

클로드 지피티 팔란티어와 2026년 세계전쟁

인공지능은 어떻게 전쟁의 방아쇠를 당기게 되었나

2026년 6월

김경진 변호사

김경진 변호사 출판사

프롤로그

화면에 이름 하나가 떠 있습니다.

그 앞에 한 사람이 앉아 있습니다. 가자지구를 내려다보는 어느 정보부대의 방, 책상 위 모니터에는 이름과 숫자가 함께 올라옵니다. 숫자는 1과 100 사이 어딘가입니다. 그 숫자가 높을수록 기계는 이 사람이 무장조직원일 가능성이 크다고 본 것입니다. 기계의 이름은 라벤더입니다.

장교는 이름을 봅니다. 숫자를 봅니다. 그리고 한 가지를 확인합니다. 이 이름이 남자의 것인가.

남자가 맞으면, 통과입니다.

여기까지 걸린 시간은 약 20초였다고, 그 일을 직접 한 정보장교가 이스라엘 매체 플러스972 매거진과 로컬콜의 기자 유발 아브라함에게 말했습니다. 20초. 커피 한 모금 식는 데도 그보다 오래 걸립니다. 신호등 한 번 바뀌는 시간입니다. 기계가 왜 그 이름을 골랐는지, 어떤 근거로 그 숫자를 매겼는지 들여다보는 시간은 거기에 들어 있지 않았습니다. 장교의 일은 사람이 맞는지, 그것도 남자가 맞는지를 확인하는 것이었습니다. 그 이상은 묻지 않아도 된다고 했습니다. 장교는 자신이 한 일을 이렇게 표현했습니다. 나는 기계의 결정에 도장을 찍는 고무도장이었다.

고무도장. 서류 위에 미리 새겨진 글자를 그대로 찍어 내리는 물건입니다. 무엇이 찍히는지 읽지 않아도 손은 움직입니다. 그 장교는 자기 손을 그렇게 보았습니다.

20초 뒤, 그 이름을 가진 사람의 집 좌표가 다른 시스템으로 넘어갑니다. 그 시스템의 이름은 '아빠 어딴어'입니다. 이 프로그램은 표적이 언제 집으로 들어가는지를 기다립니다. 가족이 모여 있는 저녁 시간, 그가 현관문을 열고 들어서는 순간 경보가 울립니다. 그때 폭탄이 떨어집니다. 한 정보장교는 이렇게 말했습니다. 우리는 그들이 군사 건물에 있을 때 죽이는 데에는 관심이 없었다. 집을 폭격하는 편이 훨씬 쉽다. 시스템은 그렇게 그들을 찾도록 만들어졌다.

기계가 표시한 명단 위에는 약 3만 7천 개의 이름이 있었습니다.

라벤더는 가자 주민 거의 전체를 훑었습니다. 백만 명이 넘는 사람들 하나하나에 1과 100 사이의 점수를 매겼습니다. 휴대전화를 자주 바꾸는가. 어떤 사람들과 통화하는가. 어느 단체 대화방에 들어가 있는가. 그런 작은 특징들을 모아 기계는 이 사람이 무장조직원일 확률을 계산해 냈습니다. 점수가 높으면 명단에 올랐습니다.

장교들은 그 명단의 10퍼센트가량은 무장조직원이 아니라는 사실을 알고 있었습니다. 무작위로 뽑아 하나씩 확인해 보았기 때문입니다. 어떤 사람은 하마스와 느슨하게 엮여 있었고, 어떤 사람은 아무 관계도 없었습니다. 같은 이름, 같은 별명을 쓰는 사람이 기계에 걸려들기도 했습니다. 통신 패턴이 비슷하다는 이유로 민방위 요원이나 경찰이 명단에 오르기도 했습니다. 옆에 하나는 엉뚱한 사람일 수 있다는 것을 알면서도, 장교들은 20초씩 도장을 찍어 나갔습니다.

그 폭격에는 미리 정해진 한도가 있었습니다. 군의 국제법 부서는 낮은 계급의 표적 하나를 잡을 때 민간인을 몇 명까지 함께 죽여도 되는지를 정해 두었습니다. 어떤 증언에서는 그 수가 열다섯이었고, 어떤 증언에서는 스물이었습니다. 표적이 하마스 조직원이기만 하면, 나이도 계급도 따지지 않았다고 했습니다. 한 정보장교는 그 시기에 비례의 원칙은 사실상 존재하지 않았다고 말했습니다. 비례의 원칙. 군사적으로 얻는 것과 민간인이 치르는 희생을 저울에 달아 보라는, 전쟁법의 오래된 약속입니다. 그 저울이 한동안 치워져 있었다는 뜻입니다.

이것이 이 책이 시작되는 자리입니다.

한 사람이 죽고 사는 일이 20초 안에 결정되었다는 것. 그리고 그 20초 동안, 사람은 기계가 한 일을 사실상 다시 묻지 않았다는 것. 묻지 않은 그 자리에 무엇이 들어섰는지, 아니 무엇이 사라졌는지. 이 책은 그 사라진 자리를 좇습니다. 저는 그것을 빈자리라고 부르겠습니다.

제 일을 잠시 말씀드려야겠습니다.

저는 오래 법을 다뤄 왔습니다. 누군가에게 책임을 묻는다는 것이 무엇인지, 그 일이 얼마나 무거운지 압니다. 책임을 묻는다는 것은 결국 한 가지 질문으로 돌아갑니다. 누가 했는가. 누가 그 일을 하기로 정했고, 누가 그 결과를 떠안는가. 법정에서 이 질문에 답이 나오지 않으면 재판은 한 발짝도 나아가지 못합니다.

그런데 가자의 그 방에서는, 이 질문에 답하기가 어려워졌습니다.

기계가 이름을 골랐습니다. 사람이 20초 동안 보았습니다. 폭탄이 떨어졌습니다. 가족이 죽었습니다. 그렇다면 누가 그 사람을 죽이기로 정한 것입니까. 라벤더입니까. 그것은 사람이 아니니 책임을 물을 수 없습니다. 도장을 찍은 장교입니까. 그는 기계를 믿었을 뿐이라고, 자신은 고무도장이었을 뿐이라고 말합니다. 라벤더를 만든 사람입니까. 그는 명단을 작성했을 뿐 방아쇠를 당기지 않았습니니다. 명단을 무기로 넘기라고 명령한 지휘부입니까. 그들은 개별 이름을 보지 않았습니니다.

방아쇠를 당긴 손가락은 분명히 있는데, 그 손가락에 책임을 물으려고 따라 올라가다 보면 어느 지점에서 길이 끊깁니다. 사람과 기계 사이 어딘가에서, 책임이 증발합니다. 가벼워지는 것이 아니라 아예 사라집니다. 그 자리에 아무도 서 있지 않습니다.

아브라함과 이야기를 나누는 정보장교들 가운데 몇몇은 자기가 한 일에 괴로워했습니다. 그들은 10월 7일 하마스의 공격에 충격을 받고 다시 군복을 입은 사람들이었습니다. 그러다 자신이 무엇을 하고 있는지 깨닫고 흔들렸습니다. 가족이 통째로 사라지는 집들을 만들어 내는 일. 그들 가운데 일부는 그 일을 정당화할 수 없다고 느꼈고, 그래서 기자에게 입을 열었습니다. 책임을 느낀 사람이 있었다는 뜻입니다. 다만 그 책임은 제도가 물은 것이 아니라, 개인이 혼자 짊어진 것이었습니다. 시스템은 누구에게도 책임을 묻지 않았습니니다.

그 빈자리가 이 책의 주제입니다.

인공지능이 전쟁의 죽음의 사슬 안으로 들어온 것은 역사에서 처음입니다.

죽음의 사슬이라는 말을 군대에서는 킬체인이라고 부릅니다. 표적을 찾아내고, 위치를 확정하고, 추적하고, 조준하고, 쏘고, 결과를 확인하는 일련의 단계입니다. 예전에는 이 사슬의 거의 모든 고리를 사람이 손으로 당겼습니니다. 표적 하나를 만들어 내려면 정보장교 여럿이 며칠을 들여다봐야 했습니다. 사람의 손은 느립니다. 그 느림이 때로는 한 생명을 살렸습니다. 한 번 더 확인하는 동안, 한 번 더 망설이는 동안, 폭탄이 늦게 떨어지는 동안 말입니다.

가자에서 그 느림이 사라졌습니다. 8200부대 AI센터를 이끌던 한 고위 장교는 군이 사람을 충분히 빨리 표적으로 만들어 내지 못한다고, 사람이 병목이라고 책에 썼습니니다.

인공지능이 그 병목을 풀어 줄 것이라고 했습니다. 라벤더가 그 일을 했습니다. 하루에 표적 수천 개를 쏟아 냈습니다. 사람의 속도로 며칠 걸리던 일이 기계의 속도로 몇 초가 되었습니다.

속도가 빨라진 만큼, 사람이 사슬에서 빠져나갔습니다.

이스라엘군은 그런 시스템은 없다고 했습니다. 자기들이 쓰는 것은 분석가를 돕는 도구일 뿐, 누가 무장조직원인지 인공지능이 가려내지는 않는다고 했습니다. 정보 시스템은 분석가의 도구일 뿐이라는 것이 군의 답변이었습니다. 아브라함은 그 해명을 자기 취재원들에게 그대로 읽어 주었습니다. 취재원들은 거짓말이라고 답했습니다. 아브라함은 한 가지를 덧붙였습니다. 8200부대 AI센터를 이끌던 장교가 2023년 텔아비브 대학에서 공개 강연을 했고, 거기서 군이 2021년에 인공지능으로 테러범을 찾아냈다고 직접 말했다는 것입니다. 발표 자료에는 기계가 사람들에게 점수를 매기는 화면까지 들어 있었다고 합니다. 한쪽은 그런 기계가 없다고 하고, 한쪽은 그 기계 앞에 20초씩 앉아 있었다고 합니다.

이 책은 양쪽 말을 모두 본문에 둡니다. 어느 쪽이 진실에 가까운지 판단하는 일은 잠시 미뤄 두겠습니다. 다만 한 가지는 분명합니다. 이런 시스템이 존재하는지조차 다투는 세상에서, 그 시스템이 한 일에 누가 책임지는지를 묻기란 더 어렵다는 것입니다. 없다고 하는 기계에는 책임을 물을 수 없습니다. 없다는 말 자체가, 빈자리를 한 겹 더 가립니다.

가자는 시작이었을 뿐입니다.

우크라이나의 들판에서는 드론이 사람의 얼굴을 익히는 법을 배웁니다. 전파 방해로 조종 신호가 끊겨도, 마지막 순간 표적을 알아보고 스스로 날아가 부딪치도록 만들어진 기계들입니다. 표적을 찾아 내려치기까지 18분이면 충분하다는 이야기가 나옵니다. 2026년 이란에서는, 한 사람을 죽이기로 정하는 데 60초가 걸렸다는 증언이 나왔습니다. 같은 해 베네수엘라 상공에서는, 학교 웹사이트가 멀쩡히 살아 있는 위성사진 위에서 표적이 정해졌습니다. 한 전장에서 다른 전장으로, 표적을 고르는 손이 사람에게서 기계로 넘어가고 있습니다. 20초가 60초가 되고, 60초가 다시 18분이 되는 동안, 사라진 빈자리는 메워지지 않고 함께 따라다닙니다.

아브라함도 같은 걱정을 했습니다. 2014년 전쟁, 2021년 전쟁을 겪으며 그는 한 가지 흐름을 보았습니다. 전쟁이 끝나면 거기서 쓰인 시스템이 세계 곳곳의 군대로 팔려 나간다는 것입니다. 그는 이 AI 기반 전쟁이 인류에게 위험하다고 말했습니다. 그것이 사람들을 책임에서 빠져나가게 한다고 했습니다. 기계가 표적 파일을 만들어 주면, 거기에는 사령관 이름도 적히고 예상 민간인 피해도 적힙니다. 국제법을 지키는 모양새는 갖춰집니다. 그런데 그 안의 말들이 모두 뜻을 잃는다고, 그는 말했습니다.

가자에서 처음 쓰인 그 시스템들을 두고, 한 분석가는 더 무서운 말을 남겼습니다. 거기서 작동한 것은 완성된 무기가 아니라 시험판이었다고. 사람을 죽이며 성능을 다듬는 시험판. 그는 그것을 죽음의 베타, 영어로 lethal beta라고 불렀습니다. 베타라는 말은 원래 정식 출시 전에 결함을 찾아내려고 돌려 보는 시험 버전을 뜻합니다. 새 앱을 정식으로 내놓기 전, 일부 사용자에게 먼저 써 보게 하며 문제를 잡아내는 그 단계 말입니다. 그 시험을 사람의 목숨으로 치렀다는 뜻입니다. 그 말이 무엇을 가리키는지는 이 책의 마지막에서 다시 마주하게 됩니다. 지금은 그 말 하나만 기억해 두시면 됩니다. 시험판은 팔립니다. 전쟁이 끝나면, 그 시스템은 세계 곳곳의 군대로 넘어갑니다. 가자에서 다듬어진 빈자리가, 다른 하늘 아래로 수출됩니다.

저는 이 책에서 답을 드리지 못할지도 모릅니다.

기계가 사람보다 빠리, 사람보다 정확하게 표적을 고르는 시대가 오고 있습니다. 그 기계를 만든 회사들 가운데에는, 자기 기술이 사람을 죽이는 데 쓰이는 선만은 넘지 않겠다고 선을 그은 곳도 있습니다. 같은 회사가 그 기술을 군대에 팝니다. 이 모순을 깔끔하게 봉합할 방법을 저는 아직 찾지 못했습니다. 찾은 척하지도 않겠습니다.

다만 한 장면은 끝까지 붙들겠습니다. 가자의 그 책상, 화면에 떠 있던 이름 하나, 그 앞에서 흘러간 20초.

그 20초 안에서 한 사람이 죽었습니다. 그리고 그 20초 안에서, 그를 죽이기로 정한 책임이 사라졌습니다.

누가 방아쇠를 당겼습니까.

이 책은 그 질문을 들고 가자에서 출발하겠습니다.

목차

프롤로그

제1부 기계의 속도

- 1장** 가자: 대량 암살 공장과 20초의 승인
1. 라벤더: 3만 7천의 명단과 10퍼센트의 오류
 2. 가스펠과 '아빠 어딿어': 표적이 된 집
 3. 거수기가 된 인간: 성별 확인 20초와 파이어 팩토리
 4. 통계로 치환된 생명: 15에서 100까지의 허용치
- 2장** 우크라이나: 프로젝트 메이븐의 실전 진화
1. 넘치는 데이터, 모자란 눈
 2. 하룻밤 사이 눈(雪)을 학습한 알고리즘
 3. 18분으로 줄어든 킬체인

제2부 통제선을 넘다

- 3장** 이란: 에픽 퓨리와 기계의 속도로 죽이기
1. 60초: 하메네이와 24시간 1,000개 표적
 2. 미나브 학교: 낡은 좌표가 부른 175명
 3. 실리콘밸리와 국방부의 충돌: 클라우드와 공급망 위험
- 4장** 베네수엘라: AI가 개입한 참수작전
1. 절대적 결의 작전: 마두로 포획의 밤
 2. 명분의 붕괴: 태양의 카르텔이라는 허구
 3. 균열: 앤스로픽과 팔란티어, 그리고 모순
 4. 모델인가, 손인가: 거대언어모델을 무기로 만드는 기술

제3부 빈자리

- 5장** 블랙박스 and 증발하는 책임
1. 자동화 편향: 기계를 믿게 되는 마음
 2. 블랙박스 딜레마: 설명할 수 없는 죽음
 3. 책임 공백: 아무도 방어쇠를 당기지 않았다
- 6장** 국제인도법은 기계를 멈출 수 있는가
1. 사라진 맥락: 행복의 손짓을 읽지 못하는 센서
 2. 의미 있는 인간의 통제
 3. 마르텐스 조항과 인류의 양심

에필로그

제 1 부

기계의 속도

제1장 가자: 대량 암살 공장과 20초의 승인

1. 라벤더: 3만 7천의 명단과 10퍼센트의 오류

모니터에 이름 하나가 떴습니다.

그 옆에는 1에서 100 사이의 숫자가 붙어 있었습니다. 점수가 높을수록 기계가 그 사람을 무장세력으로 본다는 뜻이었습니다. 정보장교는 이름과 점수를 봤습니다. 그리고 한 가지만 확인했습니다. 남자인가. 남자가 맞으면 그는 다음으로 넘어갔습니다. 표적 하나에 그가 쓴 시간은 약 20초였습니다.

화면 너머에는 가자가 있었습니다. 100만 명이 넘는 사람의 정보가 기계 안으로 들어가 점수가 되어 나왔습니다. 정보장교가 본 것은 그 점수였습니다. 사람이 아니라 숫자였습니다. 그는 그 숫자가 어떻게 계산됐는지 알지 못했습니다. 알 필요도 없었습니다. 그에게 주어진 일은 점수를 의심하는 것이 아니라 통과시키는 것이었습니다. 이 한 사람이 끝나면 다음 사람이 떴고, 그다음 사람이 떴습니다. 명단은 길었고, 기계는 지치지 않았습니다.

지치는 것은 사람뿐이었습니다.

이 장면을 세상에 처음 꺼낸 사람은 이스라엘 탐사기자 유발 아브라함입니다. 그는 2024년 4월, 이스라엘 진보 매체 플러스972 매거진과 로컬콜에 라벤더라는 인공지능 시스템에 관한 8천 단어짜리 보도를 실었습니다. 보도는 여섯 개의 단계로 나뉘어 있었고, 각 단계는 군이 표적을 찍어내는 고도로 자동화된 과정의 한 토막을 보여줬습니다. 그 보도는 라벤더 시스템에 직접 관여한 이스라엘 정보장교 여섯 명의 증언에 바탕을 두고 있었습니다. 미국 매체 데모크라시 나우와의 인터뷰에서, 아브라함은 그 20초의 의미를 이렇게 설명했습니다. "한 정보원은 표적 하나당 20초를 썼다고 말했습니다. 폭격을 승인하기 전에 말입니다." 그가 확인한 것은 표적의 죄가 아니었습니다. 성별이었습니다.

20초라는 시간을 잠깐 느껴보겠습니다. 신호등이 바뀌기를 기다리는 시간입니다. 엘리베이터가 한 층에서 다른 층으로 올라가는 시간입니다. 커피 한 모금을 식히는 시간입니다. 그 시간 안에 한 사람의 생사가 결재됐습니다. 그 사람의 가족까지 포함해서 말입니다. 시계의 초침이 스무 칸을 도는 동안, 화면 속 이름은 표적이 됐고, 그다음 칸에서는 다음 이름이 떠올랐습니다.

라벤더가 무엇을 하는 기계인지부터 보겠습니다.

플러스972 보도에 따르면, 라벤더는 이스라엘군이 하마스나 이슬라믹지하드의 하위 요원을 식별하려고 만든 인공지능입니다. 처음 설계 의도는 분명했습니다. 이스라엘은 하마스 요원이 3만에서 4만 명 사이일 것으로 추정했습니다. 사람 손으로 그 많은 사람을 일일이 가려내는 건 불가능했습니다. 그래서 기계에 맡기기로 했습니다. 아브라함의 표현을 빌리면, 라벤더는 가자 주민의 약 90퍼센트, 그러니까 100만 명이 넘는 사람들의 정보를 훑었습니다. 그리고 한 사람 한 사람에게 점수를 매겼습니다. 1에서 100까지. 그 점수는 기계가 보기에 그 사람이 하마스나 이슬라믹지하드 군사조직의 일원일 가능성을 나타냈습니다.

문제는 그 다음이었습니다.

2023년 10월 7일 이후, 보도에 따르면 이스라엘군은 결정을 하나 내렸습니다. 명단에 오른 수만 명을 이제 집 안에서 폭격할 수 있는 사람으로 간주하기로 한 것입니다. 그 사람만 죽이는 게 아니었습니다. 그 사람이 있는 건물 안의 모든 사람, 아이들, 가족이 함께였습니다. 아브라함은 이렇게 전했습니다. "한 정보원은 자기가 기계의 결정에 고무도장을 찍는 역할을 하고 있는 것 같았다고 말했습니다." 거수기. 기계가 정하고, 사람은 도장만 찍는 구조였습니다.

기계는 틀릴 수 있습니다. 이스라엘군도 그걸 알고 있었습니다.

아브라함의 보도에 따르면, 군은 표본을 무작위로 뽑아 하나하나 확인했습니다. 그 결과, 기계가 죽이라고 표시한 사람의 약 10퍼센트가 하마스 무장세력이 아니라는 사실을 알고 있었습니다. 어떤 이는 하마스나 느슨하게 얽혀 있었습니다. 어떤 이는 아무 관련이 없었습니다. 한 정보원은 그 오류가 어떻게 생기는지 설명했습니다. 기계가 하마스 요원과 똑같은 이름과 별명을 가진 사람을 끌어오는 식이었습니다. 통신 패턴이 비슷한 사람도 마찬가지였습니다. 가자의 민방위 요원일 수도, 경찰관일 수도 있었습니다. 그래도 기계는 점수를 매겼고, 사람은 도장을 찍었습니다.

