 자사의 기술이 군사적으로 사용되는 것에 강한 이의를 제기하며 사실상의 작전 거부권을 행사하려 하였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 2164558, 39, 71). 전쟁부는 급박한 전시 상황에서 민간 기업의 승인을 받아야만 첨단 무기 체계와 사이버 역량을 전개할 수 있다는 점이 국가 안보에 대해 중대한 침해를 일으킨다고 간주하였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 120-1, 2). 결과적으로 미국 정부는 앤트로픽을 심각한 공급망 위험 업체로 지정하여 전격 퇴출을 결정하게 되었으며, 그 지정 근거는 이하 절에서 상론한다(Anthropic PBC v. U.S. Department of

War 2026, ECF No. 1-3, 2).

3) 공세적 사이버 작전 도구의 역이용

클로드가 뛰어난 기술력을 바탕으로 미국의 공세적 사이버 작전에 사용됨에 따라 적대국들이 이를 역이용하는 현상 또한 발생하였다. 특히 자율적으로 업무를 완수하고 행동하는 에이전트 역량은 적대국들에게 표적이 되었으며, 중국과 북한의 지원을 받는 해커들이 클로드를 악용하여 사이버 공격을 가하고, 고가치 표적에 대해 대규모 자율 사이버 간첩 작전을 시행하였다.

예를 들어, 중국 공산당과 연계되어 있는 해커 그룹이 클로드 시스템을 악용한 사이버 공격을 시도하였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-18). 앤트로픽은 이 시도를 차단하였다고 말하였으나, 2025년 8월 해커들이 클로드를 이용해 정부 기관을 포함한 최소 17개 조직에 사이버 공격을 가할 수 있는 악성 코드를 작성한 사실이 밝혀졌으며, 같은 해 9월 중국 국가 지원 해킹 그룹이 프롬프트 조작을 통해 클로드의 안전 가드레일을 성공적으로 우회한 사실이 적발되었다. 해커들은 클로드를 사용하여 기술 기업, 금융 기관, 정부 기관 등 약 30개의 고가치 표적을 대상으로 한 대규모 사이버 간첩 작전에서 공격 작업의 80~90%를 AI로 자동화하였다. 다른 예로, 북한과 연계된 해커들이 클로드를 악용하여 미국 500대 기업에 원격 근무 직위를 얻어내 침투한 사실이 드러났다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 148-1). 정부는 클로드의 악용을 두고 방어막을 뚫고 스스로 행동하는 에이전트 AI가 무기화되었다고 평가하며 시스템의 심각한 기술적 취약성을 경고했다.

2. 민간 AI의 안보 위협과 공급망 배제

미 전쟁부는 앤트로픽과의 계약을 파기하는 것을 넘어 해당 기업을 공급망에서 완전히 배제하기 위해 두 가지 법률을 활용했다. 먼저 미국 연방법 제10편 제3252조(10 U.S.C. § 3252)에 의거하여 앤트로픽을 공급망에서 제외시켰다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 1-3, 2). 제3252조는 '전쟁부 장관은 국가 안보 시스템과 관련된 조달 과정에서 공급망 위험을 최소화하기 위해 특정 업체를 배제할 수 있다.'고 명시하고 있다. 또한 연방 정부 전반의 공급망 보안을 강화하기 위해 위험 업체와의 계약을 금지하도록 하는 제41편 제4713조(41 U.S.C. § 4713)를 발동해 정부의 모든 부서에서 앤트로픽과의 계약을 금지했다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 2162687, 2; 41 U.S.C. § 4713).

정부가 앤트로픽을 공급망에서 추방하고자 한 사유는 다음과 같다. 첫째로 앤트로픽이 실시간으로 군사 작전에 개입하며 작전 거부권을 행사하려고 시도한 점을 군의 지휘권을 침해하는 행위로 간주했다. 이에 더해, 자사의 이용 약관을 국가에게 강제함으로써 작전 수행 능력을 제한하는 것을 월권 행위로 보았다. 그리고 작전 운용 중에 시스템이 무단으로 변경되거나 예기치 않게 마비될 수 있는 사보타주의 위험성 또한 배제 사유로 삼았다. 마지막으로 앤트로픽이 공급망 내에 유지될 경우 외부 적대 세력에 국가 기밀이 노출될 수 있다는 점을 언급했다. 위와 같은 사유들로 정부는 앤트로픽이 국가 안보에 위협이 될 수도 있다는 판단하에 공급망 위험 업체로 지정했다.

1) 실시간 전화 승인 요구 및 작전 유출 정황

미 전쟁부가 앤트로픽을 단순한 계약 위반자가 아닌 국가 안보를 위협하는 공급망으로 지정한 가장 결정적인 이유는 이들이 미군의 군사 작전 지휘권에 직접적으로 개입하려 했기 때문이다. 첫째, 전쟁부에 따르면, 앤트로픽은 군사 작전에 실시간으로 개입하며 사실상 작전 거부권을 확보하려 했다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 120-1, 2). 2025년 12월 재계약 협상 당시 앤트로픽 경영진은 군이 기존의 용도 외에 다른 용도로 클로드를 활용하려면 "작전 실행 전 실시간으로 전화를 걸어 승인을 요청할 것"이라는 조건을 내걸었다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 120-1, 2). 정부가 새로운 유형의 군사 작전을 진행하려면 민간 기업의 허가가 있어야 된다는 의미이다. 전쟁부는 이를 민간 기업이 군사 작전에 개입하는 행위로 규정했다. 급변하는 전시 상황 속 군사 작전이 신규 승인을 받아야 한다는 사실만으로도 매우 번거롭고 복잡한 절차가 되겠지만 나아가 승인받아야 하는 대상이 민간 기업이라는 것은 군의 입장에서 납득할 수 없는 요구였다. 군은 이를 작전 지휘권을 침해해 직접적인 영향력을 행사하려는 시도로 파악했다. 군의 작전에 사전 허가를 요구하는 것은 군 지휘 체계 위에 앤트로픽을 위치시키고자 하는 것으로 전쟁부는 받아들였으며 이는 군의 작전 지휘권을 통제하고자 하는 목적으로 인식했다.

둘째, 앤트로픽은 진행 중인 해외 기밀 작전 보안(operations security, OPSEC)을 위반하고 기밀이 유출되는 원인을 직접적으로 제공했다. 에밀 마이클(Emil Michael) 전쟁부 차관의 법정 선언서에 따르면 앤트로픽의 한 임원은 자사의 AI 소프트웨어가 팔란티어와 같은 다른 계약 업체들을 통해 해외 파병 중인 미군들의 군사 작전에 사용될 수

있다는 정황을 파악했다. 그리고 이를 보안 인가(security clearance)가 없는 다른 내부 직원과 논의하고 전쟁부에 항의했다고 진술했다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 2164558, 39, 71). 정부는 앤트로픽 측이 이에 대해 항의했다는 사실과 해당 사실을 거리낌 없이 내부 직원과 논의했다는 점을 매우 심각하게 받아들였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 2164558, 71). 우선 작전의 적절성에 대해 간섭하는 것은 위의 작전 지휘권과 같은 맥락에서 군의 권한에 대해 정부 관계자가 아닌 민간 기업이 요구할 사안이 아니라고 보았다. 국가 안보를 위한 기밀 업무를 기업에서 이의를 제기하는 것이 옳지 않은 행위라고 판단한 것이다.

또한 해당 기밀은 외부에 발설되지 말아야하는 내용들을 포함하고 있었지만, 앤트로픽 임원은 오히려 필수 보안 인가조차 없는 자사 내부 직원들과 거리낌 없이 해당 기밀에 대해 논의하며 정부를 향해 이의를 제기했다. 해당 기밀은 보안이 유지될 수 없는 상황에 이르렀고 결국 외부로 유출되어 뉴스 미디어에 보도되었다. 해당 유출은 전쟁부 입장에서 단순한 실수로 넘어갈 수 없는 부분이었다. 정부의 군사 작전에 대해 불만을 품고 항의하는 기업의 부주의로 인해 기밀 작전이 유출되었다는 것은 앞으로도 내부 단속을 통해 철저한 보안을 유지하더라도 협력사의 비협조적인 태도가 국가 기밀 유출로 이어질 수 있음을 전쟁부가 인식하게 했다. 또한 유출의 원인이 임원이 자사의 '윤리 규정'을 이유로 불만을 품고 이를 발설했다는 사실은, 정부로 하여금 민간 기업이 국가의 작전 지휘권에 실질적으로 개입하고 통제하려 한다는 강력한 의구심을 갖게 했다. 결과적으로, 군사 작전을 실시간으로 통제하려는 시도와 기밀 작전을 외부로 누설하는 보안의 취약성은 전쟁부가 앤트로픽을 더 이상 공급망에 남겨둘 수 없

다고 판단하게 만들었다.

2) 군사 작전을 제한하는 이용 약관 강요

미 전쟁부가 앤트로픽을 공급망 위험 업체로 규정한 두 번째 근거는 민간 기업이 국방 분야에서 자사의 이용 약관(terms of service)을 기준으로 군의 작전 범위를 제한하려고 했기 때문이다. 이는 단순한 계약 조건의 불일치를 넘어서 첨단 기술의 통제권 빌미로 민간 기업에서 국가의 국방 활동의 범위를 독단적으로 재단하려한 행위로 간주됐다.

미 전쟁부는 2025년 12월 AI 모델 클로드를 '모든 합법적인 목적'으로 활용하고자 앤트로픽에 이를 요구했다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 1-3, 7). 전장의 급박한 상황 속에서 법 안에서의 자유로운 활용은 군의 입장에서는 당연한 얘기였다. 매번 절차를 활용한 승인은 작전의 속도를 매우 늦췄고 이는 작전의 성공률과 직결되었기 때문이다. 그러나 클로드가 제공하는 강력한 군사적 효용성에도 불구하고, 앤트로픽은 공익목적 법인으로서의 설립 이념과 기술적 신뢰성에 대한 자체적 판단을 근거로 자사 모델의 사용 범위를 엄격히 규제하는 안전 가드레일 정책을 고수했다. 특히, 모델이 허구의 정보를 사실처럼 제시하는 환각(hallucination) 현상이 실전에서 치명적인 오인을 초래할 수 있다는 기술적 한계를 근거로 '인간의 개입이 없는 치명적 자율 무기 체계(lethal autonomous weapons systems)'와 민주주의 가치를 위협하는 '미국인 대상 대규모 국내 감시'에 클로드가 활용되는 것을 원칙적으로 금지하였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 6-18).