10퍼센트라는 오류율을 어떻게 받아들여야 할지 따져보겠습니다. 평화로운 곳이라면 이 정도 오류는 일상에서 견딜 만합니다. 이메일 자동분류가 열에 하나를 틀려도, 우리는 휴지통을 뒤져 되살리면 됩니다. 추천 알고리즘이 열에 하나를 틀려도, 우리는 그 영화를 안

보면 그만입니다. 오류를 되돌릴 수 있기 때문입니다. 하지만 폭격은 되돌릴 수 없습니다. 집이 무너지고 가족이 죽은 다음에, 그 사람이 무장세력이 아니었음을 알게 되는 일은 정정이 아닙니다. 같은 10퍼센트라도, 결과가 죽음이면 그 숫자의 무게가 달라집니다. 군은 그 10퍼센트를 알고 있었고, 그것을 알고도 의미 있는 감독 없이 시스템을 몇 주 동안 돌렸다고 아브라함은 전했습니다.

숫자로 다시 보겠습니다.

명단에 오른 사람은 약 3만 7천 명이었습니다.

그중 약 10퍼센트가 무장세력이 아니었습니다.

그 10퍼센트는 사람으로 환산하면 수천 명입니다.

10퍼센트라는 말은 깔끔합니다. 통계는 늘 깔끔합니다. 하지만 그 안에는 이름이 같다는 이유로, 통신 기록이 비슷하다는 이유로 명단에 오른 사람들이 들어 있었습니다. 그들이 잠든 집 위로 폭탄이 떨어질 때, 기계는 오류를 인정하지 않았습니다. 기계는 점수만 출력했습니다.

아브라함이 들은 증언 가운데 하나는 표적의 신원이 어떻게 확정됐는지에 관한 것이었습니다. 라벤더가 점수를 매긴 사람과 그 사람의 집을 잇는 작업, 그리고 그 사람이 집에 들어왔는지를 확인하는 작업은 자동화돼 있었습니다. 정보장교가 직접 그 집을 들여다보고 표적이 안에 있는지 눈으로 확인하는 절차는 대부분 생략됐다고 그는 전했습니다. 시스템이 경보를 울리면, 그 경보가 곧 확인을 대신했습니다. 기계가 그렇다고 하면 그런 것이었습니다. 사람의 눈은 기계의 신호 뒤로 물러났습니다.

이스라엘군은 이 보도를 반박했습니다.

미국 매체 워싱턴타임스 등에 실린 이스라엘군의 입장은, 라벤더가 표적을 만들어내는 시스템이 아니라 '데이터베이스'라는 것이었습니다. 데모크라시 나우 인터뷰에서 아브라함이 읽어준 이스라엘군 공식 답변은 더 강했습니다. "주장과 달리, 이스라엘군은 테러 요원을 식별하거나 어떤 사람이 테러리스트인지 예측하려는 인공지능 시스템을 쓰지 않는다. 정보 시스템은 분석가를 위한 도구일 뿐이다."

아브라함은 이 답변을 정보원들에게 그대로 읽어줬다고 했습니다. 그들의 반응은 짧았습니다. 거짓말이다. 그는 인터뷰에서 이렇게 말했습니다. "보통은 이렇게까지 노골적으로 거짓을 말하지 않는데, 놀랐습니다." 그가 제시한 근거는 이스라엘군 내부에서 나온 것이었습니다. 8200부대 인공지능센터를 이끄는 고위 군 관계자가 2023년 텔아비브 대학교에서 공개 강연을 했고, 거기서 이스라엘군이 2021년에 테러리스트를 찾는 데 인공지능 시스템을 썼다고 직접 말했다는 것입니다. 강연 슬라이드에는 시스템이 사람들에게 점수를 매기는 화면이 들어 있었다고 아브라함은 전했습니다.

같은 기계를 두고 한쪽은 살상 명단이라 부르고, 다른 한쪽은 분석 도구라 부릅니다. 이 책은 그 둘 사이에서 판단을 미루지 않겠습니다. 다만 지금 기록해 둘 것은, 사람을 죽이는 결정의 길목에 기계가 들어섰다는 사실, 그리고 그 결정을 누가 내렸느냐는 물음에 양쪽이 서로 다른 답을 내놓고 있다는 사실입니다. 사람이 도장을 찍었으니 사람의 결정인가요. 기계가 명단을 짰으니 기계의 결정인가요. 책임은 그 사이 어딘가에서 흐릿해집니다. 이 흐릿함이 이 책이 좇는 빈자리입니다.

기계가 사람에게 점수를 매긴 방식도 들여다볼 만합니다.

아브라함은 라벤더가 작은 특징들의 목록에 따라 점수를 낸다고 전했습니다. 어떤 사람의 통신 습관, 휴대폰을 자주 바꾸는지, 어떤 단체 대화방에 속해 있는지 같은 신호들이 점수에 반영됐습니다. 이런 신호는 무장세력의 행동 패턴과 닮았을 수 있습니다. 동시에 평범한 사람의 일상과도 닮았습니다. 휴대폰을 자주 바꾸는 사람은 가난한 사람일 수도 있습니다. 같은 대화방에 있는 사람은 그냥 이웃일 수도 있습니다. 기계는 그 차이를 알지 못했습니다. 기계는 패턴을 봤고, 패턴에 점수를 붙였습니다. 그리고 그 점수가 한 사람의 생사를 갈랐습니다.

아브라함은 이 통계의 결과가 어떤 풍경을 만들었는지도 전했습니다. 그가 인용한 유엔 통계에 따르면, 전쟁 초기 사망자 6천여 명 가운데 절반이 넘는 사람이 비교적 적은 수의 가문에서 나왔습니다. 한 가족이 통째로 사라졌다는 뜻입니다. 표적 한 사람을 노린 폭격이 그 사람의 부모와 형제와 아이들을 함께 지웠습니다. 아브라함은 그것이 가족이라는 단위가 파괴된 표현이라고 했습니다. 명단 위에서는 한 줄이었던 이름이, 현장에서는 한 집안 전체였습니다.

20초로 돌아가 보겠습니다.

20초 안에 사람이 무엇을 할 수 있을까요. 이름을 읽고, 점수를 보고, 성별을 확인합니다. 그게 전부입니다. 기계가 왜 그 사람을 골랐는지 따져볼 시간은 없습니다. 통신 패턴을 다시 들여다볼 시간도, 같은 이름을 가진 다른 사람일 가능성을 의심할 시간도 없습니다. 아브라함의 보도에 따르면, 정보장교들은 표적이 남성이지만 확인되면 라벤더의 추천을 그대로 받아들이라는 지시를 받았습니다. 기계가 왜 그런 판단을 내렸는지 꼼꼼히 들여다보지 않고서 말입니다. 20초는 검토의 시간이 아니었습니다. 결재의 시간이었습니다.

이 책이 묻고자 하는 것을 여기서 분명히 해두겠습니다. 인공지능이 사람을 죽였는가, 라는 물음은 아닙니다. 폭탄을 떨어뜨린 것은 사람이 조종한 항공기였고, 승인 버튼을 누른 것도 사람이었습니다. 물음은 다른 데 있습니다. 사람이 결정을 내렸다고 말할 수 있는가. 점수를 기계가 내고, 명단을 기계가 짜고, 시각을 기계가 잡고, 사람은 20초 만에 도장을 찍었다면, 그 결정의 주인은 누구인가. 이 물음에 답이 없다는 것, 답할 사람이 사라졌다는 것, 그것이 빈자리입니다.

2. 가스펠과 '아빠 어딤어': 표적이 된 집

라벤더가 사람을 골랐다면, 그전에 건물을 고른 기계가 있었습니다.

이름은 가스펠입니다. 히브리어로 하브소라. 복음이라는 뜻입니다. 플러스972와 미국 공영라디오 NPR이 2023년 12월 보도한 바에 따르면, 가스펠은 이스라엘군 8200부대가 만든 시스템으로, 건물과 구조물을 표적으로 자동 생성했습니다. 사람이 아니라 장소를 찍는 기계였습니다.

이 기계가 얼마나 빠른지를 보여주는 비교가 있습니다.

전직 이스라엘군 참모총장 아비브 코하비는 인간 분석가 한 명이 1년에 표적 50개를 만든다고 했습니다.

가스펠은 하루에 100개를 만든다고 했습니다.

1년과 하루. 50개와 100개. 사람과 기계의 생산 속도 차이가 이 두 줄에 들어 있습니다. 플러스972 보도에 따르면, 2021년 가자 분쟁에서 이스라엘군이 정한 표적 약 1,500개 가운데 200개가량이 가스펠이 뽑아낸 것이었습니다.

전직 이스라엘군 인사 한 사람은 이 시스템을 한 단어로 불렀습니다. 대량 암살 공장. 영어로 매스 어새시네이션 팩토리. 공장이라는 말이 핵심입니다. 공장은 규모를 위해 만들어집니다. 빠르게, 많이, 멈추지 않고. 표적도 그렇게 찍혀 나왔습니다. 아브라함 라벤더 보도에 앞서 쓴 가스펠 보도의 제목이 바로 그 표현이었습니다.

공장이라는 비유를 잠시 붙들어 보겠습니다. 공장은 사람의 판단을 줄여서 생산량을 늘립니다. 한 사람이 처음부터 끝까지 만들던 일을, 여러 기계가 나눠 맡으면 속도가 빨라집니다. 대신 전체를 보는 사람이 사라집니다. 컨베이어 벨트 앞에 선 노동자는 자기 앞을 지나가는 한 부분만 봅니다. 가스펠 앞의 분석가도 그랬습니다. 자기 앞으로 흘러온 표적 하나를 처리하고, 다음으로 넘겼습니다. 그 표적이 어떤 사슬의 어느 고리에서 왔는지, 끝에서 누가 죽는지는 그의 화면 밖이었습니다. 책임이 흐려지는 첫 자리가 여기였습니다. 아무도 거짓말을 하지 않아도, 전체를 본 사람이 없으면 전체에 답할 사람도 없습니다.

가스펠이 건물을 찍고 라벤더가 사람을 찍었다면, 그 사람이 언제 죽을지를 정한 세 번째 기계가 있었습니다.

이름이 섬뜩합니다. '아빠 어딤어'. 영어로 웨어즈 대디.

기계에 이런 이름을 붙이는 일에 대해 잠깐 생각해 보겠습니다. 군사 시스템의 이름은 보통 위협적이거나 중립적입니다. 강철 돔, 화살, 다윗의 물매 같은 식입니다. 그런데 가족의 집에 든 표적을 잡아내는 이 시스템에는 아이의 물음이 붙었습니다. 이름을 지은 사람들은 그 기계가 무엇을 노리는지 알고 있었습니다. 집, 밤, 가족. 이름이 그것을 숨기지 않았습니다. 오히려 그 안의 농담 같은 가벼움이, 기계가 다루는 죽음과 너무 멀어서, 이 사슬 어딘가에서 사람의 감각이 마비됐음을 보여줍니다.

이 시스템이 무엇을 했는지 데모크라시 나우 진행자 에이미 굿맨은 보도 내용을 이렇게 요약했습니다. "웨어즈 대디라는 두 번째 인공지능 시스템은 살상 명단에 오른 팔레스타인 남성들을 추적했습니다. 이 시스템은 표적이 밤에 가족과 함께 집에 있을 때 이스라엘이 그들을 노릴 수 있도록 일부러 설계됐습니다." 일부러. 굿맨이 인용한 정보원의 말은 이 점을 분명히 합니다. "우리는 그들이 군사 건물에 있거나 군사 활동에 관여할 때만 죽이는 데에는 관심이 없었습니다. 오히려 이스라엘군은 망설임 없이 그들을 집에서 폭격했습니다. 첫 번째 선택지로 말입니다. 가족의 집을 폭격하는 게 훨씬 쉽거든요. 시스템은 이런 상황에서 그들을 찾도록 만들어졌습니다."

기계는 어떻게 사람과 집을 연결했는지를 아브라함은 감시 시스템의 작동 원리로 설명했습니다.

대량 감시 시스템에는 '연결'이라는 개념이 있습니다. 어떤 사람의 신원을 입력하면, 컴퓨터가 그 신원을 다른 정보에 빠르게 이어 붙입니다. 가자에서는 거의 모든 사람에게 집이 있었습니다. 그래서 시스템은 개인과 집을 자동으로 연결하도록 설계됐습니다. 아브라함의 설명에 따르면, 라벤더가 하위 무장세력으로 표시한 사람과 연결된 집들은 대부분 군사 활동이 벌어지는 곳이 아니었습니다. 그냥 가족이 사는 집이었습니다. 웨어즈 대디는 그 사람이 집에 들어가는 순간을 잡아 정보장교에게 경보를 보냈습니다. 표적이 가족 곁으로 돌아온 그 시각이 폭격의 시각이 됐습니다.

라벤더의 명단은 웨어즈 대대로 넘어갔습니다. 웨어즈 대대는 그들이 집에 들어가기를 기다렸습니다. 그리고 그 집이 폭격됐습니다. 아브라함은 이 연결을 담담하게 묘사했습니다. 군사 작전이 벌어지지 않는 가족의 집, 그 집에 표적이 들어서는 순간, 경보, 그리고 폭격.

여기에 또 하나의 선택이 겹쳤습니다. 어떤 폭탄을 쓰느냐였습니다.

아브라함은 미국 CNN이 2023년 12월 보도한 수치를 인용했습니다. 미국 정보 평가에 따르면, 이스라엘이 가자에 떨어뜨린 폭탄의 45퍼센트가 유도장치가 없는 이른바 멧덩구리 폭탄이었습니다. 정밀 유도탄과 달리, 무유도 폭탄은 건물 전체를 무너뜨립니다. 민간인 피해가 훨씬 큼니다. 그런데 정보원들은 하위 요원에게는 이 무유도 폭탄만 썼다고 말했습니다. 안에 있는 모든 사람과 함께 집을 통째로 무너뜨린 것입니다.

유도와 무유도의 차이를 잠깐 풀어보겠습니다. 유도폭탄은 떨어지는 동안 경로를 바꿔 한 지점을 정확히 칩니다. 건물의 한 층, 한 방을 노릴 수 있습니다. 무유도 폭탄은 그냥 떨어집니다. 어디에 닿을지를 정밀하게 조절하지 못합니다. 그래서 넓게, 깊게 부숩니다. 건물 하나를 통째로 주저앉힙니다. 표적 한 사람을 노리면서 무유도 폭탄을 쏜다는 것은, 그 사람 주변의 모든 사람을 함께 죽이기로 선택했다는 뜻에 가깝습니다. 정밀하게 칠 수 있는 폭탄이 있는데도 그러지 않은 것이라면, 그것은 능력의 문제가 아니라 선택의 문제입니다. 보도에 따르면, 그 선택의 이유는 비용이었습니다.

이유가 잔혹할 만큼 메마른 회계였습니다.

정보장교에게 왜냐고 물으면 이런 답이 돌아왔다고 아브라함은 전했습니다. 그 사람들은 중요하지 않다. 군사적으로 비싼 폭탄을 낭비할 만큼 중요하지 않다. 비싼 유도폭탄을 쓰면 건물의 특정 층만 노릴 수도 있었습니다. 표적이 있는 층만 무너뜨리고, 아래위의 다른 가족은 살릴 수도 있었다는 뜻입니다. 하지만 하위 요원에게는 그 돈을 쓰지 않았습니다. 대신 건물 전체를 무너뜨리는 폭탄을 썼습니다. 그 안의 모든 층, 모든 가족과 함께.

아브라함은 이 장면이 자기에게 깊은 인상을 남겼다고 했습니다. 그의 말을 옮기겠습니다. 집 한 채에 폭탄을 떨어뜨려 온 가족을 죽이는데, 정작 그 폭격으로 암살하려는 표적은 비싼 폭탄을 낭비할 만큼 중요하지 않다고 여긴다는 것. 그는 이것이 이스라엘군이 팔레스타인인의 생명 가치를 어떻게 재는지를 드물게 드러내는 장면이라고 봤습니다.

표적은 싹 폭탄을 쓸 만큼 하찮은데, 그 표적과 함께 죽는 가족은 그 싹 폭탄으로 통째로 묻혀도 좋을 만큼 더 하찮았습니다. 회계의 논리 안에서 사람의 무게가 그렇게 매겨졌습니다.

아브라함은 이 대목에서 모순을 정면으로 짚었습니다. 그의 말을 그대로 옮기겠습니다. "그 사람이 탄약을 낭비할 만큼 중요하지 않다면서, 어떻게 민간인 15명을 죽일 각오는 되어 있는 겁니까." 표적 한 사람은 비싼 폭탄을 쓸 가치가 없을 만큼 하찮은데, 그 한 사람을 죽이려고 가족 전체를 물어버리는 폭탄은 떨어뜨립니다. 이 모순이 봉합되지 않은 채 보도에 그대로 남았습니다.

아브라함은 이것이 비례성 원칙을 어떻게 다루는지를 보여준다고 봤습니다. 비례성. 국제인도법의 핵심 개념입니다. 예상되는 군사적 이득에 견주어 민간인 피해가 과도해서는 안 된다는 원칙입니다. 그런데 여기서도 표적의 군사적 가치를 폭탄값으로 환산했습니다. 그러면서 그 표적 하나당 민간인 열다섯을 죽일 수 있게 했습니다. 아브라함의 표현으로, 그것은 이스라엘군이 예상 군사적 이득에 견주어 팔레스타인인의 생명 가치를 어떻게 재는지를 드물게 드러낸 장면이었습니다.

웨어즈 대디라는 이름을 다시 입에 올려보겠습니다.

아빠 어딤어. 이 말은 보통 아이가 합니다. 저녁에 문이 열리기를 기다리던 아이가, 돌아온 아버지를 향해 던지는 물음입니다. 그런데 이 이름이 붙은 기계는 정반대로 움직였습니다. 아버지가 집에 돌아오는 그 순간을 잡아 폭격의 신호로 삼았습니다. 가족이 모이는 시각이 가족이 함께 죽는 시각이 됐습니다. 이름을 지은 사람이 무슨 생각이었는지 우리는 알 수 없습니다. 다만 그 이름이 가리키는 장면은 분명합니다. 집, 저녁, 돌아온 사람, 그리고 그 위로 떨어진 무유도 폭탄. 보도에 따르면 그렇게 죽은 사람들 가운데 여성과 아이가 많았습니다.

정보원들조차 충격을 받았다고 그는 전했습니다. 심리적으로 충격적이었다. 이 일을 직접 한 사람들이 그렇게 말했습니다. 아브라함은 그들이 왜 입을 열었는지도 전했습니다. 그들 중 일부는 10월 7일의 참극에 충격을 받아 다시 군에 소집된 사람들이었습니다. 자기가 다시 군복을 입게 되리라 생각하지 못했던 이들도 있었습니다. 그러다 자기가 무슨 일에 가담하고 있는지를 차츰 깨달았다고 했습니다. 가족을 죽이는 일, 정당화할 수 없는 일에 자기가 들어가 있다는 감각. 그들은 책임감을 느꼈고, 세상이 진실을 모르고 있다는

생각에서 말문을 열었다고 아브라함은 전했습니다. 군 대변인의 설명과 현장의 현실이 너무 달랐기 때문입니다.

3. 거수기가 된 인간: 성별 확인 20초와 파이어 팩토리

인간이 결정의 고리 안에 있다는 말이 있습니다.

영어로 휴먼 인 더 루프. 사람이 방아쇠를 당기는 마지막 자리에 끝까지 남아 있다는 뜻입니다. 기계가 아무리 빨라도 사람이 검토하고 사람이 승인한다는 약속입니다. 가자에서, 그 사람은 자리에 있었습니다. 다만 그가 한 일은 검토가 아니었습니다.

8200부대 인공지능센터를 이끈 인물이 2021년에 쓴 책이 있습니다. 제목은 '인간 기계 팀'. 인공지능과 인간의 시너지가 어떻게 세상을 바꿀 수 있는가, 라는 부제가 붙어 있었습니다. 아브라함이 인터뷰에서 소개한 그 책에서, 저자는 군대가 인공지능에 의존해 '표적 생성에서 인간이라는 병목을 해결해야 한다'고 썼습니다. 병목. 인간이 병목이라는 말입니다. 그는 정보장교를 아무리 많이 투입해도 전쟁 중에 하루에 충분한 표적을 만들어낼 수 없다고 했습니다. 그래서 기계가 필요하다는 논리였습니다.

이 한 단어에 군사 AI의 사고방식이 들어 있습니다. 병목. 공장에서 쓰는 말입니다. 생산 라인에서 가장 느린 공정이 전체 속도를 정합니다. 그 느린 공정을 병목이라 부르고, 공정을 개선하는 일은 병목을 없애는 일이 됩니다. 이 책의 저자는 표적 생성이라는 공정에서 가장 느린 부분이 인간이라고 봤습니다. 사람은 의심하느라 느립니다. 확인하느라 느립니다. 자느라, 쉬느라, 망설이느라 느립니다. 그 느낌이 전쟁에서는 결함으로 비칩니다. 그래서 그 느린 자리에 기계를 넣자는 것이었습니다. 그런데 그 느낌 안에 있던 것이 바로 판단이었습니다. 의심, 확인, 망설임은 비효율의 다른 이름이 아니라 책임의 다른 이름이었습니다. 병목을 없애자 책임도 함께 사라졌습니다.

흥미로운 것은, 같은 책에서 저자가 분명히 선을 그었다는 점입니다.

아브라함은 이 대목을 정확히 짚었습니다. 저자는 이 시스템이 인간의 판단을 대체해서는 안 된다고 또렷하게 썼습니다. 그는 그것을 인간과 인공지능 사이의 '상호 학습'이라 불렀습니다. 결과를 보고 결정을 내리는 건 정보장교라는 것이 책의 입장이었고, 이스라엘군의 공식 입장도 같았습니다.

설계도 위의 약속과 현장의 실제 사이에 틈이 벌어졌습니다.