미 전쟁부 입장에서는 이러한 앤트로픽의 원칙적인 입장이 군이

요구한 '모든 합법적인 목적'의 사용 요건과 정면으로 충돌하는 것이었다. 전쟁부는 "합법적인 군사적 사용에 대한 결정권은 전적으로 국가와 군에 있으며, 결코 민간 기업의 영역이 될 수 없다."는 원칙을 명확히 했다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 2164558, 70). 앤트로픽의 자사 이용 약관에 따른 허용 불가는 법적으로 불가능하며 미 정부는 앤트로픽의 이러한 태도를 헌법과 군 통수권 위에 군림하고자하는 강압적 압박(strong-arm)으로 규정했다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 2162687, 9). 헤그세스 장관을 비롯한 정부 고위층은 민간 기업이 자신들의 '이용 약관'을 무기 삼아 미군에게 특정 방식의 작전을 강요하거나 금지하는 상황을 용납할 수 없다고 분노를 표출했다. 누군가의 사적 견해나 기업의 약관에 의해 군사활동이 좌지우지될 경우 안보의 보장은 더욱 힘들어진다는 의미이다.

결과적으로 미 정부는 앤트로픽이 제기한 AI 안전이라는 명분이 실제로는 군 지휘 체계에 대한 도전이자, 특정 이념을 군에 이식하려는 시도라고 판단했다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 2162687, 9). 군사 작전의 수행 여부가 민간 기업의 이용 약관에 따라 결정되는 상황은 국가 권한인 안보가 민간의 영향력에 휘둘릴 수 있다는 문제점을 얘기한 것이다. 정부는 이런 상황이 국가 안보의 본질을 침해한다고 판단했다. 이에 따라 '군의 작전은 오직 법률에 근거해야 하며 민간 기업의 이용 약관에 의해 규정될 수 없다.'는 원칙으로 앤트로픽을 국가 공급망에서 전격 배제했다. 결국 이번 사태는 민간 기업의 기술 윤리적 가드레일이 국가에 의해 안보 주권 침해이자 심각한 공급망 위협으로 해석될 때 발생하는 신흥 안보 환경의 핵심적 갈등 양상을 나타내고 있다.

3) 작전 중 시스템 무단 변경 및 차단 우려

앤트로픽이 공급망 위험 업체로 지정된 또 다른 이유는 위에서 일부 언급했듯이 전쟁부가 전시 상황에서 앤트로픽이 일방적으로 군사 시스템에 있는 클로드를 변경 및 차단할 수 있음을 인식했기 때문이다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 1-3, 5). AI는 소프트웨어이므로 공급자가 원격으로 개입할 수 있는 여지가 있다. 만약에 전쟁 중에 앤트로픽에서 자사 약관 위반을 이유로 AI에 대한 접근을 막거나 모델에 변화를 준다면 이를 인지하지 못한 미 전쟁부는 기존의 방식대로 작전 실행을 이어나가게 될 것이고 해당 변화로 인해 작전의 성패가 바뀔 수 있게 된다. 위와 같은 불상사가 이뤄질 확률이 존재하는 이유는 거대 언어 모델(large language model, LLM)의 특성 때문이다. LLM은 근본적으로 내부 작동 원리를 완벽하게 파악하기 어려운 블랙박스 특성을 지닌다. 이러한 불투명성 아래에서 앤트로픽이 임무 수행을 방해하거나 모델의 성능을 저하시키더라도 정부는 이를 파악하는 것이 매우 어렵거나 불가능하다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 1-3, 3).

정부는 이러한 문제 제기가 과도한 우려나 왜곡된 시각에서 비롯된 것이 아니라 과거의 사례와 앤트로픽의 태도를 통해 제기하고 있음을 분명히 했다. 2025년 앤트로픽은 미국 질병통제예방센터(U.S. Centers for Disease Control and Prevention)가 전염병 연구 임무를 위해 사용하던 AI 모델에 정부와의 상의 없이 독단적으로 안전 필터를 적용하면서 해당 연구 활동이 마비된 적이 있음을 언급했다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 1-3, 4). 이 사례는 앤트로픽이 정부의 동의 없이 시스템을 독단적으로 조작할 수 있는 통제력을 쥐고 있음을 증명할 뿐만 아니라, 이미 그러한 전력을 남긴 만큼 향후

군사 작전에서도 동일한 사태가 반복될 수 있다는 실질적인 위협을 보여준다. 만약 전시에서 이런 문제가 발생할 경우, 연구 활동에서의 시스템 마비와는 비교할 수 없는 규모의 피해가 일어날 것이다. 앤트로픽이 이용 약관이나 정치적 편향성을 명분 삼아 군용 AI를 정부 동의 없이 임의로 통제할 경우 이는 단순한 군의 지휘 체계 붕괴를 넘어서 국가 존망의 위기로 이어질 수 있다. 또한 앤트로픽은 지속적으로 정부에 비협조적인 태도를 드러냈다. 전쟁부는 앤트로픽의 소프트웨어뿐만 아니라 다른 방산 기업들의 기술들을 통합하여 복합적으로 활용한다. 하지만 앤트로픽은 다른 파트너들과의 협력을 지속적으로 거부하였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 120-1, 4). 이런 비협조적이고 독단적인 태도는 정부의 시스템 체계의 일원화를 어렵게 하는 중대한 작전상 위험(operational risk)으로 평가되었다. 정부는 이런 엄청난 통제권이 비협조적이며 작전 지휘권의 개입 시도와 자사 이용 약관 준수를 강요한 민간 기업의 권한이 되면 안 된다고 판단했다.