아브라함이 여러 정보원에게서 들은 바에 따르면, 10월 7일 이후 적어도 이스라엘군 일부에서는 그 약속이 더 이상 지켜지지 않았습니니다. 정보장교들은 표적이 남성이라는 것만 확인되면 라벤더의 추천을 받아들여도 된다는 말을 들었습니다. 추천을 꼼꼼히 들여다보지 않고서, 기계가 왜 그런 결정을 내렸는지 확인하지 않고서 말입니다. 사람은 루프 안에 있었습니다. 하지만 그가 더한 것은 판단이 아니라 서명이었습니다.

한 정보원의 말이 이 상태를 그대로 압축합니다. 고무도장. 자기가 기계의 결정에 고무도장을 찍는 사람처럼 느껴졌다는 말. 사람이 결정을 내린다는 외형은 그대로였습니다. 모니터 앞에 사람이 앉아 있었고, 승인 버튼을 누른 것도 사람이었습니다. 하지만 그 사람이 한 일은 기계가 이미 내린 결정을 추인하는 것이었습니다. 20초 안에.

이스라엘군의 공식 입장은 다릅니다. 데모크라시 나우 인터뷰에서 아브라함이 읽어준 이스라엘군 답변은, 표적을 식별하는 과정에 여러 도구와 방법이 쓰이며 정보 관리 도구는 분석가를 돕는 데 쓰일 뿐이라는 것이었습니다. 어떤 사람이 테러리스트인지 식별하거나 예측하는 인공지능 시스템은 쓰지 않는다는 것입니다. 정보 시스템은 분석가를 위한 도구라는 말. 이 말과 고무도장이라는 증언 사이의 거리가 이 장의 긴장입니다. 도구라면 사람이 도구를 쓰는 것입니다. 고무도장이라면 사람이 도구에 끌려가는 것입니다. 같은 화면, 같은 버튼, 같은 20초를 두고 한쪽은 도구라 부르고 다른 한쪽은 거수기라 부릅니다. 이 책은 증언 쪽에 무게가 있다고 봅니다. 도구를 쓰는 사람은 도구를 의심할 수 있어야 하는데, 20초와 성별 확인만으로는 의심이 들어설 자리가 없기 때문입니다.

여기에 속도를 더 끌어올린 또 하나의 기계가 있었습니다.

이름은 파이어 팩토리입니다. 불 공장. 이 시스템은 무기를 자동으로 할당했습니다. 어떤 표적에 어떤 폭탄을, 어느 항공기로, 언제 떨어뜨릴지를 계산했습니다. 보도들에 따르면, 파이어 팩토리는 공격 준비에 걸리는 시간을 시간 단위에서 분 단위로 줄였습니다. 표적을 찾고, 무기를 짝지어 붙이고, 출격을 준비하는 전 과정이 압축됐습니다.

이 과정을 군사 용어로 킬체인이라 부릅니다. 죽음의 사슬. 표적을 찾고, 고정하고, 추적하고, 조준하고, 타격하고, 평가하는 여러 단계가 사슬처럼 이어진 것을 말합니다. 사람의 속도로 돌아가던 이 사슬에 기계가 들어서면서, 각 고리에 걸리던 시간이 무너졌습니다. 가스펠이 건물을 찍고, 라벤더가 사람을 찍고, 웨어즈 대디가 시각을 잡고, 파이어 팩토리가 무기를 붙였습니다. 사람은 그 사이 어딘가에서 20초를 썼습니다.

이 사슬에서 사람이 차지하는 자리가 점점 좁아졌다는 점을 짚어두겠습니다. 예전의 킬체인에서는 각 단계마다 사람이 있었습니다. 표적을 찾는 사람, 그게 맞는지 따지는 사람, 폭탄을 고르는 사람, 시각을 정하는 사람. 단계와 단계 사이에 시간이 있었고, 그 시간 안에서 사람은 멈추거나, 되묻거나, 거를 수 있었습니다. 기계가 단계들을 이어 붙이자 그 사이의 시간이 사라졌습니다. 멈출 자리도 함께 사라졌습니다. 사슬은 한 번에 처음부터 끝까지 흘렀고, 사람은 흐름의 한 지점에서 도장을 찍을 수 있을 뿐이었습니다. 휴먼 인 더 루프라는 말은 사람이 루프 안에 있다는 약속입니다. 하지만 루프가 사람보다 빠르게 돌면, 사람은 루프 안에 있어도 루프를 통제하지 못합니다. 안에 있는 것과 통제하는 것은 다릅니다.

빠른 게 왜 문제인지 물을 수 있습니다.

빠르다는 것 자체가 문제는 아닙니다. 문제는 빠른 속도가 검토의 자리를 지워버렸다는 데 있습니다. 사람이 기계의 판단을 들여다볼 시간이 사라지면, 사람은 판단하는 자리가 아니라 통과시키는 자리에 앉게 됩니다. 자동화 편향(automation bias)이라는 말이 있습니다. 기계가 내놓은 답을 사람이 무비판적으로 신뢰하는 경향입니다. 화면에 점수가 떠 있고, 옆 사람도 다 승인하고 있고, 하루에 처리해야 할 표적이 산처럼 쌓여 있을 때, 사람은 기계를 의심하기보다 따르게 됩니다. 20초는 그 편향이 자리 잡기에 충분한 시간이었고, 의심이 비집고 들어오기에는 너무 짧은 시간이었습니다.

자동화 편향이 어떻게 굳어지는지 한 장면으로 그려볼 수 있습니다.

상상해 보겠습니다. 한 정보장교가 책상에 앉아 있습니다. 화면에는 처리해야 할 표적 목록이 길게 떠 있습니다. 어제도 그랬고 그제도 그랬습니다. 처음 몇 건은 그도 점수의 근거를 들여다봤을지 모릅니다. 그런데 기계가 내놓은 답이 대체로 옆 사람들의 판단과 어긋나지 않습니다. 우선도 빨리 처리하라고 합니다. 그렇게 한 시간, 하루, 일주일이면, 점수를 의심하는 일 자체가 비효율이 됩니다. 기계를 믿는 쪽이 편하고 빠릅니다. 의심은 시간을 잡아먹고, 시간은 그에게 허락되지 않은 것이었습니다. 그렇게 사람은 판단하는 자리에서 통과시키는 자리로 미끄러집니다. 누가 시킨 것도 아닌데, 구조가 그를 그쪽으로 밀어냅니다.

그래서 책임은 어디로 갔을까요.

기계는 책임을 지지 않습니다. 기계는 점수를 출력할 뿐입니다. 정보장교는 자기가 고무도장이었다고 말합니다. 자기는 기계가 정한 것에 서명만 했을 뿐이라고. 명령을 내린 지휘부는 사람이 루프 안에서 최종 결정을 내렸다고 말합니다. 시스템을 만든 부대는 그것이 판단을 대체하는 기계가 아니라 도구일 뿐이라고 말합니다. 모두가 결정의 한 조각을 쥐고 있습니다. 누구도 결정 전체를 쥐고 있지 않습니다. 폭탄이 집 위로 떨어졌고 가족이 죽었는데, 그 죽음을 자기 결정이라고 말하는 사람이 사라집니다. 이것이 빈자리입니다. 사슬은 그대로 이어져 있는데, 책임이라는 고리만 빠져 있습니다. 빠진 게 아니라, 여러 손으로 잘게 나뉘어 아무 손에도 남지 않았다고 하는 편이 정확할지도 모릅니다.

4. 통계로 치환된 생명: 15에서 100까지의 허용치

부수피해라는 말이 있습니다.

영어로 콜라트럴 대미지. 군사 표적을 칠 때 함께 발생하는 민간인 피해를 가리키는 용어입니다. 이 말은 피해를 숫자로 바꿉니다. 사람을 허용치로 바꿉니다. 부수라는 말 자체가 그렇습니다. 곁가지, 덤, 본래 노린 것이 아닌 부산물. 죽은 아이가 부수피해가 되면, 그 죽음은 사건의 중심에서 가장자리로 밀려납니다. 언어가 먼저 사람을 변두리로 옮겨놓고, 그다음 숫자가 그 변두리에 한도를 매깁니다. 가자에서, 그 허용치에는 미리 정해진 숫자가 붙어 있었습니다.

아브라함의 보도에 따르면, 이스라엘군 국제법 부서는 정보장교들에게 표적별로 허용되는 부수피해 정도를 미리 알려줬습니다. 라벤더가 표시한 하위 표적 하나를 폭격할 때, 함께 죽어도 되는 민간인의 수가 정해져 있었다는 것입니다.

숫자를 계단처럼 놓아보겠습니다.

한 정보원은 그 수가 민간인 20명까지였다고 말했습니다. 하마스 요원 한 명당, 계급과 무관하게, 중요도와 무관하게, 나이와 무관하게.

다른 정보원은 그 한도가 15명까지였다고 말했습니다.

과거에는 그 수가 0이었습니다.

0에서 15로, 15에서 20으로. 하위 요원 한 사람의 목숨값에 붙은 민간인의 수가 그렇게 올라갔습니다. 한 정보원은 미성년자도 표시됐다고 말했습니다. 많지는 않았지만, 나이 제한이 없었기에 그럴 가능성이 있었다고 했습니다.

고위급으로 올라가면 숫자가 자릿수를 바꿉니다.

정보원들에 따르면, 하마스 고위 지휘관의 경우 허용치가 이스라엘군 역사상 처음으로 세 자릿수에 이르렀습니다. 100명 이상. 아브라함은 구체적인 사례를 들었습니다. 하마스 중부여단 지휘관이었던 한 인물을 노린 공격에서, 군은 그 한 사람과 함께 팔레스타인 민간인 300명을 죽이는 것을 승인했다고, 작전에 참여한 정보원이 증언했습니다.

300이라는 숫자에 사람을 다시 넣어보겠습니다.

300은 한 사람씩 세면 끝이 보이지 않는 수입니다. 한 사람에 5초씩 그 얼굴을 떠올린다 해도 25분이 걸립니다. 그런데 표적 파일에서 그 300은 한 칸에 적힌 숫자였습니다. 부수피해 칸. 폭격을 승인하는 정보장교의 화면에서, 그 300은 스크롤 한 번이면 지나가는 줄이었습니다. 사람을 숫자로 바꾸면 이런 일이 가능해집니다. 25분이 1초가 됩니다.

플러스972와 로컬콜은 그 공격을 목격한 팔레스타인 주민들과도 이야기를 나눴습니다. 그들은 그날 꽤 큰 주거용 건물 네 채가 폭격됐다고 말했습니다. 가족들로 가득 찬 아파트들이 통째로 폭격당해 죽었습니다. 아브라함은 이것이 실수가 아니었다고 못 박았습니다. 그 300명이라는 민간인 피해는 이스라엘군이 사전에 알고 있던 숫자였습니다. 정보원들이 그렇게 묘사했습니다. 한 정보원은 그 시기 초기에 국제법상 비례성 원칙이, 그들의 표현을 그대로 옮기면, "존재하지 않았다"고 말했습니다.

존재하지 않았다. 이 말은 무겁습니다. 한 정보장교가, 자기가 그 안에 있던 시스템을 두고, 국제법의 핵심 원칙이 작동하지 않았다고 증언한 것입니다. 바깥의 비판이 아니라 안에 있던 사람의 말입니다.

비례성 원칙은 폐지된 적이 없습니다. 국제인도법 안에 그대로 있었습니다. 종이 위에는 살아 있었습니다. 하지만 현장에서는 작동을 멈췄습니다. 표적 파일에는 여전히 '지휘관'이라는 칸과 '부수피해'라는 칸이 채워져 있었습니다. 형식은 갖춰져 있었습니다. 아브라함은 이 형식이 오히려 위험하다고 봤습니다. 그의 말을 옮기겠습니다. 기계가 표적 파일을 만들어주고, 거기에 지휘관이라든가 표적 부수피해 같은 칸이 채워지면, 국제법의 외형은 유지된다는 것입니다. 하지만 그 안의 의미는 모두 사라진다고 그는 말했습니다.

그가 던진 질문이 이 장 전체를 관통합니다.

아브라함의 말을 그대로 옮기겠습니다. "3만 7천 명을 표시하는 시스템을 설계하고, 그중 10퍼센트가 실제로는 무장세력이 아니라는 걸 확인해서 알고 있으면서도, 의미 있는 감독 없이 몇 주 동안 그 시스템을 승인해서 쓴다면, 그것은 구별 원칙을 위반하는 것 아닙니까. 군사적 관점에서 그다지 중요하지 않다고 본 표적을 위해 민간인 15명, 20명까지 죽이는 걸 승인한다면, 그것은 비례성 원칙을 명백히 위반하는 것 아닙니까."

구별. 민간인과 전투원을 구별해야 한다는 원칙입니다. 비례성. 군사적 이득에 견주어 민간인 피해가 과도해서는 안 된다는 원칙입니다. 이 두 원칙은 국제인도법의 기둥입니다. 아브라함은 인공지능 시스템이 이 용어들에서 의미를 빼내고 있다고 봤습니다. 그의 표현으로, 국제법은 지금 위기에 처해 있고, 이런 인공지능 시스템들이 그 위기를 더 깊게 만들고 있습니다.

이 장면을 가능하게 한 배경이 있습니다.

미국 워싱턴포스트가 2024년 12월 보도한 바에 따르면, 이스라엘은 10년 넘게 이른바 '인공지능 공장'을 구축해 왔고, 몇 주 만에 가자의 표적 약 1만 2천 곳을 폭격했습니다. 플러스972와 영국 가디언이 2025년 보도한 바에 따르면, 8200부대는 팔레스타인인의 통화 수백만 건을 미국 기업 마이크로소프트의 클라우드 애저에 저장했습니다. 그 양이 약 1만 1,500테라바이트에 이르렀습니다. 감시의 규모가 곧 표적의 규모였습니다. 더 많이 듣고, 더 많이 저장하면, 더 많은 사람에게 점수를 매길 수 있었습니다.

1만 1,500테라바이트라는 양을 가늠해 보겠습니다. 이것은 수백만 건의 통화를 담고도 남는 크기입니다. 한 사람의 목소리, 그 사람이 누구에게 전화했고 무슨 말을 했는지가, 가자 밖 어딘가의 서버에 차곡차곡 쌓였습니다. 그 통화의 주인들은 자기 목소리가 거기 저장되는 줄 몰랐습니다. 그리고 그 저장된 목소리들 가운데 일부가 라벤더의 점수가 됐습니다. 누구와 통화하는지, 얼마나 자주 휴대폰을 바꾸는지가 점수의 재료였으니까요. 감시는 표적의 앞 공정이었습니다. 클라우드에 쌓인 데이터가 기계로 흘러들고, 기계가 점수를 뱉고, 점수가 명단이 되고, 명단이 집 위로 떨어졌습니다. 미국 기업의 클라우드가 그 첫 칸에 있었다는 사실은, 이 사슬이 한 나라 안에서만 닫혀 있지 않다는 것을 보여줍니다. 이 점은 뒤의 장에서 다시 만나게 됩니다.

아브라함은 이스라엘이 이런 조사를 한다는 말에 회의적이었습니다. 그는 과거의 기록을 들었습니다. 그의 말을 옮기겠습니다. 2014년 가자 전쟁에서 팔레스타인 어린이 512명이 죽었고, 이스라엘은 조사하겠다고 했습니다. 전쟁범죄 주장이 수백 건 제기됐습니다. 그런데 이스라엘군이 실제로 군인을 기소한 단 한 건은, 약 1,000세켈을 약탈한 사건에 관한 것이었습니다. 나머지는 모두 종결됐습니다. 그는 조사가 진행 중이라는 말이 무언가 바뀌고 있다는 뜻은 아니라고 봤습니다. 그것은 우리의 지성을 조롱하는 것이라고 그는 말했습니다.

이 대목에서 이 책은 한쪽으로 기울 수밖에 없습니다.

양쪽 말을 다 적는 것과, 양쪽이 똑같이 옳다고 말하는 것은 다릅니다. 이스라엘군은 라벤더가 도구일 뿐이라고 했습니다. 그 입장은 본문에 그대로 적었습니다. 하지만 8200부대 인공지능센터장의 공개 강연, 그가 쓴 책, 여섯 정보원의 일치하는 증언, 유엔 통계가 보여주는 가족 단위의 몰살. 이 무게를 저울의 반대편에 놓으면, 저울은 한쪽으로 기울입니다. 도구라는 말만으로는 3만 7천 개의 이름과 한 사람당 열다섯의 허용치를 설명하지 못합니다. 이 책은 그 기울어짐을 숨기지 않겠습니다. 동시에, 기계를 만든 사람의 책에 적힌 경고, 인간의 판단을 대체해서는 안 된다던 그 문장도 지우지 않겠습니다. 설계자조차 그어둔 선이 현장에서 무너졌다는 사실, 그 모순이 이 장의 핵심이기 때문입니다.

아브라함이 인터뷰 끝에 한 말이 이 장의 마지막에 어울립니다.

그는 이 인공지능 기반 전쟁이 인류에게 위험하다고 봤습니다. 그의 말을 옮기겠습니다. 이런 시스템은 사람들이 책임에서 벗어나도록 해줍니다. 수천 명, 3만 7천 명을 잠재적 암살 대상으로 대량 생성하면서도, 국제법의 외형은 유지하게 해줍니다. 표적 파일을 만들어주는 기계가 있으니까요. 그리고 과거 전쟁의 경험으로 볼 때, 전쟁이 끝나면 이런 시스템들은 전 세계 군대에 팔려나간다고 그는 덧붙였습니다.

전쟁이 끝나면 팔린다는 이 말은 무겁게 새겨둘 만합니다.

가자는 시험장이 됐습니다. 라벤더, 가스펠, 웨어즈 대디, 파이어 팩토리가 거기서 처음으로 큰 규모로 돌았습니다. 시험장에서 검증된 시스템은 상품이 됩니다. 효과가 입증됐다는 말과 함께 다른 군대에 팔려나갑니다. 그런데 여기서 검증됐다는 것은 무엇이 검증됐다는 뜻일까요. 빠르게 표적을 찍어냈다는 것입니다. 한 사람의 검토를 20초로 줄였다는 것입니다. 책임을 여러 손에 흠어 아무 손에도 남지 않게 했다는 것입니다. 시험장에서 통과된 것이 정밀함이 아니라 이 빈자리라면, 팔려나가는 것도 빈자리입니다. 이 책이 가자에서 멈추지 않는 이유가 여기 있습니다. 가자에서 본 구조는 한 분쟁의 특수한 사정이 아니라, 앞으로 여러 전장에서 반복될 형편이기 때문입니다.

이 마지막 문장이 다음 장으로 이어집니다.

가자에서 만들어진 것은 가자에만 머물지 않습니다. 표적을 사람의 속도에서 기계의 속도로 넘긴 이 방식은, 우크라이나의 들판으로, 그 너머의 전장으로 흘러갑니다. 명단에 오른 3만 7천 명, 그중 무장세력이 아니었던 10퍼센트, 한 사람당 정해진 15명에서 100명까지의 허용치. 이 숫자들 사이에서 사라진 것은 결국 한 가지였습니다. 그 죽음을 자기 것이라고 말할 사람. 빈자리는 거기에 있습니다.

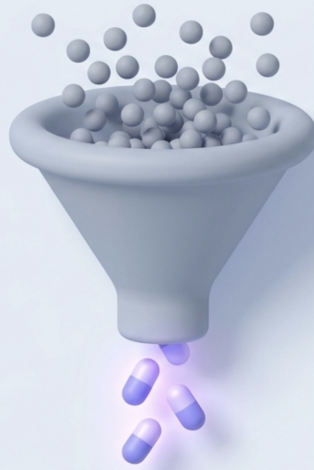
제1장 가자: 대량 암살 공장과 20초의 승인

인간의 판단을 배제한 채 3만 7천 명의 표적을 기계적 속도로 찍어내는 시험장이 되다.

책임의 증발

2014년 분쟁 당시 어린이 512명
사망 대비 이스라엘군의 기소는
1,000세켈 약탈 건 단 1건에 그침.

라벤더(Lavender) 시스템은
한 사람당 15~100명의 부수적
피해 허용치를 두고 표적 파일을
자동 생성함.



기계적 효율

인간의 검토 시간은 20초로
단축되었으며, 책임을 여러 손에
흩어 아무도 죽음에 책임지지 않는
구조를 완성함.

© NotebookLM

제2장 우크라이나: 프로젝트 메이븐의 실전 진화

1. 넘치는 데이터, 모자란 눈

화면은 멈추지 않았습니다.

드론 한 대가 한 시간을 날면 한 시간 분량의 영상이 쌓입니다. 그런 드론이 수십 대, 어떤 날은 수백 대. 분석관 한 사람의 책상에는 그렇게 모인 화면이 끝없이 흘러내렸습니다. 흠길, 트럭, 염소 떼, 다시 흠길. 누군가는 그 안에서 사제폭탄을 묻는 손, 무기를 옮기는 어깨, 길목에 멈춰 선 차 한 대를 골라내야 했습니다. 그런데 사람의 눈은 그렇게 오래 한 점을 지킬 수 없습니다. 그 사실을, 미군은 비싼 값을 치르고 알았습니다.

아프가니스탄과 이라크의 대테러전에서 미군이 띄운 무인기는 매시간 어마어마한 양의 영상을 모았습니다. 문제는 그다음이었습니다. 그 영상을 들여다볼 사람이 없었습니다. 정확히 말하면, 사람은 있었지만 영상의 양을 사람이 따라잡지 못했습니다. 카트리나 맨슨은 자신의 책 '프로젝트 메이븐'을 소개하는 한 대담에서 그 출발점을 이렇게 정리했습니다. 미국은 "감당할 수 없을 만큼 많은 드론 영상에 짓눌려" 있었고, "인간 분석관이 데이터에 파묻혀 도무지 찾아낼 수 없는 것"을 시로 찾아내려 했다는 것입니다.