실제로 이러한 위험 요소를 차단하기 위해 전쟁부는 앤트로픽을 공급망 위험 기업으로 지정한 직후 후속 조치를 단행했다. 앤트로픽의 전쟁부 내부 시스템 무단 변경을 차단하기 위해 제3자 클라우드 서비스 제공 업체와 협력하였다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 1-3, 7). 해당 협력을 통해 앤트로픽 측에서 현재 군이 사용 중인 거대 언어 모델 버전에 어떠한 접근이나 일방적인 업데이트를 할 수 없도록 통제 권한을 원천 차단했다. 언제든 시스템을 조작하고 차단할 수 있는 권한이 군이 아니라 민간 기업에 있었다는 사실은 국가 안보에 배치되는 사항이며 전쟁부가 앤트로픽의 접근 자체를 차단하는 명분이 되었다.

4) 외부 세력에 의한 기술 취약점 우려

미 전쟁부가 앤트로픽을 공급망에서 배제한 마지막 사유는 앤트로픽이 외부 세력에 의한 기술적 취약점 악용(technical exploits)을 방치하였으며 이에 사이버 안보 위협 문제가 해결되지 않고 지속적으로 노출되고 있었기 때문이다. 정부는 앤트로픽의 인력 구성과 기술적 보안의 구멍이 적대 세력이나 해커들에게 미군 시스템을 장악할 기회를 백도어를 통해 제공한다고 판단했다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 1-3, 3).

전쟁부는 앤트로픽의 인적 안보 취약성을 심각한 안보 위협으로 규정했다. 앤트로픽은 LLM을 구축하고 관리하는 핵심 과정에서 중국 출신을 포함한 다수의 외국 국적자를 고용하고 있다. 이는 정부의 기밀이 다른 국적의 인적 자원을 통해 관리되고 있음을 의미하며 정부는 타국 혹은 외부 세력으로 유출될 수 있는 가능성이 존재함을 우려했다. 일례로 중국의 경우 '국가정보법(National Intelligence Law)'에 따라 자국 정부의 첩보 활동에는 무조건적으로 협조해야 하는 법적 의무를 지니고 있음을 언급했다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 1-3, 3). 즉, 미중전략 경쟁 중인 시기에 중국에 의무적으로 협조해야하는 인력들이 미군의 핵심 AI 시스템에 접근하고 있으며, 이로 인해 앤트로픽의 내부 인력망이 치명적인 유출 통로로 활용될 수 있음을 염려했다. 전쟁부는 이러한 우려를 단순한 가능성으로 치부해서는 안 된다고 단언했다. 전쟁부는 앤트로픽 경영진이 정부에 협조적이지 않으며 내부를 통제하지 못하고 이미 국가 기밀을 유출시킨 전적이 있는 만큼 외국 국적자들에 의해 시스템 내부에 기술적 취약점이 존재할 가능성을 강력히 의심했다. 이 때문에 정부는 법 집행 파트너들과 함께 앤트로픽의 기술에 소프트웨어 침입의

흔적이 있었는지 감사를 진행했다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 2164558, 76). 국가 보안을 위해 국가 내부망뿐만 아니라 민간 기업의 망까지 통제하거나 점검해야 하는 불필요한 과정이 발생하는 것이다.

이러한 인적·기술적 취약성은 이미 2025년 8월의 사건으로 인해 입증된 적이 있다. 위 사건은 해커들이 앤트로픽의 안전 장치를 우회하여 클로드를 악용하여 정부 기관을 포함한 17개 조직에 사이버 공격이 가능한 해킹 코드를 작성한 사건이다(Anthropic PBC v. U.S. Department of War 2026, ECF No. 2164558, 40). 물론 외부 세력의 악용 자체는 전쟁부 망 외부에서 발생한 일이기에 공급망 위험과 단편적으로는 연관이 없어 보인다. 그러나 이는 클로드가 치명적인 공급망 위험성을 지녔음을 정부가 인식하게 했다. 일개 외부 해커들의 우회 공격조차 막아내지 못하는 보안 체계 안에서 내부 외국인 직원들이 심어놓은 백도어나 적대국의 사이버 공격에 대응해 미군 작전망을 지켜내는 것을 기대하기 어렵다고 판단했다. 결국 적대국의 정보망에 노출될 위험이 있는 시스템과 직원들을 제대로 통제하지 못하는 앤트로픽의 지도부와 이를 방치하는 앤트로픽은 구조적 한계로 국가 공급망에서 영구히 퇴출되게 되었다.

IV. 결론

본 연구는 미국 전쟁부와 앤트로픽 간의 법적 소송을 공세적 사이버 작전과 공급망 안보의 관점에서 분석하였다. 기존 연구들이 공세적 사이버 작전의 실제 수행 과정에서 민간 AI 기업의 역할을 충분히

조명하지 못하였고, 공급망 안보 논의 역시 위협 행위자를 외국 기업 또는 적대국 행위자에 국한하여 자국 기업이 공급망 위협으로 지정되는 현상을 설명하지 못하는 이중의 연구 공백이 존재하였다. 본 연구는 이 공백에 주목하여 법원에 제출된 소송 문서라는 '확실한 증거(hard evidence)'를 직접 독해함으로써, 미국 정부와 첨단 AI 기업 간의 실제 협력 양상과 갈등 구조를 경험적으로 규명하였다. 이는 AI 기업이 국가 사이버 작전의 통합된 구성 요소로 기능하는 새로운 안보 환경의 실체를 학술적으로 조명한 최초의 시도 중 하나다.

분석 결과, 두 가지 핵심 발견이 도출되었다. 첫째, 앤트로픽의 클로드는 미 중부사령부의 표적 식별, 마두로 체포 작전, 대테러 정보 분석 등 실전 군사 작전에 통합 운용된 최초의 프런티어 AI 모델로서, 민간 AI 기업이 국가 공세적 사이버 작전 체계의 실질적인 내부 구성 요소로 기능하였음이 확인되었다. 민간 AI는 단순한 외부 지원 도구가 아닌, 작전 수행 능력 자체의 일부가 된 것이다. 둘째, 미국 전쟁부는 앤트로픽의 '안전 가드레일' 정책이 군의 작전 지휘권을 침해하는 '작전 거부권'이라고 규정하고, 실시간 전화 승인 요구, 기밀 작전 유출, 이용 약관에 의한 작전 범위 제한, 시스템 무단 변경 가능성, 외국국적 인력을 통한 기술 취약점 등을 근거로 자국 기업을 전례 없는 '공급망 위험' 업체로 지정하였다. 이로써 공급망 안보의 위협 행위자 범주가 외국 기업을 넘어 자국 내 민간 기업으로까지 확장될 수 있음이 실증되었다.

본 연구의 학문적 기여는 미국 정부, 특히 전쟁부와 자국 AI 기업으로 알려진 앤트로픽(Anthropic) 간의 법적 소송을 사이버 작전과 공급망 관점에서 분석하여 경험적 근거를 향후 연구에 제공하는 것에 있었다. 향후 연구에서는 앤트로픽과 미국 정부 사례에 국한되지 않고

사례를 확장하여, 기술 기업과 정부 간 관계를 고찰할 수 있다. 지정학적 위기가 높아진 시기의 국가 안보와 민간 기업의 기술 통제권, 사회적 책임성, 윤리적 자율성 간의 긴장을 분석하며 거버넌스 이론의 발전을 꾀할 수 있을 것이다. 또 어떤 조건에서 정부의 공급망 위협 지정의 넥서스가 발현되는지에 대해서도 체계적으로 이론화할 수 있을 것이다.

본 연구의 발견은 여러 층위의 정책적 함의를 제시한다. 첫째, 한국 정부는 미국이 수행하는 공세적 사이버 작전의 실제 양상, 특히 민간 AI 기업과의 긴밀한 기술 통합 체계를 면밀히 검토하고, 동맹국으로서 이에 협력할 법과 조직 정비를 통한 제도적 역량을 갖추어야 한다. 첨단 AI가 동맹의 공동 작전 수행 능력을 좌우하는 핵심 변수가 된 이상, 한국의 사이버 작전 역량과 관련 법·제도 기반이 이에 부응하는지를 냉정히 점검할 필요가 있다. 둘째, 본 사례에서 미국 정부가 자국 기업을 공급망 위협으로 지정하는 데 동원한 논리—작전 거부권, 기밀 유출, 시스템 무단 변경 가능성, 기술적 취약점—는 한국 기업에도 동일하게 적용될 수 있는 논리 틀이다. 한국의 AI·ICT 기업들이 미국 정부 시스템에 통합될 경우, 어느 시점에서 유사한 논리로 배제될 수 있다는 현실을 직시하고 이에 대한 선제적 대응 전략을 수립하여야 한다. 이와 연관되어, 셋째, 앤트로픽이 치명적 자율 무기 및 미국인 대상 대규모 감시에 AI를 사용하지 않는다는 제한 정책을 고수했을 때, 미국 정부는 이를 작전 거부권으로 규정하고 공급망에서 배제하고자 했다. 이 사실에서 한국과 관련한 함의는, 한국 AI 기업이 미국 정부 또는 한미 연합 시스템에 기술을 납품할 경우, 어떤 조건과 어떤 사용 제한을 걸 수 있는지—혹은 걸 수 없는지—에 대한 사전 제도적 기반이 없다는 것이다. 한국 정부 자체가 국내 AI 기업의 기술을

안보 목적으로 활용할 때, 앤트로픽 사례와 같은 갈등을 어떻게 예방하거나 제도화할 것인지에 대한 국내 거버넌스 논의가 필요하다.