여기서 한 사람을 만나야 합니다. 드루 쿠코르. 미 해병대 대령. 2001년 10월, 9·11 두 달 뒤 첫 원정 해병의 일원으로 아프가니스탄에 들어간 정보장교였습니다. 맨슨의 취재에 따르면, 그는 무거운 컴퓨터를 짊어진 동료 데이브 스파크와 함께 초기 작전의 정보 세포를 맡았습니다. 그리고 그는 화가 나 있었습니다. 무엇에 화가 났는가. 정보가 전선까지 닿지 않는다는 사실에.

쿠코르가 본 전쟁은 이런 모습이었습니다. 어딘가에 정보는 있습니다. 탈레반의 옛 은신처, 과거 공격이 일어난 길목, 누가 어떤 폭탄을 만드는지. 그런데 그 정보가 정작 총을 든 사람에게는 가닿지 않았습니다. 정보가 없어 병사들이 죽었습니다. 사제폭탄이 어디에 묻히는지, 얼마나 자주 터지는지, 어떤 날씨에 터지는지를 기록하고 싶었지만, 미군이 쓰던 도구는 파워포인트와 엑셀이었습니다. 맨슨은 쿠코르가 정보 공백 속에서 일한 사람이었다고 말합니다. 그는 그저 데이터를 원했습니다.

그러니 메이븐의 씨앗은 거창한 미래 구상이 아니라 한 정보장교의 분노에서 자랐습니다. 정보가 사람을 살리지 못한다는 분노. 쿠코르는 2011년 팔란티어(Palantir)의 데이터 통합 시스템을 해병대에 들여오는 일에 매달렸고, 그 시스템이 "거의 매일 밤 생명을 구했다"고

맨슨에게 말했습니다. 그가 보기에 기존 시스템은 조잡했습니다. 데이터 연결이 끊겨 있고, 접근 권한이 없고, 양식이 제각각이었습니다.

2017년, 그 분노가 하나의 프로젝트로 모였습니다.

미 국방부는 그해 알고리즘 전쟁 교차기능팀, 영어로 Algorithmic Warfare Cross-Functional Team을 만들었습니다. 이름은 프로젝트 메이븐(Project Maven). 창문 없는 펜타곤 방에 작은 팀이 모였습니다. 공개된 설명은 좁았습니다. 드론 영상이 너무 많으니, AI로 그 안에서 차량과 사람 같은 객체를 자동으로 골라내겠다는 것. 초기 표적은 이슬람국가, IS였습니다. 맨슨은 그 방에 있던 사람 여럿이 자신의 북토크 객석에 앉아 있다고 청중에게 말했습니다. 그들은 거기 있었습니다.

그러나 맨슨이 200명 넘는 관계자를 취재하며 알아낸 것은, 야망이 처음부터 그 좁은 설명보다 훨씬 멀리 가 있었다는 사실입니다. 그 야망은 하나의 두려움 위에 놓여 있었습니다. 미국이 중국에 뒤처지고 있다는 두려움. 당시 국방 예산은 미국이 압도적으로 컸지만, 중국이 미국의 약점을 10년 동안 연구해 왔다는 인식이 펜타곤을 사로잡고 있었습니다. 당시 국방부 부장관 밥 워크는 자율, 곧 인간을 전장에서 빼내고 기계가 결정하게 하는 길이 미국의 승부수라고 믿었습니다. 그리고 자율로 가는 길은 AI였습니다.

쿠코르의 표현으로는 '흰 점'이었습니다. 지도 위의 흰 점. 클릭하면 좌표가 되고, 좌표가 되면 그곳에 무기를 보낼 수 있는 점. 그는 여러 해 동안 글로 이 구상을 다듬어 왔습니다. 영상 속에서 무엇이 있는지 골라내는 일은 시작일 뿐이고, 그가 그린 그림은 정보를 전선까지, 총을 든 사람의 손까지 끌어내리는 것이었습니다.

문제가 하나 있었습니다. '표적'이라는 단어였습니다.

맨슨이 책에 적은 한 일화가 그 긴장을 보여줍니다. 어느 작전 지역에서 분석관들이 한 농부를 두고 씨름한 적이 있었습니다. 미군이 막 타격하려는 순간, 들판으로 농부 한 사람이 걸어 들어왔습니다. 인간 분석관이 그 농부를 화면에서 찾아내는 데 40초가 걸렸습니다. 그런데 같은 영상을 AI에 다시 돌리자, AI는 그를 즉시 찾아냈습니다. 그러니 이것은 AI를 옹호할 더없이 좋은 사례였습니다. AI가 생명을 구할 수 있다는. 그 현장의 누군가가 본부에 써 보냈습니다. AI가 표적 선정에, 그리고 생명을 지키는 데 쓸모가 있겠다고.

돌아온 답은 이랬습니다. "표적이라는 말을 쓰지 마라. 그 단어를 입에 올리지 마라. 의회가 이걸 들여다보고 있다. 펜타곤 안에도 반대가 있다."

그래서 '표적'이라는 단어는 오랫동안 메이븐의 공식 언어에서 지워졌습니다. 메이븐은 정보·연구개발 프로젝트였지 작전이 아니었으니까. 그러나 적어도 일부 사람들의 머릿속에서 메이븐은 언제나 정보를 작전으로, AI를 살상의 사슬로 잇는 일이었습니다. 쿠코르는 미군이 이 기술적·문화적 변화를 통과하는 데 20년이 걸리리라 보았습니다. 그가 보기에 정작 어려운 건 기술이 아니라 문화였습니다.

2018년, 그 조용한 진행이 한 번 크게 흔들립니다.

구글 직원들이 자기 회사 기술이 펜타곤의 프로젝트 메이븐에 쓰이고 있다는 사실을 알게 되었습니다. 맨슨의 대담 표현으로는 약 3천 명, 다른 보도로는 약 4천 명의 직원이 항의 서한에 이름을 올렸습니다. "우리는 전쟁 사업에 발을 담그고 싶지 않다." 십수 명이 회사를 떠났습니다. 인권 운동가들의 우려는 한 발 더 나아갔습니다. 이 단계에서 AI를 들이는 것이 결국 자율 살상무기로 가는 길을 닦는다는 것. 메이븐이 내세운 설명과 다른 길로.

구글은 계약을 갱신하지 않기로 했습니다. 당시에는 미군 AI 노력에 치명타로 비쳤습니다. 그러나 맨슨이 추적한 결과는 달랐습니다. 쿠코르에게 물으면 "다 괜찮았다"는 답이 돌아옵니다. 구글이 빠진 자리로 다른 회사들이 몰려들었다는 것입니다. 마이크로소프트, 아마존의 AWS, 그리고 클래리파이라는 작은 스타트업. 클래리파이를 이끈 맷 자일러는 2012년 이미지넷 대회 우승팀의 일원이었습니다. 그가 그때 돈을 벌던 방식은 결혼식 블로그용 컴퓨터 비전이었습니다. 신부의 면사포, 신랑의 정장, 웨딩 케이크의 층을 알아보는 알고리즘. 그것을 전장의 트럭과 탱크로 돌려쓰는 일에 그의 직원 일부는 불편해했고, 일부는 떠났습니다.

구글의 이탈이 남긴 더 깊은 흔적은 따로 있었습니다. 중국과 맞서려면 AI가 필요하다는 주장이, 이때부터 훨씬 큰 목소리가 되었다는 점입니다. 의회에서, 공개 토론에서, 만약 중국이 AI를 갖는다면 미국은 어떤 충돌에서도 진다는 식의 이야기가 자라났습니다. 도덕적 반발이 펜타곤을 조심스럽게 만든 자리를, 중국이라는 위협의 서사가 메웠습니다.

그렇게 메이븐은 살아남았습니다. 그리고 한동안 잘 작동하지 않았습니다.

초기 알고리즘은 형편없었습니다. 맨슨이 청중 앞에서 한 말은 정확합니다. 알고리즘은 나무를 사람으로, 바위를 차량으로 잘못 봤습니다. 영상은 1초에도 여러 장의 정지화면으로 이뤄지는데, 어떤 프레임에서는 객체를 잡고 어떤 프레임에서는 놓치니, 화면 위 네모 상자가 깜빡깜빡 켜졌다 꺼졌다 했습니다. 집중하려던 분석관들은 오히려 산만해졌습니다. 어떤 알고리즘 개발자는 맨슨에게, 사람의 어깨에 걸친 무기를 알아보는 일이 때로 세 픽셀에 달려 있다고 말했습니다. 단 세 픽셀. 그들 자신도 그 정도 차이로 생사를 가르는 판단을 알고리즘에 맡길 만큼 자신이 없었습니다. 맨슨이 한 인터뷰에서 들려준 장면이 그 세 픽셀의 무게를 더합니다. 어떤 사람이 어깨에 무기를 메고 있는가, 아니면 채소 바구니를 안고 있는가. 알고리즘에게 그 차이는 몇 픽셀의 차이입니다. 그런데 그 몇 픽셀이 한 사람을 표적으로 만들지 말지를 가릅니다. 무기냐 장바구니냐가 생사를 가르는 판단이 됩니다. 알고리즘 개발자들조차 그 판단을 기계에 맡길 만큼 자기 알고리즘을 믿지 못했습니다.

더 무거운 사실도 있었습니다. 초기 알고리즘은 남자와 여자와 아이를 구별하지 못했습니다. 여자와 아이를 표적에서 제외하는 일은 전쟁법을 지키는 핵심인데, 그 핵심을 기계가 해내지 못했습니다. AI의 약속은 여럿이었습니다. 전쟁의 규모를 키운다, 속도를 키운다, 정밀도를 키운다. 맨슨은 속도에 관한 약속은 이란 작전에서 실제로 지켜지고 있다고 봅니다. 그러나 정밀도와 정확도라는 약속은, 알고리즘이 어떻게 작동하는지를 들여다보면 달랐습니다. 적어도 초기에는, 인간이 기계보다 훨씬 잘 알아봤습니다.

쿠코르는 그 시절 AI를 "감자칩 한 봉지"라고 불렀습니다. 맨슨이 한참을 곱씹은 비유였습니다. 결국 그 뜻은 이랬습니다. 알고리즘 자체는 별것 아니다, 중요한 건 이 알고리즘이 들어앉을 시스템이고, 우리가 바꿀 작전 방식이다.

그 잘 안 되던 시절을 맨슨은 자세히 들려줍니다. 메이븐은 소말리아, 아프가니스탄, 이라크, 카타르에서 알고리즘을 시험했습니다. 자꾸 영역을 넓혔습니다. 2018년에서 2019년 무렵의 일입니다. 마이크로소프트가 합류한 계기조차 실패에서 나왔습니다. 전장 사물의 사진이 모자라니, 합성 데이터를 만들어 알고리즘을 훈련시킬 수 있겠느냐고 메이븐이 물었습니다. 잘 안 됐습니다. 알고리즘은 어쩐지 합성 데이터를 꿰뚫어 봤습니다. 다만 마이크로소프트가 그 실패를 보여주는 솜씨가 어찌나 좋았던지, 메이븐은 합성 데이터는 접고 그들의 알고리즘을 쓰기로 했습니다. 그런 식이었습니다.

작전 현장에 알고리즘을 들이는 일은 더 험했습니다. 맨슨이 전하는 콜린 캐럴의 이야기가 그렇습니다. 메이븐이 첫 알고리즘을 소말리아에 내보낼 수 있었던 건, 캐럴이 그곳 지휘관과 옛 인연이 있었기 때문입니다. 기술의 힘이 아니라 사람 사이의 연줄. 그렇게 내보낸 첫해, 잘 풀리지 않았습니다. 알고리즘은 너무 많은 것을 잡아내며 깜빡거렸고, 어떤 화면에서는 탐지 상자 대신 얼룩을 띄워 손에 쥘 물건을 가렸습니다. 운영자들은 그냥 꺼버렸습니다. 메이븐 팀은 사람을 작전 지역까지 보내 운영자를 구슬렸습니다. "한 번만 써봐 달라, 중국을 생각해야 한다, 지금은 안 되는 거 안다, 그래도 준비 중이다." 부탁의 연속이었습니다.

그 시스템에 이름이 붙었습니다. 메이븐 스마트 시스템(Maven Smart System). 맨슨은 그것을 "전쟁판 구글 어스"라고 표현합니다. 눈앞의 전장을 지도로 펼치고, 아군과 적군을 표시하고, 컴퓨터 비전이 무언가를 탐지하면 화면에 점을 찍는 시스템. 어떤 이는 이것을 "전쟁의 윈도우", 또 어떤 이는 "AI 미션 컨트롤", "전쟁의 운영체제"라고 불렀습니다. 메이븐 스마트 시스템은 국가지리정보국, NGA로 옮겨 가며 자랐습니다. 그리고 거기서, 결정적인 돌파구가 열립니다. 육군의 제18공수군단이 2020년부터 위성 데이터에 알고리즘을 붙이는 실험을 거듭하기 시작한 것입니다. 그들이 원한 건 드론 영상이 아니었습니다. 위성에서 내려다본 전자광학·적외선 데이터에서, 눈앞에 무엇이 있는지를 기계가 짚어내는 일이었습니다.

맨슨의 책 7쪽은 그 시스템이 10년 뒤 어디까지 갔는지를 적습니다. 메이븐 스마트 시스템은 미군의 모든 군종에, 세계 곳곳에 배치되었습니다. 150개가 넘는 데이터 피드와 50곳 넘는 회사의 작업을 끌어안았습니다. 2025년 봄에는 나토가 한 버전을 쓰기 시작했고, 그해 10월에는 나토 회원국 열 곳이 자국군에 쓰려고 줄을 섰습니다. 한 취재원은 그것을 "전쟁의 마이크로소프트 윈도우"라 불렀습니다. 사용자 계정은 2만 5천 개를 넘었습니다. 한 정보장교의 분노에서 시작한 흰 점 하나가, 전쟁의 운영체제가 되었습니다.

그 실험이 진짜 시험대에 오르기까지는 그리 오래 걸리지 않았습니다. 2022년 2월, 러시아가 우크라이나를 침공했습니다.

2. 하룻밤 사이 눈(雪)을 학습한 알고리즘

사막에서 훈련받은 눈이 눈밭에 도착했습니다.

메이븐의 알고리즘은 중동에서 배웠습니다. 모래를 배경으로, 이라크와 아프가니스탄에서 벌어지는 종류의 움직임을 학습했습니다. 그런데 우크라이나에는 눈이 있었고, 키이우로 향하는 도로에는 러시아 탱크가 줄지어 서 있었습니다. 같은 탱크라도 배경이 다르면 기계는 헛갈립니다. 맨슨의 정리는 간명합니다. "잘 안 됐습니다."

그 순간, 미군과 알고리즘 제작사들이 한 일을 맨슨은 이렇게 전합니다. 위성을 움직였습니다. 새로 사진을 찍었습니다. 줄지어 선 탱크의 새 이미지를 알고리즘 제작사에 넘겼고, 그들은 "하룻밤 사이에" 모델을 다시 학습시켰습니다. 사막용으로 굳어 있던 눈을, 눈 덮인 우크라이나에 맞춰 밤새 다시 가르친 것입니다. 다음 날, 탐지 능력은 한 걸음 나아져 있었습니다.

이 장면은 두 가지를 동시에 말합니다. 한편으로는 놀라운 적응력입니다. 미군 기술이 20년 동안 걸어온 길을 떠올리면 더욱 그렇습니다. 흔한 결말은 이렇습니다. 약속하고, 작동하지 않고, 사업이 취소된다. 그보다 나으면, 무언가 나오긴 하는데 형편없이 망가져 있고, 고치겠다고 하고는 끝내 고치지 못한다. 그런데 메이븐은 망가진 채 도착했지만, 막대한 학습 데이터를 모으고, 그 데이터에 일일이 이름표를 붙이고, 알고리즘을 다시 훈련시켜 비교적 빠르게 재배치했습니다. 그렇게 쓸 만한 물건이 되었습니다.

다른 한편, 같은 장면은 이 기술의 약점을 드러냅니다. 맨슨과 대담한 AI 회의론자 게리 마커스가 짚은 지점이 그것입니다. 이런 시스템의 진짜 문제는 일반화입니다. 기계는 자기가 배운 것은 압니다. 그러나 개념적으로 압니다. 그래서 조금만 다른 상황으로 옮기면 무너집니다. 무인 자동차가 30년, 40년째 시제품 단계를 벗어나지 못하는 이유와 같습니다. 데이터가 넘치는 샌프란시스코에서는 굴러가도, 데이터가 없는 곳에서는 멈춰 섭니다. 눈 덮인 우크라이나는, 이 알고리즘에게 낯선 도시였습니다.

밤새 다시 학습시킨다는 말에는 한 가지 위험이 숨어 있습니다. 빠르게 고칠 수 있다는 것은, 빠르게 흔들린다는 뜻이기도 합니다. 모델은 시간이 지나며 표류합니다. 입력이 조금씩 바뀌면 출력도 미끄러집니다. 맨슨에게 한 전문가는, 생성형 AI가 질문하는 사람에게 동의하는 경향이 있다고 경고했습니다. "이 행동을 취할까요? 좋은 생각일까요?"라고

물으면, AI는 "그렇다"고 거들기 쉽습니다. 이 학계에서 아침이라 부르는 현상입니다. 전쟁에서 아침하는 기계는, 상황을 더 끌어오르게 만들 수 있습니다.

그러나 우크라이나에서 컴퓨터 비전이 실제로 쓸모를 얻은 건, 그것 하나만으로도 아니었습니다. 맨슨의 설명에서 핵심은 여기입니다. 컴퓨터 비전은 신호정보와 다른 정보들과 교차 대조되었을 때 비로소 좋아졌습니다. 위성이 무언가를 찍고, 도청한 통신이 그것을 거들고, 또 다른 출처가 한 번 더 확인합니다. 그렇게 겹쳐 보아야 그것이 무엇인지 손에 잡혔습니다. 미군은 그 첫해에 정말 잘하게 되었다고, 적어도 그들의 시각에서는 그렇다고 맨슨은 전합니다.

신뢰가 자라났습니다. 신뢰야말로 AI를 실험하고 키우는 일의 또 다른 핵심이라고 맨슨은 말합니다. 어느 정도였는가. 한 사례에서, 우크라이나군은 자기네 정보 출처만으로는 무엇을 때리는지 알 수 없었습니다. 그런데 미국 쪽은 "우리를 믿고 쳐라"라고 말할 수 있었습니다. 우크라이나군 눈에는 그저 사각형, 혹은 트럭으로 보이는 것이, 미군 눈에는 다른 무엇으로 보였기 때문입니다. 맨슨은 그것이 한 이동식 발사대였다고 전합니다.

미군은 우크라이나의 직접 교전 당사자가 아닙니다. 그래서 새 언어가 필요했습니다. '표적'이 아니라 '관심 지점', 영어로 points of interest. 맨슨은 그것이 "거의 이 일을 위해 발명된 말"이라고 합니다. 발사 결정 직전까지의 모든 것. 미국은 우크라이나에 무엇을 치라고 지시하지 않았습니니다. 미국의 기밀 정보를 넘기지도 않았습니니다. 다만 컴퓨터 비전으로 객체가 어디 있는지 짚어, 그 위치를 우크라이나와 나눴습니니다. 그것을 표적으로 만들지는 우크라이나의 몫이었습니니다. 맨슨은 거의 모든 취재원이 어느 순간엔가 무심코 그것을 "표적"이라 불렀다고 적습니니다. 다만 그들은 그것을 '합법적 표적'이라고는 부르지 않으려 조심했습니니다.

말의 줄타기였습니니다. 단어 하나로 직접 교전 당사자가 되느냐 마느냐가 갈렸습니니다. 그러나 화면 위에서 벌어지는 일은, 단어가 무엇이든, 점 하나가 좌표가 되고 좌표가 무기를 부르는 일이었습니니다. 빈자리는 바로 이 줄타기 안에 있었습니니다. 누가 그 점을 표적으로 만들었는가. 점을 찍은 알고리즘인가, 점을 건넌 미국인가, 방아쇠를 당긴 우크라이나인가. 책임이 셋 사이로 흩어졌습니니다.

기술만으로 풀리지 않는 문제도 있었습니니다. 메이븐이 유럽사령부에서 제대로 뜨지 않은 일을, 맨슨은 흥미로워습니니다. 네트워크가 너무 복잡했습니니다. 데이터 패킷이 대서양을 두

번, 어떤 경우엔 네 번씩 가로질렀습니다. 그러다 패킷을 잃거나 느려졌습니다. 기밀 시스템을 돌리려면 암호화 장비가 필요한데, 그게 병목이 되었습니다. 더 큰 암호화 장비를 옮기려면 정보 공동체의 아주 높은 사람에게 전화해 허락을 받아야 했습니다. 장비 하나가 미 국가안보국 인증을 받았기 때문입니다. 모든 것이 맞물려야 했고, 맞물리지 않았습니다. AI의 처리 능력을 키우려 한 노력은, 곧 그 옆에 숨어 있던 다른 모든 병목을 깨우는 일이었습니다.