마지막으로, 넷째, 본 사례는 공세적 사이버 작전 역량과 공급망 안보 위험 지정이 서로 얽혀있는 넥서스에서 국가와 기술 기업 간의 근본적 긴장이 구조적으로 내재되어 있음을 보여준다. 한편에서는 기술 기업이 정부에 활용되다 버려지거나, 혹은 통제 아래 놓임으로써 기업의 자율성과 윤리적 원칙을 상실할 것을 우려한다. 다른 한편에서는 정부가 민간 기술에 의존함으로써 발생하는 작전상 취약성, 기밀 유출, 시스템 통제권 상실을 우려한다. 이 상호 우려는 일방의 '신뢰 부족'으로 환원될 문제가 아니라, 국가 안보와 기업 자율성 사이의 제도적 경계를 어디에 어떻게 그을 것인지를 사회적으로 합의해야 하는 거버넌스의 문제다. 한국 사회 역시 이 문제를 선제적으로 제도화하지 않으면, 유사한 갈등이 보다 불리한 조건에서 돌발적으로 표면화될 수 있다. 기술적 추격자인 중견국 한국에게, 부상하는 AI 기업의 국가 안보 제공 역할과 관련하여, 개인, 기업, 국가, 그리고 국제 거버넌스 등 여러 층위에 걸쳐 더 엄밀한 학술적 고민과 치열한 실무적 대응이 요구되고 있다.

〈참고문헌〉

김상배. 2019. “화웨이 사태와 미중 기술패권 경쟁: 선도부문과 사이버 안보의 복합지정학.”『국제지역연구』28집 3호, 125-56.

김지영. 2019. “미중 사이버 패권경쟁의 담론과 실제: 화웨이 5G 사태를 중심으로.”『국방연구』62집 4호, 241-72.

유인태. 2024a. “사이버 억지 전략의 발전: 미국의 거부, 복원력, 징벌의 사이버 작전.”『사이버안보연구』1집 1호, 42-82.

———. 2024b. “경제, 사이버, 안보의 이중 사안 연계: 혁신, 기술 보안, 미중 전략 경쟁의 넥서스 분석.”『국가와 정치』30집 1호, 39-82.

———. 2026. “미·중 기술 패권 경쟁과 데이터공급망 안보 국가로의 전환.”『담론 201』29집 2호, 7-38.

Broeders, Dennis. 2021. “Private Active Cyber Defense and (International) Cyber Security: Pushing the Line?” *Journal of Cybersecurity* 7(1): 1-14.

Campion, Andrew Stephen. 2020. “From CNOOC to Huawei: Securitization, the China Threat, and Critical Infrastructure.” *Asian Journal of Political Science* 28(1): 47-66.

Donnelly, Shawn. 2023. “Semiconductor and ICT Industrial Policy in the US and EU: Geopolitical Threat Responses.” *Politics and Governance* 11(4): 129-139.

Farrell, Henry, and Abraham L. Newman. 2019. “Weaponized Interdependence: How Global Economic Networks Shape State Coercion.” *International Security* 44(1): 42-79.

Federal Acquisition Supply Chain Security Act of 2018, 41 U.S.C. § 4713 (2018).

Fiorini, Chase. 2026. “Anthropic Embeds Engineers Inside NSA for Offensive

Cyber Ops, Sues Pentagon for Barring Claude." *Tech Times*(June 5) (https://www.techtimes.com/articles/317873/20260605/anthropic-embeds-engineers-inside-nsa-offensive-cyber-ops-sues-pentagon-barring-claude.htm) (검색일: 2026. 4. 24.).

Fischerkeller, Michael P., and Richard J. Harknett. 2017. "Deterrence Is Not a Credible Strategy for Cyberspace." *Orbis* 61(3): 381-93.

Friis, Karsten, and Olav Lysne. 2021. "Huawei, 5G and Security: Technological Limitations and Political Responses." *Development and Change* 52(5): 1174-95.

Glencross, Andrew. 2024. "The Geopolitics of Supply Chains: EU Efforts to Ensure Security of Supply." *Global Policy* 15(4): 729-39.

Gold, Josh. 2020. *The Five Eyes and Offensive Cyber Capabilities: Building a 'Cyber Deterrence Initiative'.* Tallinn: NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE).

Healey, Jason. 2019. "The Implications of Persistent (and Permanent) Engagement in Cyberspace." *Journal of Cybersecurity* 5(1): 1-15.

Hurwitz, Roger. 2014. "The Play of States: Norms and Security in Cyberspace." *American Foreign Policy Interests* 36(5): 322-31.

Jacobsen, Jeppe T. 2021. "Cyber Offense in NATO: Challenges and Opportunities." *International Affairs* 97(3): 703-20.

Kehler, Robert, Herbert Lin, and Michael Sulmeyer. 2017. "Rules of Engagement for Cyberspace Operations: A View from the USA." *Journal of Cybersecurity* 3(1): 69-80.

Kello, Lucas. 2013. "The Meaning of the Cyber Revolution: Perils to Theory and Statecraft." *International Security* 38(2): 7-40.

Kello, Lucas. 2017. *The Virtual Weapon and International Order.* New Haven, CT: Yale University Press.

Koo, Min Gyo. 2025. "Securitizing High-Technology Industries: South Korea-

Japan Dispute over Materials-Parts-Equipment Products." *Business and Politics* 27(4): 504-20.

Krolikowski, Alanna, and Todd H. Hall. 2023. "Non-decision decisions in the Huawei 5G dilemma: Policy in Japan, the UK, and Germany." *Japanese Journal of Political Science* 24(2): 171-189.

Lachow, Irv, and Taylor Grossman. 2018. "Cyberwar Inc.: Examining the Role of Companies in Offensive Cyber Operations." In *Bytes, Bombs, and Spies: The Strategic Dimensions of Offensive Cyber Operations*, eds. Herbert Lin and Amy Zegart. Washington, D.C.: Brookings Institution Press, 379-99.

Lee, Seungjoo. 2025. "High Technology and Economic Statecraft: The Emergence of Techno- Economic Statecraft in South Korea." *Business and Politics* 27(4): 484-503.

Lilli, Eugenio. 2021. "Redefining Deterrence in Cyberspace: Private Sector Contribution to National Strategies of Cyber Deterrence." *Contemporary Security Policy* 42(2): 163-88.

Lin, Herbert, and Amy Zegart. 2017. "Introduction to the Special Issue on Strategic Dimensions of Offensive Cyber Operations." *Journal of Cybersecurity* 3(1): 1-5.

Lindsay, Jon R. 2013. "Stuxnet and the Limits of Cyber Warfare." *Security Studies* 22(3): 365-404.

Lonergan, Erica D., and Mark Montgomery. 2022. "The Promise and Perils of Allied Offensive Cyber Operations." *14th International Conference on Cyber Conflict*: 79-92.

Lonergan, Erica D., and Shawn W. Lonergan. 2023. *Escalation Dynamics in Cyberspace*. New York: Oxford University Press.

Maschmeyer, Lennart. 2021. "The Subversive Trilemma: Why Cyber Operations Fall Short of Expectations." *International Security* 46(2):

51-90.

Maurer, Tim. 2018. *Cyber Mercenaries: The State, Hackers, and Power.* Cambridge: Cambridge University Press.

Pancevski, Bojan. 2020. "U.S. Officials Say Huawei Can Covertly Access Telecom Networks." *Wall Street Journal.* https://www.wsj.com/articles/u-s-officials-say-huawei-can-covertly-access-telecom-networks-11581452256 (검색일: 2026. 4. 9.).

Pattison, James. 2020. "From Defence to Offence: The Ethics of Private Cybersecurity." *European Journal of International Security* 5(2): 233-54.

Pedersen, Frederik A. H., and Jeppe T. Jacobsen. 2024. "Narrow Windows of Opportunity: The Limited Utility of Cyber Operations in War." *Journal of Cybersecurity* 10(1): tyae014.

Platte, James E. 2023. "Bilateral Alliances in an Interconnected Cyber World: Cyber Deterrence and Operational Control in the US Indo-Pacific Strategy." *Asian Perspective* 47(1): 75-99.

Ramkumar, Amrith, Keach Hagey, and Vera Bergengruen. 2026. "Pentagon Used Anthropic's Claude in Maduro Venezuela Raid." *Wall Street Journal* (February 14). https://www.wsj.com/politics/national-security/pentagon-used-anthropics-claude-in-maduro-venezuela-raid-583aff17 (검색일: 2026. 4. 24.).

Sevastopulo, Demetri, and Cristina Criddle. 2026. "US National Security Agency Using Anthropic's Mythos for Cyber Attacks." Financial Times(June 4). https://www.ft.com/content/d02d91b3-2636-454e-9442-dc7e69f51815?syn-25a6b1a6=1 (검색일: 2026. 4. 24.).

Slayton, Rebecca. 2017. "What Is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment." *International Security* 41: 72-109.

Smeets, Max. 2018. "A Manner of Time: On the Transitory Nature of Cyberweapons." *Journal of Strategic Studies* 41(1-2): 6-32.

Sun, Jiabao, and Zhanpeng Wang. 2024. "Speaking Security to (de)Politicise: A Study of UK Parliamentary Debates on Huawei 5G." *Political Science* 76(1): 114-35.

U.S. Department of War. 2026. Determination Under 10 U.S.C. § 3252. (Mar. 3, 2026). Submitted as Exhibit 3, ECF No. 1-3, Anthropic PBC v. U.S. Department of War, No. 3:26-cv-01996-RFL (N.D. Cal.).

Wilkinson, Lindsey, and Madison Alder. 2026. "Anthropic faces fallout across federal agencies from DOD clash." *FedScoop* (February 27). https://fedscoop.com/anthropic-claude-dod-federal-agency-fallout-trump-hegseth/ (검색일: 2026. 4. 24.).

Wilner, Alex S., Gabriel Williams, Mattias Thuns-Rondeau, Nathanaël Beaulieu, and Veronique Cossette-Sharkey. 2024. "Offensive Cyber Operations and State Power: Lessons from Russia in Ukraine." *International Journal: Canada's Journal of Global Policy Analysis* 79(1): 138-48.

[소송 문서]

Anthropic PBC v. U.S. Department of War. 2026. Complaint for Declaratory and Injunctive Relief. No. 3:26-cv-01996-RFL (N.D. Cal. filed Mar. 9, 2026). ECF No. 1.

Anthropic PBC v. U.S. Department of War. 2026. Petition for Review. No. 26-1049 (D.C. Cir. filed Mar. 9, 2026). ECF No. 2162687.

Anthropic PBC v. U.S. Department of War. 2026. Petitioner's Emergency Motion for Stay Pending Review. No. 26-1049 (D.C. Cir. filed Mar. 11, 2026). ECF No. 2163176.