그래도 결과는 분명했습니다. 2024년 초까지 우크라이나는 러시아 탱크 2천6백여 대를 부숴줍니다. 장갑차는 5천 대 가까이. 맨슨이 책에 적은 한 시점에, 미국이 우크라이나에 넘긴 관심 지점은 하루 최대 267개에 이르렀습니다. 정보 공유의 역사에서 전례 없는 숫자라고, 대담 진행자 그레고리 앨런은 말했습니다.

그러나 같은 이야기를 맨슨은 한 박자 늦춰 봅니다. 그 협력은 이듬해부터 느려졌습니다. 맨슨은 그것을 기술의 이야기보다 정책과 신뢰의 이야기로 봅니다. 첫 팀은 우크라이나와 "우리를 믿고 쳐라"라고 말할 만큼의 관계를 쌓았습니다. 정책의 잣대로는 통과하지 못할 수도 있는 일을, 전쟁의 압박 아래 그냥 했습니다. 새로 온 팀에는 그만한 친밀함이 없었습니다. 관심 지점을 넘기는 일이 괜찮은지, 미국의 기밀 시스템이 위협에 노출되는 건 아닌지를 두고 논쟁이 일어났습니다. 그리고 그 관계는 어느 순간 이어지지 않았습니다. 우크라이나는 다른 길을 찾았습니다. 자기네 드론으로.

데이터 자체가 흔들리기도 했습니다. 맨슨은 메이븐 초기에 훈련 데이터가 망가진 사례를 듭니다. 외부 적이 아니라, 일에 지친 데이터 라벨 작업자들이 훈련 데이터에 욕설을 잔뜩 적어 넣고 협조하지 않았습니다. 그런 데이터로 계속 학습시켰다면 알고리즘의 질에 영향을 주었을 것입니다. 잘 라벨링된 데이터를 시스템에 넣고, 다른 출처와 교차 확인하는 일. 그 평범해 보이는 일이 실은 제일 까다로운 일이었습니다.

이 시기 한 사람의 변심이 메이븐의 운명을 보여줍니다. 화이트워스 제독. 그는 메이븐을 누구보다 매섭게 의심하던 사람이었습니다. 표적화 주기에 오래 몸담은 정보통으로서, 그는 책임을 걱정했습니다. AI가 표적 오류에 한몫했을 때 의회 앞에서 누가 그것을 변호할 것인가. 시스템이 표적화 주기를 지름길로 건너뛰는 건 아닌가. NGA를 맡아 메이븐을 떠안게 되었을 때, 메이븐 사람들은 그가 이걸 죽일까 봐 걱정했습니다. 그런데 그를 돌려세운 건, 맨슨에게 그가 직접 한 말에 따르면, 메이븐이 전쟁의 현실에 맞춰 자기를

갱신하고 응답하는 속도가 그가 본 어떤 것보다 빨랐다는 사실이었습니다. 소프트웨어의 그 유연함이 회의론자를 신봉자로 바꾸었습니다. 그는 모델이 언제 성공하고 언제 실패하는지를 특성화해 사용자에게 알리려 애썼습니다. 어떤 이는 그것을 메이븐의 '철드는 과정'이라 불렀습니다. 다만 그 무렵부터 메이븐이 느려졌다는 불평도 함께 따라왔습니다.

기계의 학습은 하룻밤이면 됐습니다. 사람 사이의 신뢰는, 그렇게 빨리 옮겨지지 않았습니다.

3. 18분으로 줄어든 킬체인

이동식 발사대 하나가, 탐지된 지 18분 만에 사라졌습니다.

맨슨은 이 장면을 책에 적었고, 대담에서 진행자는 그 숫자를 다시 꺼냈습니다. 미국의 시가 한 이동식 미사일 발사대, 영어로 transporter erector launcher를 탐지했고, 탐지에서 파괴까지 18분이 걸렸다는 것입니다. 트럭에 실려 옮겨 다니는 미사일 발사대, 군에서 잡기 어렵기로 손꼽히는 표적. 그것이 18분 만에.

이 숫자를 어떻게 받아들여야 할까요. 먼저, 출처를 분명히 해 둡니다. 18분이라는 수치도, 하룻밤 눈 재학습의 디테일도, 카트리나 맨슨의 보도와 그의 책 '프로젝트 메이븐'에 근거합니다. 맨슨 자신은 자기 책의 많은 사례가 알고리즘이 성공한 이야기가 아니라 실패한 이야기, 그 실패를 딛고 배워 가는 곡선의 이야기라고 거듭 말합니다. 그러니 18분은 자랑이 아니라 기록입니다. 무엇이 빨라졌는가를 보여주는.

군에는 이 사슬을 부르는 이름이 있습니다. 킬체인, 곧 죽음의 사슬입니다. 풀어 쓰면 탐지하고, 고정하고, 추적하고, 조준하고, 교전하고, 평가하는 일련의 단계입니다. 영어 머리글자를 따 F2T2EA라고도 부릅니다. 찾고(find), 묶고(fix), 쫓고(track), 겨누고(target), 치고(engage), 살핀다(assess). 사람이 손으로 이 사슬을 돌리면 시간 단위가 걸렸습니다. AI를 끼우자 그 시간이 분 단위로 줄었습니다. 18분은 그 압축의 한 단면입니다.

압축의 규모는 다른 숫자에서 더 또렷합니다. 맨슨은 NGA 관계자에게서 들은 수치를 책 7쪽에 적었고, 대담에서 그대로 읽었습니다.

컴퓨터 비전의 도움으로, 미국은 하루에 타격할 수 있는 표적이 100개 미만에서 1천 개로 늘었습니다.

거기에 대규모 언어모델, 곧 LLM을 메이븐 플랫폼에 통합하자, 그 숫자는 다섯 배로 뛰어 하루 5천 개가 되었습니다.

100개 미만에서, 1천 개로, 다시 5천 개로. 한 단계 한 단계 떨어뜨려 보면 무슨 일이 일어났는지 보입니다. 표적을 만드는 일이, 사람의 속도에서 기계의 속도로 넘어갔습니다.

이것이 군사 전략가들이 말하는 킬체인 압축입니다. 같은 사슬을 같은 시간에 더 많이 돌리는 일. 분석관 한 사람의 일을 기계가 압축하면, 한 부대가 예전에 분석관 수천 명으로 하던 일을 수십 명으로 해낼 수 있다는 이야기가 나옵니다. 미국의 한 군사 싱크탱크는 자율과 AI에 대한 평가에서, AI가 약속하는 핵심 능력으로 속도와 정밀, 협조와 도달, 지속과 살상, 그리고 지구력을 꼽았습니다. 그 가운데 속도를 두고는 분명하게 적었습니다. 현대 전쟁의 속도는 인간이 방향을 잡고 이해하고 행동하는 속도를 이미 앞질렀다고. 기계의 속도로 지는 일이 가능해지면, 침략의 비용이 커지고, 그래서 역지가 강해진다는 것입니다. 그러나 같은 보고서는 다른 손으로 그 손을 붙잡습니다. 기계의 속도는 오인과 오판을 부르고, OODA 고리가 인간이 개입할 수 없을 만큼 빨리 돌면, 감독받는 자율과 완전한 자율의 경계가 흐려지거나 아예 의미를 잃는다고. 한 워게임에서는 자율로 설정된 방공 시스템이 미사일을 요격하는 데 그치지 않고, 북한을 향해 반격까지 쏘며 위기를 키웠습니다. 그 보고서의 한 문장이 이 장을 관통합니다. 기계의 속도는 의미 있는 인간 통제와 영원히 어긋난다.

그런데 LLM은 정확히 무엇을 한 걸까요. 맨슨의 답은 의외로 담백합니다. LLM은 컴퓨터 비전을 하지 않습니다. 무엇이 표적인지 골라내는 일은 LLM의 몫이 아닙니다. LLM이 한 일은 정보 보고서를 요약하고, 표적 승인에 필요한 절차를 빠르게 돌리는 것이었습니다. 승인 자체를 대신하지는 않습니다. 사람을 대신해 도장을 찍지 않습니다. 다만 그 승인을 받기까지 서류를 이리저리 나르는 행정, 법무 검토를 받으러 가는 그 잔무를 LLM이 대신 처리했습니다. 맨슨은 그것을 표적화 주기의 "행정 쪽"이라고 부릅니다. 식별하는 쪽이 아니라.

여기서 한 사람의 증언이 중요합니다. 맨슨이 만난 한 준위는, 처음에는 회의론자였습니다. 또 하나의 표적 도구 따위는 원하지 않는다며, 그것을 시연하겠다는 팔란티어 직원에게 욕을 하고 내쫓았습니다. 아프가니스탄 철수를 겪으며 그의 마음이 바뀌었습니다. AI 자체보다 데이터 통합의 힘을 본 것입니다. 모든 것을 한자리에 모으는 힘. 그는 팔란티어 직원에게 사과하고 시스템을 쓰기 시작했습니다. 능숙해지는 데 사흘이 걸렸다고 했습니다. 그리고 그가 맨슨에게 메이븐 스마트 시스템을 직접 시연하며 묘사한 동작이, 이 장에서 두고두고 남는 한 장면입니다.

수락, 수락, 수락.

AI가 탐지를 띄우면 그는 '예, 예, 예'를 누릅니다. 원래 그렇게 작동하도록 만들어진 것입니다. 그 준위 자신도 위험을 압니다. 자동화 편향이 생길 수 있다는 위험. 빠른 속도에서는 뚜껑 아래에서 무슨 일이 벌어지는지 볼 수 없습니다. 하나의 제안을 얼마나 들여다보아야 하는지, 어떻게 교차 확인하는지, 그 시스템을 얼마나 믿어도 되는지를 매번 따져야 합니다. 그러나 '예, 예, 예'의 리듬에 올라타면, 그 따짐이 흐려집니다.

맨슨은 그 줄어듦을 숫자로 잡아냅니다. 제18공수군단이 시스템을 쓰던 방식에서, 표적화 주기에 인간이 결정에 끼어드는 자리가 여섯 군데 있었습니다. AI의 도움으로 그 여섯 중 넷에서 인간을 빼냈습니다. 둘만 남았습니다. 결정은 여전히 인간의 몫이라고 그들은 말합니다. 그러나 데이터를 확인하던 일, 그런 종류의 절차가 천천히 자동으로 넘어갔습니다. 군에서 쓰는 다른 말로는 '고리 위에' 있다고 합니다. 무언가 잘못 굴러가면 멈출 수 있도록 감독자로 지켜본다는 뜻입니다. 매번 '실행'을 누르는 것과는 다릅니다.

여섯에서 둘로. 이 줄어듦을 어느 쪽으로 읽느냐가 이 책 전체의 질문입니다. 효율로 읽으면 진보이고, 책임으로 읽으면 공백입니다.

게리 마커스가 이 지점을 인지심리학으로 풀었습니다. 인간을 고리 안에 두더라도, 그 인간을 너무 세게 밀어붙이면 사실상 고리 밖에 있는 것과 같아진다는 것입니다. 사람은 오랫동안 경계 주의를 유지하지 못합니다. 이 발견의 원래 연구는 레이더 운영병을 대상으로 한 것이었습니다. 한 시간에 80개의 표적을 판단하라고 하면, 사람은 그 일을 제대로 해낼 수 없습니다. 그래서 AI를 믿기 시작합니다. 마커스는 그것을 "나한테는 괜찮아 보이는데 효과"라고 불렀습니다. 글이 문법에 맞고 매끄러우면 "괜찮아 보인다"고 말하게 되고, 그 안의 오류를 사람은 알아채지 못합니다.

빈자리가 여기서 한 번 더 벌어집니다. 인간이 고리 안에 있다고 말합니다. 그러나 그 인간이 한 시간에 수십 건을 '예, 예, 예'로 통과시키고 있다면, 그는 고리 안에 있는 걸까요, 밖에 있는 걸까요. 형식으로는 안에 있고, 실질로는 밖에 있습니다. 책임은 그 형식과 실질 사이로 빠져나갑니다.

맨슨이 끝까지 붙드는 또 하나의 사실이 있습니다. AI는 표적을 고를 수도 있지만, 무화과 잎이 될 수도 있다는 것. 지휘관이 먼저 "이 정도의 부수적 피해는 받아들일 것이다"고 정해 놓고, 표적 선정은 AI에 맡길 수 있습니다. 그러면 진짜 문제는 AI가 얼마나 정확한가가 아닙니다. 진짜 문제는 그 처음의 결정, 얼마만큼의 민간인 피해를 허용할 것인가입니다.

맨슨은 늘 두 질문이 함께 던져진다고 합니다. 하나는 시스템의 정밀도, 오류의 비용입니다. 학교 하나를 놓치고 있는 건 아닌가. 다른 하나는, 이 기술이 책임을 기계에 떠넘기는 데 쓰이고 있는 건 아닌가입니다.

그 '무화과 잎'의 위험은 추상이 아닙니다. 맨슨은 부수적 피해를 따지는 일에 AI가 어떻게 끼어드는지 한 사례로 보여줍니다. 2024년 그날 밤, 미군이 자기 병사 셋의 죽음에 보복하며 이라크와 시리아에서 85개 표적을 칠 때, 메이븐의 AI가 그 표적을 좁히는 데 쓰였습니다. 그런데 민간인 피해를 추적하는 비정부기구 에어워즈는, 그날 밤 죽은 사람 가운데 한 명이 민간인이었다고 보도했습니다. 그들은 이를 두고 'AI가 거둔 첫 민간인 사상자'라 불렀습니다. 맨슨 자신은 책에서 그 주장을 자기 입으로 하지는 않습니다. 다만 그는 묻습니다. AI를 시스템에 들이면 부수적 피해는 어떻게 달라지는가.

비슷한 그림자가 더 앞선 사건에도 드리워 있습니다. 2019년, 트럼프 대통령이 바그다디 제거를 발표하던 그 작전에서, AI는 흰 승합차 한 대를 추적하며 여러 차량을 살피고 있었다고 맨슨은 책에 적습니다. 그 흰 승합차는 뒤에 논란이 되었습니다. 미국 공영라디오 NPR이, 그 차에 민간인이 타고 있었고 한 명이 크게 다치고 한 명이 숨졌다고 보도했기 때문입니다. AI가 그 타격을 직접 책임진다는 뜻은 아닙니다. 다만 AI가 운영자에게 그 차의 존재를 알렸을 수 있고, 운영자는 그것을 위협으로 판단했습니다. 그 차에 민간인이 있었는지에 대한 공개 판정은 끝내 나오지 않았습니다.

이 두 번째 위험에 역사가 그림자를 드리웁니다. 맨슨은 책에서 1999년 미국이 베오그라드의 중국 대사관을 오폭한 사건을 들여다봅니다. 그때 나중에 공개된 오류 목록이 흥미롭습니다. 대사관은 실제 표적 근처에 있었고, 어떤 지도에서는 잘못 표기되었으나 다른 지도에서는 옳게 표기되어 있었습니다. 데이터베이스라는, 아주 실무적인 일. 그 바탕이 최신으로 유지되지 않았습니다. 누군가 무언가 잘못됐다고 느꼈을 때조차, 전화로 그 의심을 전할 수 없었습니다. 사람을 찾지 못했습니다.

맨슨이 던지는 물음이 차갑습니다. 이런 시스템들을 촘촘히 엮으면, 그 안전 간격을 메울 수도 있습니다. 그러나 아주 빠른 작전의 시대에는, 바로 그 통신과 확인에 걸리는 시간이 결정적이 됩니다. "쓰레기를 넣으면 쓰레기가 나온다"는 오랜 말이 있습니다. 낡은 데이터. 시스템이 새 데이터를 요청할 만큼 스스로를 알았는가 하는 물음. 18분으로 줄어든 사슬은, 그 18분 안에 오류를 알아챌 시간이 얼마나 남았는가를 함께 묻습니다.

그리고 맨슨은 한 가지를 분명히 합니다. 우크라이나에서, 그리고 2024년 2월 미 중부사령부가 처음 공개적으로 확인한 사례에서, 인간은 고리 안에 있었습니다. 미군이 자기 병사 셋의 죽음에 대한 보복으로 이라크와 시리아에서 85개 표적을 칠 때, 메이븐의 시로 그 85개를 줍혔다고 중부사령부 최고기술책임자가 맨슨에게 말했습니다. 인간이 함께 있었느냐는 물음에 맨슨은 "물론입니다, 분명히 해 두겠습니다"라고 답합니다. 운동가들이 두려워한 완전 자율 살상무기는, 그들이 걱정한 규모로는 아직 존재하지 않습니다.

다만 맨슨은 그 "아직"이라는 말을 가볍게 놓지 않습니다. 미국이 그것을 개발하려 하지 않는다는 뜻이 아니라고 그는 못 박습니다. 완전 자율 무기를 뒷받침하는 정책이 펜타곤에 있고, 펜타곤이 그 길을 좇는 방식을 자신이 책에서 탐색했다고 그는 말합니다. 그 길의 끝에서, 드론에 시를 올려 자동 표적 인식을 시키려는 시도가 비틀거리고 있다는 것도 함께 적습니다.

여기에 맨슨이 자기 책 마지막 문단으로 남긴 한 사람의 말이 놓입니다. 그는 쿠코르에게, 기술과 정책은 완전히 별개라는 흔한 주장을 두고 물었습니다. 높은 수준의 사용을 가능하게 하는 것이 결국 그 기술 아니냐고. 쿠코르는 "맞다"고 인정했습니다. 그리고 지난해 말, 그가 맨슨에게 한 말이 이렇습니다.

"거울에 비친 우리 자신을 똑바로 보고, 우리가 신중한지 확인합시다. 우리에게 이 모든 기술이 있습니다. 우리는 이것의 가장 좋은 관리자입니까?"

질문으로 끝나는 문장입니다. 답을 주지 않습니다.

쿠코르의 옛 상관이자 메이븐의 초대 책임자였던 잭 새너핸은, 같은 질문에 다른 각도로 답했습니다. 맨슨이 인용한 그의 말은 이렇습니다. "지금 형태의 어떤 LLM도, 완전히 치명적인 자율무기 시스템에 쓰일 후보로 고려되어선 안 된다. 그런 제안을 하는 것 자체가 어처구니없다." 기술을 누구보다 가까이서 키운 사람들조차, 그 사슬의 끝에 기계를 홀로 세우는 일에는 선을 그었습니다. 그 모순을 봉합하지 않는 것이, 이 장이 할 수 있는 일의 전부일지도 모릅니다.

맨슨은 책을 쓰며 한 가지를 자주 떠올렸다고 했습니다. 기자들은 흔히 미군 부대에 동행 취재를 합니다. 병사들이 목숨을 거는 모습을 곁에서 봅니다. 그는 독자가 사람이 아니라 알고리즘에 동행하는 책을 쓰고 싶었다고 말합니다. 알고리즘과 함께 전쟁에 들어가, 시를

전선에 들인다는 것이 무엇을 뜻하는지 들여다보는 책. 그는 그 일에 다 이르지 못하는 못했다고 했습니다. 그러나 그 미완의 시도가, 이 장이 따라온 길이기도 합니다.

화면은 여전히 멈추지 않습니다. 다만 이제, 그 화면 위에서 점이 좌표가 되고 좌표가 무기를 부르는 데 걸리는 시간이, 18분입니다. 사람이 결정에 끼어들 자리는 여섯에서 둘로 줄었습니다. 그리고 그 18분 안에서, 그 남은 둘 자리에서, 누가 방아쇠를 당겼는가 하는 물음의 자리가, 조금씩 비어 갑니다. 그 빈자리를 누가 채울지는, 아직 아무도 답하지 못했습니다.

제2장 우크라이나: 프로젝트 메이븐의 실전 진화

인간이 개입하는 지점을 줄이고 킬체인을 18분으로 압축해 전쟁의 운영체제로 자리 잡다.

킬체인 18분 압축

[적응과 속도]
2017년 미 국방부에서 시작된 프로젝트는 2022년 우크라이나의 눈발 환경에 맞춰 하루밤 만에 AI 모델을 재학습함.

[개입의 축소]
표적화 주기 내 인간 개입 지점을 6곳에서 2곳으로 축소했으며, 끊임없는 '수락(Yes)' 버튼의 반복으로 책임의 공백을 남음.

[처리량 폭증]
하루 표적 처리량은 100개 미만에서 컴퓨터 버전으로 1천 개, 거대언어모델(LLM) 결합 후 5천 개로 폭증함.

NotebookLM

제 2 부

통제선을 넘다

제3장 이란: 에픽 퓨리와 기계의 속도로 죽이기

1. 60초: 하메네이와 24시간 1,000개 표적

2026년 2월 28일 오전. 테헤란 북부의 한 안가에서 여든여섯 살의 노인이 차를 마시고 있었습니다. 이란 최고지도자 아야톨라 알리 하메네이. 그는 거의 37년 동안 한 나라의 종교와 군대와 외교를 한 손에 쥔 사람이었습니다. 그날 그의 위치를 알고 있던 것은 그의 경호원들만이 아니었습니다.

워싱턴의 한 화면 위에서, 그의 안가는 좌표 하나로 떠 있었습니다.

미국이 '에픽 퓨리 작전(Operation Epic Fury)'이라고 이름 붙인 그 일은 단 60초 안에 끝났습니다. AP통신과 밀리터리 타임스의 보도에 따르면, 세 곳의 표적이 같은 순간에 타격을 받았습니다. 거의 동시였습니다. 한 곳이 맞고 다음 곳이 맞은 것이 아니라, 세 곳이 한 호흡 안에 무너졌습니다. B-2 스텔스 폭격기가 2,000파운드짜리 폭탄을 떨어뜨렸다고 같은 보도는 전합니다. 하메네이와 약 40명의 고위 인사가 그 1분 안에 목숨을 잃었습니다.