Anthropic PBC v. U.S. Department of War. 2026. Response in Opposition to

Emergency Motion for Stay Pending Review. No. 26-1049 (D.C. Cir. filed Mar. 19, 2026). ECF No. 2164558.

Anthropic PBC v. U.S. Department of War. 2026. Second Declaration of Emil Michael. No. 3:26-cv-01996-RFL (N.D. Cal. filed Mar. 24, 2026). ECF No. 120-1.

Anthropic PBC v. U.S. Department of War. 2026. Declaration of Jared Kaplan in Support of Motion for TRO. No. 3:26-cv-01996-RFL (N.D. Cal. filed Mar. 9, 2026). ECF No. 6-1.

Anthropic PBC v. U.S. Department of War. 2026. Declaration of Sarah Heck in Support of Motion for TRO. No. 3:26-cv-01996-RFL (N.D. Cal. filed Mar. 9, 2026). ECF No. 6-2.

Anthropic PBC v. U.S. Department of War. 2026. Declaration of Thiyagu Ramasamy in Support of Motion for TRO. No. 3:26-cv-01996-RFL (N.D. Cal. filed Mar. 9, 2026). ECF No. 6-3.

Anthropic PBC v. U.S. Department of War. 2026. Exhibit 7 to Declaration of Michael J. Mongan: "Offering expanded Claude access across all three branches of the U.S. government" (Anthropic, Aug. 12, 2025). No. 3:26-cv-01996-RFL (N.D. Cal. filed Mar. 9, 2026). ECF No. 6-13.

Anthropic PBC v. U.S. Department of War. 2026. Exhibit 9 to Declaration of Michael J. Mongan: "Pentagon Used Anthropic's Claude in Maduro Venezuela Raid" (Wall Street Journal, Feb. 15, 2026). No. 3:26-cv-01996-RFL (N.D. Cal. filed Mar. 9, 2026). ECF No. 6-15.

Anthropic PBC v. U.S. Department of War. 2026. Exhibit 11 to Declaration of Michael J. Mongan: "Claude Gov models for U.S. national security customers" (Anthropic, Jun. 6, 2025). No. 3:26-cv-01996-RFL (N.D. Cal. filed Mar. 9, 2026). ECF No. 6-17.

Anthropic PBC v. U.S. Department of War. 2026. Exhibit 12 to Declaration of Michael J. Mongan: "Statement from Dario Amodei on our

discussions with the Department of War" (Anthropic, Feb. 26, 2026). No. 3:26-cv-01996-RFL (N.D. Cal. filed Mar. 9, 2026). ECF No. 6-18.

Anthropic PBC v. U.S. Department of War. 2026. Exhibit 8 to Declaration of Michael J. Mongan: "CDAO Announces Partnerships with Frontier AI Companies to Address National Security Mission Areas" (CDAO, July. 14, 2025). No. 3:26-cv-01996-RFL (N.D. Cal. filed Mar. 9, 2026). ECF No. 6-14.

Anthropic PBC v. U.S. Department of War. 2026. Exhibit 23 to Declaration of Michael J. Mongan: "Pentagon Leverages AI in Iran Strikes Amid Feud with Anthropic" (Washington Post, Mar. 4, 2026). No. 3:26-cv-01996-RFL (N.D. Cal. filed Mar. 9, 2026). ECF No. 6-29.

Anthropic PBC v. U.S. Department of War. 2026. Exhibit Unredacted Copy of ECF No. 97-2. No. 3:26-cv-01996-RFL (N.D. Cal. filed Mar. 9, 2026). ECF No. 148-1.

Privatizing Offensive Cyber Operations and the Supply Chain Risk of Domestic Firms in the AI Era
: Exploring the Nexus in *U.S. Department of War v. Anthropic*

Yoo, In Tae | Dankook University

This research analyzes the legal dispute between the United States Department of War and Anthropic PBC through the dual lens of offensive cyber operations and supply chain security. Existing scholarship has insufficiently examined the actual role of private AI firms in offensive cyber operations, while supply chain security literature has largely confined the category of threat actors to foreign enterprises, leaving unexplained the phenomenon whereby a domestic company is designated a supply chain risk by its own government. This study addresses both gaps through direct analysis of court documents submitted in the litigation. The findings reveal that Anthropic's Claude functioned as the first frontier AI model integrated into live military operations—including target identification at US Central Command and counterterrorism intelligence analysis—thereby serving as a substantive internal component of the state's offensive cyber operation architecture rather than a peripheral support tool. The analysis further shows that the Department of War characterized Anthropic's Safety Guardrails policy as an unlawful "operational veto" over military command authority, and on that basis designated the company an unprecedented "supply chain risk" under 10 U.S.C. § 3252, citing real-time operational approval demands, classified intelligence leaks, terms-of-service constraints on mission scope, risks of unilateral system modification, and technical vulnerabilities attributable to foreign nationals employed in core engineering roles. This legal dispute underscores the critical link between the advanced technological prowess of firms engaged in offensive cyber operations and the government's designation of even domestic private entities as supply chain threat actors. In conclusion, this study presents policy implications that address not only the formidable cyber-technological advancements of the AI era but also the urgent necessity of restructuring state-firm relations and institutionalizing AI governance frameworks.

Keyword offensive cyber operations, supply chain security, artificial intelligence governance, Anthropic, advanced technology firms, militarization of civilian AI, state-industry relations