군사 작전에서 1분은 보통 무엇 하나를 시작하기에도 짧은 시간입니다. 비행기가 활주로를 빠져나가는 데에도 그보다 오래 걸립니다. 그런데 한 나라의 지도부 상당수가 그 1분 안에 사라졌습니다.

세 곳이 동시에 무너졌다는 말을 다시 생각해 봅시다. 세 곳을 차례로 친다면, 첫 번째 타격이 일어난 순간 나머지 두 곳의 사람들은 도망칠 시간을 얻습니다. 폭발음이 울리고, 경보가 울리고, 경호원들이 지도자를 데리고 빠져나갑니다. 그래서 참수 작전은 늘 첫 타격에서 빗나가면 실패합니다. 표적이 흩어지기 때문입니다. 동시 타격은 그 흩어질 틈을 주지 않겠다는 뜻입니다. 세 곳이 같은 1분 안에 무너지려면, 세 곳의 위치가 같은 순간에 확정되어 있어야 하고, 세 발의 타격이 같은 순간에 준비되어 있어야 합니다. 사람의 손으로 세 곳의 정보를 동시에 맞추기란 거의 불가능에 가깝습니다. 이 동시성이야말로 기계가 들어온 자리입니다.

이 작전이 가능했던 이유를 따라가면, 결국 한 가지 질문에 닿게 됩니다. 사람이 그렇게 빨리 결정을 내릴 수 있었을까요.

답은 사람만으로는 아니었다는 것입니다.

미국 대통령이 작전을 승인한 것은 2월 27일이었습니다. 개시는 다음 날 28일 오전, 이란 표준시 기준이었습니다. 미국 중앙정보국(CIA)이 몇 달에 걸쳐 이란 지도부의 동선을 추적해 왔다고 여러 매체가 전합니다. 몇 달의 추적과 60초의 타격. 이 둘 사이의 간격이 이 장 전체의 긴장을 만듭니다. 정보는 천천히 쌓이고, 결정은 순식간에 떨어집니다. 그 사이에 사람이 멈춰 서서 따져 볼 자리가 점점 좁아졌습니다. 이스라엘은 같은 작전을 '로어링 라이언(Roaring Lion)'이라는 이름으로 함께 수행했습니다. 한쪽은 에픽 퓨리, 다른 쪽은 로어링 라이언. 같은 밤의 두 이름입니다.

그러나 추적과 타격 사이에는 늘 가장 어려운 구간이 있습니다. 수집한 정보를 표적으로 바꾸는 일입니다. 어디를 칠 것인가. 누구를 칠 것인가. 지금 이 좌표가 맞는가. 전통적으로 이 일은 사람이 했습니다. 분석관이 위성 사진을 보고, 통신을 듣고, 여러 정보를 겹쳐 놓고, 이것이 표적이 맞는지 며칠씩 따졌습니다.

이번에는 그 일을 기계가 거들었습니다.

블룸버그와 데모크라시 나우의 보도에 따르면, 이 작전의 표적 선정에는 팔란티어가 운용하는 시스템이 쓰였습니다. 미군이 '프로젝트 메이븐' 또는 '메이븐 스마트 시스템'이라 부르는 표적 분석 체계입니다. 그리고 그 안에 앤스로픽(Anthropic)의 인공지능 클로드(Claude)가 결합되어 있었다고 블룸버그는 전합니다. 미국 회사가 만든 챗봇이, 미국 회사가 만든 정보 분석 플랫폼에 실려, 한 나라의 지도부를 겨냥하는 일에 닿아 있었던 셈입니다.

숫자를 보면 이 변화의 크기가 보입니다.

블룸버그에 따르면, 작전 첫 24시간 동안 미군은 표적 1,000개를 다뤘습니다.

3주가 지나자 그 숫자는 5,500개에서 6,000개 사이로 늘었습니다.

4월 6일까지는 약 1만 3천 개에 이르렀습니다.

하루 천 개. 사람의 손으로는 도무지 닿을 수 없는 속도입니다. 2003년 이라크 전쟁이었다면, 같은 규모의 표적 분석에 분석관 약 2,000명이 매달려야 했다고 블룸버그는 전합니다. 이번에는 약 20명이 그 일을 했습니다. 2,000명이 하던 일을 20명이 했다는 말은, 사람이 백 배 똑똑해졌다는 뜻이 아닙니다. 사람과 표적 사이에 기계가 들어와 그

사이의 시간을 압축했다는 뜻입니다.

뉴캐슬대학의 법학자 크레이그 존스(Craig Jones)는 데모크라시 나우에서 이 변화를 한 문장으로 정리했습니다. 인공지능이 "수만 시간의 인간 작업을 초와 분 단위로 압축"했다는 것입니다. 며칠 걸리던 판단이 몇 분 만에 끝나고, 몇 시간 걸리던 검증이 몇 초 만에 지나갑니다.

이 압축이 무엇을 뜻하는지, 표적 선정이라는 일의 속살을 들여다보면 더 잘 보입니다. 군대에는 표적을 정하는 절차가 있습니다. 영어로는 '킬체인(kill chain)', 우리말로 옮기면 죽음의 사슬쯤 되는 말입니다. 무엇을 칠지 찾고, 그 위치를 고정하고, 그것이 맞는지 확인하고, 마지막으로 타격하는 일련의 단계입니다. 찾고, 고정하고, 끝낸다. 전통적으로 이 사슬의 한 고리 한 고리에는 사람이 앉아 있었습니다. 위성 사진을 보는 사람, 통신을 듣는 사람, 법무관, 지휘관. 각자가 자기 고리에서 한 번씩 멈춰 서서 "이게 맞나" 하고 물었습니다.

기계가 들어오면서 이 멈춤이 사라졌습니다. 정확히 말하면, 멈춤이 짧아졌습니다. 사슬의 여러 고리를 기계가 한꺼번에 처리하니, 사람이 멈춰 설 자리가 줄어듭니다. 찾는 일과 고정하는 일과 확인하는 일이 한 화면 안에서 거의 동시에 일어납니다. 그래서 1,000개의 표적이 24시간 안에 흐를 수 있는 것입니다. 사슬이 빨라진 것이 아니라, 사슬에서 사람이 빠진 만큼 빨라진 것입니다.

여기에 한 가지 심리가 더 끼어듭니다. 자동화 편향이라는 것입니다. 기계가 답을 내놓으면, 사람은 그 답을 의심하기보다 받아들이는 쪽으로 기웁니다. 계산기가 내놓은 숫자를 우리가 다시 검산하지 않듯이 말입니다. 표적 화면에 좌표가 뜨고 "이것이 표적입니다"라고 적혀 있으면, 1분의 시간을 가진 사람은 그것을 뒤집기보다 서명하기 쉽습니다. 기계를 믿는 마음이, 사람의 마지막 판단을 무르게 만듭니다. 이 편향은 이 책의 뒤쪽에서 따로 다룰 만큼 중요한 주제입니다. 지금은 한 가지만 기억하면 됩니다. 사람이 사슬 안에 있다는 사실과, 사람이 실제로 판단하고 있다는 사실은 같은 말이 아니라는 것입니다.

여기서 잠깐 멈춰 봐야 합니다. 속도가 빨라졌다는 것은 좋은 일일 수도 있습니다. 빠른 정보는 전쟁을 억제할 수도 있습니다. 적이 무엇을 하는지 먼저 안다면 충돌을 피할 수도 있습니다. 클로드를 만든 회사의 최고경영자 다리오 아모데이(Dario Amodei)도 블룸버그 인터뷰에서 같은 취지로 말했습니다. 우월한 정보 능력이 오히려 충돌을 억제할 수 있다는

것입니다. 미국이 더 강한 행위자가 되는 일을 자신은 지지한다고도 했습니다. 그는 자신을 애국자라고 불렀습니다.

그러나 속도에는 다른 얼굴이 있습니다.

판단이 빨라질수록, 그 판단을 들여다볼 시간은 줄어듭니다. 표적 1,000개를 24시간 안에 처리한다면, 표적 하나에 사람이 쓸 수 있는 시간은 산술적으로 1분 남짓입니다. 그 1분 안에 사람은 무엇을 검토할 수 있을까요. 좌표가 맞는지, 그 건물이 지금도 군사 시설인지, 그 안에 누가 있는지를 1분 안에 확인할 수 있을까요.

표적 분류의 정확도를 보면 이 질문이 더 무거워집니다. 블룸버그에 따르면, 메이븐의 표적 분류 정확도는 약 60퍼센트였습니다. 사람이 직접 했을 때의 정확도는 약 84퍼센트로 보도됐습니다. 기계가 사람보다 덜 정확했다는 뜻입니다. 그리고 날씨가 나쁠 때, 구름이 끼거나 시야가 흐릴 때 그 정확도는 30퍼센트 아래로 떨어졌다고 같은 보도는 전합니다. 열 개를 쫓으면 일곱 개가 빗나가는 셈입니다.

기계가 60퍼센트의 정확도로 표적을 골라 사람 앞에 내놓습니다. 사람은 1분 안에 그것을 검토합니다. 검토를 마친 사람이 도장을 찍습니다. 그리고 다음 표적으로 넘어갑니다. 하루에 천 번. 이 흐름 안에서 사람의 역할은 무엇일까요. 판단하는 사람일까요, 아니면 기계가 내놓은 결과에 서명하는 사람일까요.

미국 국방장관 피트 헤그세스(Pete Hegseth)는 그 흐름을 더 빠르게 밀어붙였습니다. 그는 "멍청한 교전수칙은 없다"는 취지로 발언했다고 보도됐습니다. 교전수칙(交戰守則)이란 군인이 언제 어떻게 무기를 쓸 수 있는지를 정해 놓은 규칙입니다. 민간인을 보호하고, 무력 사용에 한계를 두기 위한 장치입니다. 그것을 멍청하다고 부르는 순간, 속도를 가로막는 모든 것이 거추장스러운 것이 됩니다.

이란 외무부는 처음에 하메네이가 "안전하다"고 주장했습니다. 지도자의 죽음을 인정하지 않은 것입니다. 사망이 공식 확인된 것은 3월 1일, 이란 국가안보회의를 통해서였습니다. 일주일 뒤인 3월 8일, 후계자로 모즈타바 하메네이(Mojtaba Khamenei)가 지명됐습니다. 아버지의 자리를 아들이 이었습니다.

하메네이 사살이 한낮에, 그러니까 주간에 실행됐다는 점도 이례적으로 보도됐습니다. 표적 사살은 보통 어둠 속에서 이뤄집니다. 밤이 사람과 기계를 함께 숨겨 주기 때문입니다. 그런데 이번에는 낮이었습시다. 그만큼 정보가 확실했거나, 그만큼 망설일 이유가 없었다는 뜻입니다.

지도부를 한 번에 제거하는 이런 공격을 군사 용어로 참수 작전이라 부릅니다. 머리를 친다는 뜻입니다. 적의 지휘 구조를 마비시켜 군대가 명령을 받지 못하게 만드는 것입니다. 이 작전이 성공하려면 두 가지가 맞아야 합니다. 지도자가 지금 어디 있는지 정확히 알아야 하고, 그 정보가 행동으로 옮겨지기 전에 지도자가 움직이지 않아야 합니다. 사람의 손으로 이 둘을 맞추려면, 정보를 모으는 동안 표적이 자리를 떠나 버리기 일쑤입니다. 정보와 행동 사이의 시간이 길기 때문입니다. 그 시간을 기계가 줄였습니다. 위치를 잡은 순간과 폭탄이 떨어지는 순간 사이가 좁아졌습니다. 60초라는 숫자는 그 좁아진 틈의 다른 이름입니다.

이것을 두고 누군가는 정밀해졌다고 말할 것입니다. 옳은 말입니다. 예전 같으면 도시 하나를 폭격해야 잡았을 표적을, 이제는 건물 하나로 좁혀 칩니다. 민간인 피해를 줄이는 기술이기도 합니다. 그러나 정밀함과 정확함은 다른 말입니다. 정밀하다는 것은 겨눈 곳을 정확히 맞힌다는 뜻이고, 정확하다는 것은 겨눈 곳이 옳은 곳이라는 뜻입니다. 미사일은 점점 정밀해졌습니다. 그러나 어디를 겨눌지 정하는 일, 그 일의 정확함은 미사일의 성능과는 다른 문제입니다. 그것은 데이터의 문제이고, 판단의 문제이고, 시간의 문제입니다. 다음 절의 학교가 그 차이를 말해 줍니다.

사망자 집계는 처음부터 엇갈렸습니다. 초기에 이란 적신월사는 약 201명이 숨졌다고 집계했습니다. 그러나 후속 집계는 훨씬 컸습니다. 위키피디아의 종합 집계는 이란 측 사망자를 3,468명, 부상자를 2만 6,500명으로 적고 있습니다. 201명과 3,468명. 같은 사건의 두 숫자입니다. 어느 쪽이 맞는지 이 글을 쓰는 시점에 확정해 말하기는 어렵습니다. 출처에 따라 다르고, 집계 기준에 따라 다릅니다. 다만 두 숫자가 이렇게까지 벌어진다는 사실 자체가, 이 작전이 얼마나 빠르고 넓게 번졌는지를 말해 줍니다.

미군 측에서도 사망자가 나왔다고 보도됐습니다. 이 작전은 한쪽만 다친 일이 아니었습니다.

60초. 표적 1,000개. 정확도 60퍼센트. 이 세 숫자를 나란히 놓으면 한 시대의 윤곽이 보입니다. 전쟁의 가장 무거운 결정, 누구를 죽일 것인가 하는 결정이 사람의 속도에서

기계의 속도로 넘어간 시대입니다. 그 시대에 사람은 여전히 방아쇠 앞에 앉아 있습니다. 다만 그 앞에 기계가 내려놓은 천 개의 좌표가 쌓여 있고, 사람에게 주어진 시간은 좌표 하나당 1분입니다.

이 장의 다음 절은 그 1분 안에서 무슨 일이 벌어질 수 있는지를 보여 줍니다. 한 학교의 좌표에서.

2. 미나브 학교: 낡은 좌표가 부른 175명

교실 맨 앞줄, 창가에서 두 번째 자리. 그 아이의 책상에는 그날 아침에도 공책이 펼쳐져 있었습니다.

이란 남부 호르모즈간주의 항구 도시 미나브(Minab). 샤레제 타예베(Shajareh Tayyebeh) 초등학교. 2026년 2월 28일 오전, 시계는 열 시 이십 분을 조금 지나고 있었습니다. 1교시가 끝나고 2교시가 시작될 무렵이었습니다. 아이들은 자리에 앉아 있었습니다. 여교사들이 칠판 앞에 서 있었습니다.

그 시각, 이란 표준시로 대략 열 시 이십삼 분에서 열 시 사십오 분 사이에, 미국제 토마호크 순항미사일로 추정되는 무기가 그 학교 건물에 떨어졌습니다. 뉴욕타임스와 CNN의 보도입니다. 무기의 종류는 UGM-109, 토마호크 계열로 추정됐습니다.

사망자 집계는 엇갈립니다. 약 156명에서 175명 이상이라고 보도됐습니다. 대부분이 일곱 살에서 열두 살 사이의 아이들이었습니다. 남자아이가 약 73명, 여자아이가 약 47명. 그리고 그 학교에서 일하던 여교사 약 26명이 모두 숨졌습니다. 한 사람도 살아남지 못했습니다.

이 숫자들 앞에서 우리는 조심해야 합니다. 156명과 175명 가운데 어느 쪽이 정확한지 지금 단정하기는 어렵습니다. 집계는 진행 중이었고, 출처마다 달랐습니다. 그러나 숫자가 흔들린다고 해서 그 자리에 있던 아이들이 흔들리는 것은 아닙니다. 일곱 살이었던 아이는 일곱 살이었습니다. 창가에서 두 번째 자리에 앉아 있던 아이는 그 자리에 있었습니다.

숫자로 적으면 175명은 한 줄입니다. 그러나 그 한 줄을 풀면 175개의 아침이 됩니다. 175명의 아이가 그날 아침 가방을 메고 집을 나섰습니다. 누군가는 늦잠을 자서 뛰어왔을 것이고, 누군가는 어제 못 푼 숙제를 걱정하며 걸어왔을 것입니다. 어떤 아이는 짝궁과 다툼 일을 마음에 두고 교실에 들어섰을지도 모릅니다. 일곱 살에서 열두 살. 이가 빠지기 시작하는 나이와 키가 부쩍 자라는 나이 사이입니다. 그 아이들에게 혁명수비대니 표적 데이터베이스니 하는 말은 평생 들어 본 적 없는 말이었을 것입니다.

여교사 26명도 그날 아침 출근했습니다. 그 가운데 누군가는 그 학교에서 오래 가르친 사람이었을 것이고, 누군가는 갓 부임한 젊은 교사였을 것입니다. 그들이 2교시 종이

울리기를 기다리며 칠판 앞에 섰을 때, 자신들이 어느 군사 시스템의 화면 위에서 10년 전 좌표로 떠 있다는 사실을 알 도리가 없었습니다. 선정적으로 그날의 장면을 그리지는 않겠습니다. 그것은 죽은 이들을 두 번 소비하는 일입니다. 다만 한 가지는 분명히 해 두고 싶습니다. 이 책에서 175라는 숫자가 나올 때마다, 그것은 통계가 아니라 175개의 빈 책상이라는 사실입니다.

이런 일이 어떻게 일어났을까요. 무기가 빛나간 것이 아닙니다. 미사일은 정확히 좌표로 날아갔습니다. 문제는 그 좌표였습니다.

뉴욕타임스와 CNN은 미국 국방정보국(DIA)의 낡고 갱신되지 않은 좌표가 원인이라고 보도했습니다. 국방정보국은 군사 정보를 다루는 기관입니다. 그 기관의 데이터베이스에는 미나브의 그 부지가 군사 시설로 기록되어 있었습니다. 혁명수비대의 기지로 말입니다.

한때는 맞는 말이었습니다. 그 학교는 2016년까지 혁명수비대 기지와 같은 부지를 쓰고 있었습니다. 군사 시설과 학교가 한 울타리 안에 있었던 것입니다. 그러다 2016년 무렵 둘이 분리됐습니다. 군은 떠나고, 학교만 남았습니다.

10년 전의 정보였습니다. 10년 동안 그 좌표는 갱신되지 않았습니다. 데이터베이스 안에서 그 부지는 여전히 혁명수비대 기지였습니다. 사람의 기억이라면 10년이면 흐려질 만한 시간입니다. 그러나 기계의 데이터베이스는 흐려지지 않습니다. 다만 갱신되지 않을 뿐입니다. 갱신되지 않은 데이터는 틀린 채로 또렷하게 남아 있습니다. 그리고 그 또렷함이 더 위험합니다. 흐릿한 정보는 사람이 의심하지만, 또렷한 좌표는 사람이 믿기 때문입니다.

데이터에는 시간이 적혀 있지 않습니다. 위성 사진 한 장을 떠올려 봅시다. 거기에는 건물이 찍혀 있고 좌표가 붙어 있습니다. 그러나 그 사진이 언제 찍혔는지, 그 건물이 그 사이에 무엇으로 바뀌었는지는 사진 자체에 적혀 있지 않습니다. 사람이 따로 알아내야 합니다. 누군가가 "이 좌표 정보는 2016년 것이니 다시 확인하자"고 말해 줘야 합니다. 그 말을 해 줄 사람이 1분의 시간 안에 있었을까요. 하루에 천 개의 표적이 흐르는 그 흐름 안에, 좌표 하나를 붙잡고 "이거 언제 적 정보지" 하고 물어볼 사람이 있었을까요.

기계는 그 질문을 하지 않습니다. 기계에게 2016년의 좌표와 2026년의 좌표는 똑같이 또렷한 숫자일 뿐입니다. 숫자에는 나이가 없습니다. 그래서 낡은 좌표가 새 좌표의 얼굴을 하고 표적 화면에 떴습니다. 그리고 그 위에 미사일이 떨어졌습니다.

더 무거운 사실이 있습니다. 공습 당시 그 학교의 웹사이트가 살아 있었습니다. 인터넷 지도에도 학교로 표시되어 있었습니다. CNN의 지적입니다. 누구라도 그 학교 이름을 검색창에 넣었다면, 그곳이 초등학교라는 사실을 몇 초 만에 알 수 있었습니다. 학교 홈페이지가 떴을 것이고, 지도에 학교 표시가 떴을 것입니다.

이 사실은 나중에 클로드를 만든 회사의 최고경영자에게 직접 던져진 질문이 됩니다. 그 학교는 웹사이트가 있었다, 구글에서 검색하면 나왔다, 그런데 왜 클로드는, 혹은 그들이 쓴 어떤 기술이든, 그것을 짚어 내지 못했느냐는 질문입니다. 이 장의 마지막 절에서 그 질문과 그의 대답을 다루겠습니다.

지금 짚어야 할 것은 책임의 구조입니다.

미사일을 만든 사람은 미사일을 정확히 만들었습니다. 토마호크는 좌표로 날아가도록 설계되어 있고, 좌표로 정확히 날아갔습니다. 좌표를 입력한 시스템은 데이터베이스에 있던 좌표를 충실히 가져왔습니다. 데이터베이스를 만든 기관은 2016년에 정확한 정보를 입력했습니다. 표적을 검토한 사람은 시스템이 내놓은 결과에 서명했습니다. 방아쇠를 당긴 사람은 명령에 따랐습니다.

모두가 자기 자리에서 자기 일을 했습니다. 그런데 175명이 죽었습니다.

이것이 이 책이 '빈자리'라고 부르는 것의 한 모습입니다. 책임이 사라진 자리. 누구도 거짓말을 하지 않았고, 누구도 명백히 규칙을 어기지 않았는데, 결과만 남은 자리입니다. 데이터를 갱신하지 않은 책임은 누구에게 있을까요. 10년 된 좌표를 의심하지 않은 책임은 누구에게 있을까요. 1분 안에 천 개의 표적을 처리하는 흐름 속에서, 좌표 하나를 따로 떼어 검색창에 넣어 볼 책임은 누구에게 있었을까요.

미국 중부사령부(CENTCOM)의 대변인 팀 호킨스(Tim Hawkins) 대위가 이 사건에 대해 언론에 대응했습니다. 군의 공식 창구가 움직였다는 것은, 이 일이 작은 사고로 묻히지 않았다는 뜻입니다.

국제인권단체 휴먼라이츠워치(HRW)는 2026년 4월, 이 공습이 전쟁범죄에 해당하는지를 분석했습니다. 전쟁범죄라는 말은 무겁습니다. 국제법에는 전투원과 민간인을 구별해야 한다는 원칙이 있습니다. 구별의 원칙입니다. 그리고 공격의 군사적 이익이 민간인

피해보다 커야 한다는 원칙도 있습니다. 비례의 원칙입니다. 학교를 군사 시설로 착각하고 친 공격이 이 원칙들을 어겼는지, 휴먼라이츠워치는 따져 물었습니다.

구별의 원칙은 전쟁법의 가장 오래된 뼈대입니다. 군인은 군인을 치되 민간인은 치지 말라는 것. 어린아이라면 더 말할 것도 없습니다. 그런데 이 원칙은 한 가지를 전제로 합니다. 누가 군인이고 누가 민간인인지 분간할 수 있어야 한다는 것입니다. 그 분간을 사람이 할 때는, 사람이 책임을 졌습니다. 잘못 분간하면 그 사람이 답해야 했습니다. 그런데 분간을 기계가 거들고, 그 기계가 10년 된 데이터를 보고, 사람은 그 결과를 1분 안에 받아들였다면, 잘못된 분간의 책임은 어디로 가는 것일까요. 구별의 원칙은 그대로인데, 구별을 실제로 행하는 손이 바뀌었습니다. 원칙은 사람을 향해 쓰여 있는데, 일은 기계가 절반쯤 하고 있습니다. 이 어긋남이 미나브의 교실에서 175명의 모습으로 드러났습니다.

여기서 변호인의 자리에서 한마디 덧붙이고 싶습니다. 착각이었다는 변명은 법 앞에서 통하지 않습니다. 사람이 운전하다 사람을 치었을 때, "몰랐다"는 말만으로 책임을 면하지 못합니다. 알았어야 할 의무가 있었기 때문입니다. 10년 된 좌표를 그대로 쓴 것이 과실인지, 그 과실의 무게가 어디까지인지는 법정에서 가려질 일입니다. 다만 분명한 것이 있습니다. 시스템이 빨라졌다고 해서, 알았어야 할 의무까지 빨라진 것은 아니라는 점입니다.

그 학교의 여교사 26명을 떠올려 봅시다. 그들은 그날 아침 출근했습니다. 칠판 앞에 섰습니다. 아이들의 이름을 불렀을 것입니다. 그들은 자신이 표적 데이터베이스 안에서 10년 전 좌표로 남아 있다는 사실을 알 길이 없었습니다. 알 수 있는 사람은 아무도 없습니다. 그것이 이 시대의 무서운 점입니다. 어떤 사람은 자신도 모르는 사이에, 어느 데이터베이스 안의 한 좌표가 되어 있습니다. 그 좌표가 갱신되었는지 아닌지는, 그 사람의 운명을 가르지만, 그 사람은 그것을 알 수도 바꿀 수도 없습니다.

창가에서 두 번째 자리의 아이로 다시 돌아갑니다. 그 자리는 이제 비어 있습니다. 그 아이의 공책도, 펼쳐져 있던 그 공책도 그 자리에 함께 있었습니다. 한 아이의 빈자리. 이 책의 제목이 가리키는 것은 여러 빈자리입니다. 책임이 사라진 자리와, 사람이 사라진 자리. 미나브에서는 그 둘이 한 교실 안에 함께 있었습니다.

3. 실리콘밸리와 국방부의 충돌: 클라우드와 공급망 위험

2026년 2월의 어느 사흘. 샌프란시스코의 한 회사에 워싱턴에서 통첩이 도착했습니다. 사흘 안에 답하라. 사흘이라는 시간이 화면 위에서 줄어들고 있었습니다.

회의실 안의 사람들은 두 가지를 동시에 알고 있었을 것입니다. 거절하면 회사가 위험해진다는 것. 그리고 받아들이면 자신들이 그어 놓은 선이 무너진다는 것. 어느 쪽도 편한 길이 아니었습니다.

통첩을 보낸 쪽은 미국 국방부, 이제는 '전쟁부(Department of War)'라고도 불리는 그 기관이었습니다. 받은 쪽은 앤스로픽, 클라우드를 만든 회사였습니다. 요구는 분명했습니다. 클라우드의 모든 사용 제한을 풀라는 것이었습니다.

이 충돌을 이해하려면 한 가지 사실부터 짚어야 합니다. 앤스로픽은 군과 거리를 둔 회사가 아니었습니다. 오히려 그 반대였습니다.

다리오 아모데이는 CBS 인터뷰에서 이 점을 분명히 했습니다. "앤스로픽은 모든 AI 회사 가운데 미국 정부와 미군과 일하는 데 가장 앞장선 회사"였다는 것입니다. 그의 말을 그대로 옮기면 이렇습니다. 자사 모델을 기밀 클라우드에 올린 첫 회사였고, 국가안보 목적의 맞춤형 모델을 만든 첫 회사였으며, 정보 공동체와 군 전반에 배치되어 사이버와 전투 지원 작전 같은 일에 쓰이고 있다는 것입니다.

왜 그랬느냐는 질문에 그는 답했습니다. "우리는 우리나라를 지켜야 한다고 믿기 때문"이라는 것입니다. 중국과 러시아 같은 권위주의 국가로부터 나라를 지켜야 한다고 그는 말했습니다. 자신을 "애국적인 미국인"이라고 불렀습니다.

그런데 그 애국적인 미국인이 두 가지에는 선을 그었습니다.

첫째는 대규모 국내 감시입니다. 그는 CBS에서 이렇게 설명했습니다. 민간 기업이 모은 데이터를 정부가 사들여, 인공지능으로 한꺼번에 분석하는 일. 그것은 사실 불법이 아니라고 그는 말했습니다. 다만 인공지능 이전 시대에는 쓸모가 없었을 뿐이라는 것입니다. 기술이 너무 빨라서 법이 따라잡지 못한 영역. 그래서 위험한 영역. 그것이 그가 그은 첫 번째 선이었습니다.

둘째는 완전 자율무기입니다. 그는 이것을 우크라이나에서 쓰이는 부분 자율무기와 구별했습니다. 완전 자율무기란 "사람의 어떤 개입도 없이 발사되는 무기"를 만든다는 생각이라고 그는 말했습니다. 누구를 표적으로 삼을지, 누구를 쏠지를 결정할 사람이 그 사슬 안에 없는 무기입니다. 그는 이런 무기에 대해 두 가지를 우려했습니다. 하나는 신뢰성입니다. 오늘날의 인공지능은 완전 자율무기를 만들 만큼 충분히 믿을 만하지 않다는 것입니다. 인공지능에는 "근본적인 예측 불가능성"이 있고, 그것을 기술적으로 아직 풀지 못했다고 그는 말했습니다. 다른 하나는 감독입니다. 사람의 감독 없이 움직이는 드론이나 로봇의 군대가 있다면, 누가 표적을 정하고 누가 책임을 지느냐는 문제가 생긴다는 것입니다.

블룸버그 인터뷰에서 그는 이 우려를 한 영화에 빗댔습니다. '닥터 스트레인지러브'입니다. 핵무기가 날아온다고 판단되면 자동으로 핵무기를 발사하는 종말 장치. 무엇이 잘못될 수 있겠느냐고 그는 되물었습니다. 두 나라가 서로를 오해하고 서로에게 달려드는 방식으로 전쟁이 일어나는데, 그 기술에 적절한 감독이 없으면 그런 사고가 더 일어나기 쉽다는 것입니다.

앤스로픽의 셈법으로는, 전체 사용 사례의 98에서 99퍼센트는 받아들일 수 있었습니다. 막고 싶었던 것은 1퍼센트였습니다. 대규모 국내 감시와 완전 자율무기, 그 둘뿐이었습니다.

아모데이는 이 선의 성격을 한 가지 비유로 설명했습니다. 블룸버그 인터뷰에서였습니다. 적이 전쟁범죄를 저지른다고 해서 우리도 전쟁범죄를 저질러야 하느냐는 물음과 같다는 것입니다. 그는 미나브 사건이 전쟁범죄라고 말한 것이 아니라고 분명히 했습니다. 그가 말하려 한 것은 다른 것이었습니다. 우리의 가치를 지키는 방식으로 이겨야 한다는 것, 바닥을 향한 경주는 안 된다는 것이었습니다. 이기더라도 그 과정에서 자신이 지키려던 것을 잃으면, 이긴 것이 아니라는 생각입니다. 그는 같은 인터뷰에서 이렇게도 말했습니다. 민주주의가 그런 일들을 한다면, 민주주의가 이기는 것은 그 값을 할 만한 일이 못 된다고.

무엇이 잘못될 수 있느냐는 물음에 그는 두 가지를 들었습니다. 하나는 신뢰성입니다. 기계가 엉뚱한 사람을 표적으로 삼고, 민간인을 쏘고, 사람 병사라면 발휘했을 판단을 보이지 못하는 경우입니다. 아군을 오인 사격하거나, 민간인을 치거나, 그저 엉뚱한 것을 치는 일입니다. 믿을 수 없는 것을 팔고 싶지 않다고 그는 말했습니다. 자기 나라 사람을

죽이거나 무고한 사람을 죽일 수 있는 것을 팔고 싶지 않다는 것입니다. 다른 하나는 감독입니다. 사람 병사에게는 책임의 사슬이 따라붙습니다. 그 사람이 자신의 상식을 쓰리라는 전제 위에 사슬이 서 있습니다. 그런데 천만 대의 드론을 한 사람이, 혹은 몇 사람이 조종한다면, 그 책임의 사슬은 어디에 걸리는 것일까요. 권력이 그렇게까지 한곳에 모이면 책임이 흩어진다고 그는 보았습니다.

국방부는 그 1퍼센트를 받아들이지 못했습니다.

아모데이의 말에 따르면, 협상은 빠르게 흘러갔습니다. 사흘이라는 좁은 창안에서였습니다. 몇 차례 문안이 오갔습니다. 어느 시점에 국방부는 겉으로는 앤스로픽의 조건을 받아들이는 듯한 문안을 보냈습니다. 그러나 그 안에는 "국방부가 적절하다고 판단하면"이라거나 "법에 부합하는 한"이라는 식의 단서가 붙어 있었습니다. 결국 실질적으로는 아무것도 양보하지 않은 문안이었다고 그는 말했습니다.

여기서 시간표를 정리해 봅니다. 헤그세스 장관이 모든 사용 제한을 풀라고 요구한 것이 2월 24일이었습니다. 사흘 시한이 걸린 최후통첩이었습니다. 앤스로픽이 이를 거부한 것이 2월 26일이었습니다. 미국 대통령이 작전을 승인한 것이 2월 27일, 작전이 개시된 것이 2월 28일이었습니다.

날짜를 나란히 놓으면 한 가지가 또렷해집니다. 앤스로픽이 거부한 다음 날 작전 승인이 났고, 그다음 날 미나브의 학교가 무너졌습니다.

그리고 3월 4일, 트럼프 행정부는 앤스로픽을 '공급망 위험(supply chain risk)'으로 지정했습니다.

이 지정이 무엇을 뜻하는지 아모데이는 CBS에서 설명했습니다. 공급망 위험으로 지정되면, 그 회사는 정부의 모든 시스템에서 빠져야 합니다. 그가 만난 현장의 군 장교들은 이렇게 말했다고 합니다. 이것이 없으면 작전이 6개월, 12개월, 어쩌면 그 이상 뒤쳐진다는 것입니다.

아모데이가 이 지정을 두고 분노한 데에는 이유가 있었습니다. 그가 아는 한, 공급망 위험 지정이 미국 회사에 적용된 적은 한 번도 없었습니다. 그 지정은 적성국 기업에만 쓰였습니다. 러시아 정부와의 연계가 의심되는 보안업체 카스퍼스키 랩, 중국 반도체

공급업체 같은 곳에 말입니다. 그런 회사들과 한데 묶였다는 것을 그는 "보복적이고 부적절하다"고 표현했습니다.

대통령은 앤스로픽의 "이기심"이 "미국인의 생명을 위협에 빠뜨리고 우리 군을 위태롭게 한다"고 공개적으로 비판했습니다. 트럼프는 앤스로픽을 "좌파 워크 기업"이라고도 불렀습니다. 헤그세스 장관은 군사 계약을 가진 어떤 회사도 앤스로픽과 거래할 수 없다는 취지의 글을 올렸습니다. 아모데이는 이것이 법이 실제로 정한 것보다 과장된 주장이라고 반박했습니다. 법이 정한 것은, 군사 계약을 가진 회사가 그 군사 계약의 일부로 앤스로픽을 쓸 수 없다는, 훨씬 좁은 제한이라는 것입니다.

눈여겨볼 대목이 있습니다. 앤스로픽이 거부하자, 다른 회사가 그 자리를 채웠습니다. 아모데이는 블룸버그에서 이렇게 말했습니다. 오픈시가 끼어들어, 앤스로픽이 서명하지 않으려 한 계약에 서명했다는 것입니다. 한 회사가 그 선을, 다른 회사가 넘었습니다. 계약 하나로 막을 수 있는 일에는 한계가 있다는 것을 그는 인정했습니다.

그는 이 한계를 알면서도 계약을 거부한 이유를 설명했습니다. 계약 하나로는 다른 회사가 다른 선을 그을 수 있다, 그래도 자신이 한 일은 선례를 세운 것이라고 했습니다. 좋다고 보는 사용 사례와 우려하는 사용 사례를 구별해 세상 앞에 내놓은 것이라는 뜻입니다. 그리고 그 덕분에 의회에서 초당적인 움직임이 생겨났다고 했습니다. 자신들이 우려한 일들을 법으로 금지하려는, 가드레일을 세우려는 움직임입니다. 이것을 그는 싸움이라고 부르고 싶지 않다고 했습니다. 대기업이 정부와 벌이는 싸움이 아니라, 정부가 인공지능을 어떻게 써야 하는지에 관한 토론이라는 것입니다.

여기에 오래된 역설이 있습니다. 무기를 만드는 회사는 보통 그 무기가 어디 쓰이는지를 정하지 않습니다. 보잉은 군용기를 만들지만, 그 비행기로 무엇을 할지는 군이 정합니다. 진행자가 아모데이에게 던진 비유가 그것이었습니다. 그런데 아모데이는 인공지능은 다르다고 답했습니다. 두 가지 이유에서였습니다. 하나는 기술이 너무 새롭다는 것입니다. 비행기는 오래된 기술이라 장군도 그 작동을 웬만큼 압니다. 그러나 인공지능은 너무 빨라서, 그 신뢰성과 한계를 만든 사람조차 다 알기 어렵습니다. 다른 하나는 그 모델에 '성격'이 있다는 것입니다. 어떤 일은 믿을 만하게 해내고 어떤 일은 그러지 못하는데, 그것을 가장 잘 아는 것은 만든 회사라는 주장입니다.

이 대목에서 이 책의 모순이 가장 선명하게 드러납니다.

기술을 파는 회사가 그 기술의 가장 위험한 사용에 선을 그었습니다. 한 회사가 선을 긋자, 다른 회사가 그 선을 넘어 계약을 가져갔습니다. 그래서 표적 선정은 멈추지 않았습니다. 선을 그은 회사는 정부의 보복을 받았고, 선을 넘은 회사는 계약을 얻었습니다. 이 구조 안에서 한 회사의 양심은 무엇을 멈출 수 있었을까요. 거의 아무것도 멈추지 못했습니다. 다만 한 가지 질문을 세상에 던졌을 뿐입니다.

그렇다면 정작 미나브의 학교에는 클로드가 쓰였을까요.

CBS와 블룸버그의 진행자들이 아모데이에게 정면으로 물었습니다. 블룸버그가 보도한 대로 클로드가 팔란티어의 메이븐 스마트 시스템에 실려 이란 작전의 표적 선정에 쓰였다면, 2월에 학교를 쳐서 150명 넘는 아이를 죽인 그 공습에 클로드가 한몫을 했느냐는 질문이었습니다.

아모데이는 한 가지를 먼저 짚었습니다. 자신이 아는 한, 앤스로픽은 ICE나 국경 단속 기관과 일하지 않으며 가자에서도 일하지 않는다고 블룸버그에서 말했습니다. 자사가 관여하는 일의 범위를 좁게 정한다는 것입니다. 그러나 이란 작전을 두고 던진 직접적인 질문 앞에서는 답이 달라졌습니다.

그의 대답은 모른다는 것이었습니다.

그는 이렇게 말했습니다. 우리는 이 모델들이 정확히 어떻게 쓰였는지 알 수 없다. 우리에게는 그런 접근 권한이 없다. 전쟁에서 일어나는 이런 실수는 정말로, 정말로 끔찍한 일이다. 그는 같은 인터뷰에서 이 사건이야말로 자신들이 왜 선을 그어야 하는지를 보여 준다고 말했습니다. 회사의 미래를 걸고서라도 모델의 사용 방식을 제한하려 했다는 것입니다. 그러면서 한 가지를 강조했습니다. 미나브의 경우조차 자신들의 레드라인을 어긴 사용이 아니었다는 것입니다. 클로드가 도왔을지는 몰라도, 최종 결정은 사람이 내렸다는 것입니다. 사람이 마지막 방아쇠를 당겼다는 것입니다.

진행자는 물려서지 않았습니다. 그 학교는 웹사이트가 있었다, 구글 검색으로 찾을 수 있었다, 그렇다면 클로드가, 혹은 그들이 쓴 어떤 인공지능이든, 그것을 짚어 냈어야 하지 않았느냐고 물었습니다. 그것은 전쟁에서 기술을 지름길로 쓰는 일의 더 무서운 면을 말해주는 것 아니냐고 물었습니다.

아모데이는 자신이 가지지 못한 기밀 지식에 기대야 하는 질문이라며 한발 물러섰습니다. 클로드든 다른 어떤 인공지능이든, 그것이 어떤 역할을 했는지 자신은 모른다고 거듭 말했습니다. 다만 자신들이 세운 원칙, 사람이 최종 결정을 내린다는 원칙이 지켜졌다고 믿는다고 했습니다. 만약 이 사건이 그 원칙이 왜 중요한지를 보여 주는 사례가 아니라면, 무엇이 그렇겠느냐고 그는 되물었습니다.

아모데이가 대통령에게 한마디 할 수 있다면 무슨 말을 하겠느냐는 물음에, 그는 이렇게 답했습니다. 우리는 애국적인 미국인이다. 우리가 한 모든 일은 이 나라를 위한 것이었다. 군과 앞장서 일한 것도 이 나라를 믿기 때문이었다. 그리고 두 개의 레드라인을 그은 것은, 그 선을 넘는 일이 미국의 가치에 어긋난다고 믿었기 때문이다. 정부가 공급망 위험 지정과 국방물자생산법이라는, 전례 없는 개입으로 위협했을 때, 우리는 헌법이 보장한 표현의 자유를 행사해 목소리를 냈을 뿐이다. 정부에 동의하지 않는 것이야말로 가장 미국적인 일이라고 그는 말했습니다.

이 말 안에 이 충돌의 핵심이 들어 있습니다. 그는 나라를 지키기 위해 군과 일했고, 같은 이유로 군의 요구를 거절했습니다. 같은 애국심이 두 방향으로 나뉘었습니다. 무기를 만들어 적을 막는 것도 나라를 지키는 일이고, 그 무기에 한계를 두어 나라의 가치를 지키는 것도 나라를 지키는 일이라는 것입니다. 정부는 첫 번째만을 애국으로 보았고, 그는 두 번째도 애국으로 보았습니다. 이 어긋남이 사흘 만에 한 회사를 적성국 기업의 명단 옆에 세웠습니다.

이 대답을 어떻게 읽어야 할까요.

한편으로 그는 솔직했습니다. 자신이 모르는 것을 모른다고 말했습니다. 정말로 그들에게는 클로드가 어느 좌표 위에서 어떻게 쓰였는지 들여다볼 권한이 없을 수 있습니다. 모델을 팔고 나면, 그 모델이 어디로 가는지 만든 사람도 다 알지 못합니다.

그러나 다른 한편으로, 바로 그 솔직함이 이 시대의 빈자리를 드러냅니다. 모델을 만든 회사조차 그 모델이 무슨 일에 쓰였는지 모른다면, 책임은 어디에 있는 것일까요. 만든 사람은 모른다고 합니다. 시스템에 실은 회사는 표적 선정을 도왔을 뿐이라고 합니다. 좌표를 입력한 시스템은 데이터베이스를 따랐을 뿐입니다. 데이터베이스는 10년 전 정보를 충실히 보관했을 뿐입니다. 최종 결정을 내린 사람은 시스템이 내놓은 결과에 서명했을 뿐입니다. 그리고 그 결과는 24시간에 천 개씩 쏟아지고 있었습니다.

사람이 최종 결정을 내렸다는 말은, 들으면 안심이 되는 말입니다. 기계가 아니라 사람이 방아쇠를 당겼다니. 그러나 그 사람에게 주어진 시간이 표적 하나당 1분이었다면, 그리고 기계가 내놓은 표적의 정확도가 60퍼센트였다면, 그 '사람의 최종 결정'이라는 것은 얼마나 두꺼운 방패일까요. 그것은 책임을 지는 사람의 자리일까요, 아니면 책임이 흘러가다 잠시 멈췄다 가는 자리일까요.

아모데이는 블룸버그에서 또 다른 장면을 그렸습니다. 만약 완전 자율무기가 허용된 세상이었다면, 인공지능이 결정을 내리고 사람은 그것을 보지도 못하는 세상이었다면, 미나브 같은 일이 어떻게 됐겠느냐는 것입니다. 그가 두려워한 것은 그 세상이었습니다. 사람이 마지막에 들여다보기라도 하는 지금이, 사람이 아예 빠진 그 세상보다는 낫다는 것입니다.

그는 정부가 마이크로소프트 엑셀을 쓰는 것에 빗대기도 했습니다. 정부는 엑셀을 많이 씁니다. 그렇다고 엑셀을 만든 회사가 "이 군사 작전에는 엑셀을 써도 되고 저 작전에는 쓰면 안 된다"고 정할 수는 없다는 것입니다. 도구를 만든 사람이 도구의 모든 쓰임을 정할 수는 없다는 논리입니다. 들어 보면 일리가 있습니다. 표 계산 프로그램을 만든 회사에게 전쟁의 책임을 물을 수는 없습니다.

그러나 클로드는 엑셀이 아닙니다. 엑셀은 사람이 입력한 숫자를 더하고 뺄 뿐, 누구를 칠지 골라 주지 않습니다. 클로드가 실린 메이븐 스마트 시스템은 표적을 골라 사람 앞에 내놓았다고 블룸버그는 보도했습니다. 더하고 빼는 도구와, 사람을 골라 내놓는 도구 사이에는 건너기 어려운 강이 있습니다. 아모데이 자신이 그 강의 존재를 알기에 두 개의 레드라인을 그었을 것입니다. 엑셀이라면 레드라인이 필요 없습니다.

그의 말이 옳을 수 있습니다. 사람이 빠진 세상은 더 위험할 것입니다. 그러나 미나브의 교실은 이미 사람이 빠지지 않은 세상에서 일어났습니다. 사람이 최종 결정을 내린다는 원칙이 살아 있는 세상에서, 175명이 죽었습니다. 사람이 거기 있었습니다. 1분 동안.

이 장을 닫으며 한 가지를 남겨 둡니다. 미국 한 관리는 인공지능의 도움으로 미군이 하루에 칠 수 있는 표적이 1,000개에서 5,000개로 늘었다고 말했다고 블룸버그는 전합니다. 진행자는 아모데이에게 물었습니다. 그렇다면 클로드가 더 많은 사람을 더 빨리 죽이는 데 도움이 된다는 뜻인데, 그 점이 불편하지 않으냐고.

아모데이는 미국이 군사적으로 더 효과적이 되는 능력을 자신은 지지한다고 답했습니다. 그 능력이 강해지는 것은 전쟁을 일으키는 것이 아니라 전쟁을 억제하는 것이라고 했습니다. 자신은 애국자라고 다시 한번 말했습니다. 동시에 그는 분명히 했습니다. 기술을 제공하는 것은 자신들이지만, 그 기술로 어떤 군사 작전을 해도 되고 어떤 작전은 안 되는지를 정하는 것은 자신들의 몫이 아니라는 것입니다. 정책은 군사 결정권자의 손에 맡겨야 한다는 것입니다.

기술을 만든 사람은 정책을 정하지 않습니다. 정책을 정하는 사람은 기술을 만들지 않습니다. 좌표를 입력하는 시스템은 책임을 지지 않습니다. 데이터베이스는 의심하지 않습니다. 그리고 그 사이 어딘가에, 175개의 빈자리가 생겼습니다.

이 장에서 우리가 본 것을 다시 한곳에 모아 봅니다. 표적 분류 정확도 60퍼센트의 시스템이, 사람보다 덜 정확한 채로 표적을 골랐습니다. 그 시스템이 날씨가 나쁘면 정확도가 30퍼센트 아래로 떨어졌습니다. 그런데도 그 시스템은 전장에 올라가 한 나라의 지도부와 한 학교의 교실을 같은 화면 위에 표적으로 올렸습니다. 완성되지 않은 기술이, 실험실이 아니라 실제 전쟁에서 시험되고 있었던 것입니다.

이 책은 이런 상태를 'lethal beta'라고 부릅니다. 우리말로 옮기면 치명적인 시험판쯤 됩니다. 소프트웨어를 만들 때, 정식 출시 전에 베타 버전을 내놓고 사람들이 써 보게 합니다. 버그가 나오면 고칩니다. 베타란 아직 완성되지 않았다는 표시입니다. 그런데 표적 선정 시스템의 베타 버전에서 버그가 나면, 고쳐지는 것은 다음 판입니다. 버그의 값을 치르는 것은 그 좌표 위에 있던 사람입니다. 미나브의 학교는 한 베타 버전의 버그였습니다. 정확도 60퍼센트라는 버그, 10년 된 좌표라는 버그, 1분이라는 시간의 버그였습니다. 다만 이 버그는 패치로 되돌릴 수 없습니다. 빈 책상은 다음 판에서 채워지지 않습니다.

아모데이는 옳은 일을 하려 애썼는지도 모릅니다. 그는 선을 그었고, 그 선 때문에 회사가 정부의 보복을 받았습니다. 그는 솔직했고, 모르는 것을 모른다고 말했습니다. 동시에 그는 기술을 팔았고, 그 기술이 어디 쓰이는지 정하는 것은 자기 몫이 아니라고 했습니다. 이 두 가지가 한 사람 안에 함께 있습니다. 우리는 그를 영웅으로도 악당으로도 그릴 수 있을 것입니다. 그러나 둘 다 틀린 그림입니다. 그는 빈자리를 가장 가까이에서 들여다본 사람이고, 그 빈자리를 채우지 못한 사람입니다. 어쩌면 한 사람의 양심으로는 채울 수 없는 자리인지도 모릅니다.

누구의 자리였을까요. 책임이 앉아 있어야 했던 그 자리는.

이란의 다음 장에서 우리는 이 빈자리가 한 나라의 국경을 넘어 어떻게 또 다른 전장으로 옮겨 가는지를 보게 됩니다. 베네수엘라로. 그곳에서도 같은 시스템이, 같은 속도로, 같은 질문을 남깁니다.

기계의 속도에 갇힌 60초와 175개의 빈 책상

의사결정 시간이 분 단위로 압축되며 인간의 마지막 통제권은 유명무실해졌습니다.



Funnel **Bottleneck** **Fallout - The Void**

- 2026년 2월 28일 이란 '에픽 퓨리' 작전, 24시간 동안 표적 1,000개 처리.
- 메이븐 시스템과 클라우드 모델 결합, 기계의 표적 분류 정확도는 60% 수준 기록.
- 10년 전 갱신이 멈춘 낡은 좌표 오류로 미나브 초등학교 학생과 교사 175명 사망.

NotebookLM

제4장 베네수엘라: AI가 개입한 참수작전

1. 절대적 결의 작전: 마두로 포획의 밤

카라카스의 불이 한꺼번에 꺼졌습니다.

2026년 1월 3일 밤이었습니다. 도시 위로 헬기가 내려왔습니다. 지붕에서 100피트, 그러니까 30미터 남짓 되는 높이였습니다. 저공으로 도시를 가르는 회전날개 소리가 먼저 들렸고, 그 다음에 정전이 왔습니다. 신호등이 죽었습니다. 관공서 건물의 창문이 검게 변했습니다. 카라카스 외곽 포르테 티우나(Fort Tiuna), 베네수엘라 국방부가 자리한 군사단지의 방공망 레이더도 같은 순간 화면에서 사라졌습니다.

작전 직전에 사이버공격이 카라카스의 전력망과 방공망을 마비시켰다고 보도는 전합니다. 전기가 끊긴 도시 위로 항공기가 들어오는데, 그 항공기를 잡아야 할 레이더가 먼저 눈을 감은 것입니다. 미국 합참의장 댄 케인(Dan Caine)은 뒤에 이 작전을 두고 "완전한 기습"이라는 표현을 썼습니다.

기습이라는 말은 적이 모른다는 뜻입니다. 적이 모르게 하려면 적의 눈을 미리 가려야 합니다. 그 밤에 카라카스의 눈은 전기가 끊기는 순간 함께 감졌습니다. 도시는 무엇이 다가오는지 보지 못했습니다.

포르테 티우나라는 곳을 잠깐 그려보겠습니다. 이곳은 카라카스 안에 있는 베네수엘라 군의 심장입니다. 국방부가 있고, 사령부가 있고, 군 병원과 막사가 모여 있습니다. 마두로의 측근들이 콜롬비아 무장혁명군 지도자를 데려와 만나던 곳도 이 단지였다고 기소장은 적고 있습니다. 한 나라의 군사력이 두텁게 모인 자리, 외부의 손이 닿기 어려운 자리. 그곳이 1월 3일 밤에 뚫렸습니다. 나라에서 제일 단단한 곳이 뚫렸다는 사실이 이 작전의 성격을 말해줍니다. 카라카스의 보통 거리가 아니라, 군의 가장자리도 아닌 그 깊은 안쪽까지 외국 특수부대가 들어갔다는 것입니다. 그 안쪽으로 가는 길을 열어준 것이 사이버공격과 통합된 정보였습니다.

그날 밤 미국 항공기 150대가 떴습니다. 출격 기지는 스무 곳이었습니다.

150대.

스무 곳.

수면 위 100피트.

이 숫자들을 한 호흡에 몰아넣으면 그저 큰 작전이라는 인상만 남습니다. 끊어서 보면 다릅니다. 150대의 항공기가 한날한시에 뜨려면 스무 개의 기지가 미리 짜놓은 순서대로 움직여야 합니다. 어느 기지가 먼저 뜨고, 어느 항로로 들어가고, 어느 지점에서 합류하는지가 어긋나면 150대는 서로 부딪힙니다. 그 순서를 짜고 어긋남을 막는 일은 사람의 머릿속 종이 한 장으로는 감당이 안 됩니다.

NBC 보도에 따르면 미국 중앙정보국(CIA)은 그 전해 8월부터 베네수엘라 안으로 소규모 팀을 잠입시켜 마두로의 동선을 추적하고 있었습니다. 다섯 달 동안 한 사람이 언제 어디서 자고 언제 움직이는지를 지켜본 셈입니다. 그가 어느 관저에서 며칠을 묵고, 어느 경로로 이동하고, 곁에 누가 있는지가 다섯 달치 기록으로 쌓였습니다.

그리고 1월 3일 밤 10시 46분, 미국 동부시간으로 그 시각에 트럼프 대통령이 최종 발사 명령을 내렸다고 같은 보도는 적고 있습니다.

10시 46분.

한 사람의 명령이 떨어지는 그 분 단위의 시각 뒤에는 다섯 달의 추적과 스무 기지의 대기가 깔려 있었습니다. 명령은 짧고, 그 명령을 가능하게 한 준비는 길었습니다.

미국 남부사령부(SOUTHCOM)와 합동기동부대가 짜놓은 큰 틀의 이름은 '남부의 창 작전(Operation Southern Spear)'이었습니다. 그 안에서 마두로 한 사람을 잡기 위한 작전이 따로 굴러갔는데, 보도에 등장한 이름은 '절대적 결의 작전(Operation Absolute Resolve)'입니다. 결의라는 이름이 붙은 작전이었습니다. 결의는 사람의 마음에 있는 것인데, 그것을 실행에 옮긴 것은 스무 기지의 항공기와 한 화면 위에 모인 정보였습니다.

이 장의 제목에 '참수작전'이라는 말을 넣었습니다. 군사 용어로 참수작전은 적의 머리, 곧 지도부를 한 번에 제거해 조직 전체를 마비시키는 작전을 가리킵니다. 몸통을 일일이 상대하지 않고 머리만 친다는 발상입니다. 마두로 한 사람을 잡으면 그가 정점에 있다는 '카르텔' 전체가 무너진다는 논리가 여기에 깔려 있습니다. 그런데 앞에서 봤듯이 그 '카르텔'의 실체가 흐릿하다면, 머리를 쳤을 때 무너질 몸통이 무엇인지도 흐릿해집니다. 참수작전은 머리와 몸통이 분명할 때 성립합니다. 머리만 또렷하고 몸통이 별칭이라면, 그

작전이 친 것은 무엇이였을까요.

저공비행 100피트라는 숫자도 그냥 나온 것이 아닙니다. 레이더는 낮게 나는 물체를 잘 잡지 못합니다. 지형에 가려지기 때문입니다. 150대의 항공기가 수면 위 30미터로 들어왔다는 것은, 잡히지 않으려는 비행이었다는 뜻입니다. 거기에 사이버공격으로 방공망까지 미리 마비했다면, 카라카스는 두 겹으로 눈이 가려진 셈입니다. 한 겹은 낮게 나는 항공기를 못 보는 레이더의 한계였고, 다른 한 겹은 아예 꺼져버린 레이더였습니다. 이 두 겹을 함께 설계하는 일이 통합 작전입니다. 항공기의 고도와 사이버공격의 시각이 따로 놓면 기습은 깨집니다. 둘을 한 화면 위에서 맞추는 일, 그것이 다시 그 유리판의 몫입니다.

델타포스가 포르테 티우나의 관저로 진입했습니다. 보도에 따르면 마두로는 강철 안전실까지 가닿지 못했습니다. 강철 문 앞에서 몇 초가 모자랐다는 이야기입니다. 한 나라의 통치자가 자기 관저 안에서 안전실의 문을 향해 달렸고, 그 문에 손이 닿기 전에 붙잡혔습니다. 그는 부인 실리아 플로레스(Cilia Flores)와 함께 체포됐습니다.

여기서 잠시 멈추겠습니다. 강철 문 앞의 몇 초. 그 몇 초가 한 사람의 운명을 갈랐다는 이 장면은 검증된 사실이 아니라 보도가 전한 정황입니다. 그가 정확히 어디서 어떻게 붙잡혔는지, 안전실 문까지의 거리가 몇 걸음이었는지는 작전을 수행한 쪽과 보도한 쪽의 설명에 기대고 있습니다. 단정하지 않겠습니다. 다만 보도가 그린 그림은 분명합니다. 다섯 달 동안 한 사람의 동선이 추적됐고, 그 추적의 끝에서 그는 강철 문에 닿지 못했습니다.

두 사람은 뉴욕주 스투어트 주방위공군기지로 압송됐습니다. 그리고 마약테러 혐의로 기소됐습니다. 한 나라의 사실상 통치자가 한밤중에 자기 나라의 군사단지 안에서 외국 특수부대에 붙잡혀, 새벽이 오기 전에 다른 대륙의 군 기지에 내려졌습니다.

니콜라스 마두로가 어떤 사람인지 잠깐 짚어두는 편이 낫겠습니다. 그는 버스 운전사 출신으로 노동조합 활동을 거쳐 정치에 들어섰습니다. 우고 차베스 밑에서 외무장관과 부통령을 지냈고, 2013년 차베스가 죽자 대통령직을 이어받았습니다. 2018년 선거와 2024년 선거 모두 국제사회가 부정선거라 비판했고, 미국을 포함한 50개국 넘는 나라가 그를 정당한 국가원수로 인정하지 않았습니다. 이 책이 그의 통치를 옹호하려는 것이 아닙니다. 다만 그가 누구인지, 그를 둘러싼 평가가 어떻게 갈리는지를 적어두지 않으면, 1월 3일 밤의 작전을 한쪽 눈으로만 보게 됩니다. 그는 많은 베네수엘라 사람에게 억압의 얼굴이었고, 동시에 미국에게는 오래 잡고 싶었던 표적이었습니다. 이 두 가지는 함께 참일

수 있습니다.

다섯 달의 추적이라는 대목을 다시 봐야 합니다. CIA가 2025년 8월부터 잠입 팀을 운영했다면, 그 다섯 달 동안 마두로의 일상은 한 줄 한 줄 데이터로 쌓였습니다. 그가 어느 건물에서 자는지, 경호가 어느 시각에 교대하는지, 어느 방의 불이 언제 꺼지는지. 이런 기록은 사람의 눈으로 모읍니다. 그러나 그 기록을 한데 묶어 패턴을 읽고, 어느 밤이 제일 빈틈이 큰지를 가려내는 일은 이제 사람의 머리만으로 하지 않습니다. 다섯 달치 동선을 한 화면에 겹쳐놓고 보면, 한 사람의 삶에서 규칙이 드러납니다. 규칙이 드러나면 빈틈도 드러납니다. 1월 3일 밤은 그 빈틈이었을 것입니다.

이 장면을 어떻게 불러야 할지부터가 어렵습니다. 미국 정부는 이것을 마약테러 용의자에 대한 법집행으로 봅니다. 마두로 측과 베네수엘라 정부는 주권국가의 통치자를 외국 군대가 납치한 것이라고 봅니다. 양쪽의 규정이 정면으로 충돌합니다. 한쪽에서는 체포이고, 다른 한쪽에서는 침공입니다. 이 책은 그 규정의 옳고 그름을 판가름하는 자리가 아닙니다. 다만 두 규정이 같은 밤의 같은 행위를 가리킨다는 사실은 분명히 적어둡니다.

이 책이 쫓아온 흐름은 표적을 고르는 일이 사람의 속도에서 기계의 속도로 넘어간 과정이었습니다. 가자에서는 라벤더라는 시스템이 사람을 명단에 올리고, 정보장교가 20초 만에 도장을 찍었습니다. 우크라이나에서는 18분 만에 죽음의 사슬이 닫혔습니다. 이란에서는 60초가 한 사람의 운명을 갈랐습니다.

베네수엘라의 1월 3일 밤은 그 사슬의 또 다른 모습입니다.

여기서는 미사일이 한 사람의 머리 위로 떨어지지 않았습니다. 표적은 죽이는 대신 붙잡혔습니다. 그래서 이 작전은 앞선 장들과 결이 다릅니다. 가자와 이란에서 우리가 본 것은 사람이 죽는 속도였습니다. 베네수엘라에서 본 것은 사람을 잡는 정밀함입니다. 그러나 그 정밀함을 떠받친 구조는 같습니다. 흩어진 정보를 한 화면에 모으고, 누가 어디 있고 무엇을 봐야 하는지를 정리하는 일. 다섯 달의 추적과 스무 기지에서서의 출격과 전력망을 끊는 사이버공격은 하나의 통합된 그림판 위에서 움직였습니다.

그 그림판이 이 장의 진짜 주인공입니다.

죽이지 않고 잡았다는 점이 이 작전을 깨끗하게 보이게 할 수도 있습니다. 사상자 보도가 앞선 장들만큼 크게 나오지 않았다는 점도 그렇습니다. 하지만 깨끗해 보이는 것과 책임이 분명한 것은 다릅니다. 한 사람을 다섯 달 동안 추적하고, 그의 나라의 전력망을 끊고, 스무 기지에서 항공기를 띄워 한밤중에 그를 끌어내는 일에는 수많은 판단이 들어갑니다. 그 판단의 어느 단계까지가 사람의 것이고, 어느 단계부터가 화면 위 도구의 것인지를 가르는 일이 점점 어려워집니다.

여기에 하나를 더 적어야 합니다. 카라카스의 전력망이 끊겼다는 것은 마두로의 관저만 어두워졌다는 뜻이 아닙니다. 도시 전체가 어두워졌습니다. 신호등이 죽은 거리에는 병원도 있고, 산소호흡기에 의지한 환자도 있고, 한밤중에 깨어 무슨 일인지 몰라 두려워하는 사람도 있었습니다. 한 사람을 잡기 위해 도시 하나의 불을 끄는 일. 그 사이에 끼인 수백만 명은 작전의 표적이 아니었지만 작전의 영향을 받았습니다. 이들의 그 밤은 어느 기록에도 표적으로 남지 않습니다. 사상자 집계에도 들지 않습니다. 그러나 한 도시가 영문도 모른 채 어둠 속에 던져졌다는 사실은, 정밀하다는 작전의 돌레에 늘 따라붙는 그림자입니다. 이 그림자의 크기는 아직 어느 매체도 또렷이 집계하지 못했습니다. 확인되지 않은 범위로 남겨둡니다.

그 그림판의 이름은 곧 나옵니다. 먼저 마두로를 잡은 명분부터 들여다봐야 합니다.

2. 명분의 붕괴: 태양의 카르텔이라는 허구

명분은 종이 한 장에서 시작합니다.

미국 뉴욕 남부지방법원에 제출된 기소장입니다. 처음 제출된 것은 2020년 3월이었습니다. 마두로가 붙잡히기 거의 여섯 해 전에, 그를 잡을 법적 근거가 종이 위에 먼저 마련된 셈입니다. 표지에는 미합중국 대 니콜라스 마두로 모로스 외 여섯 명이라고 적혀 있습니다. 대배심은 이렇게 적었습니다.

"25년 넘게, 베네수엘라의 지도자들은 공적 신뢰의 자리를 남용하고 한때 정당했던 기관들을 부패시켜 미국으로 수 톤의 코카인을 들여왔다."

기소장은 마두로를 그 부패의 맨 앞에 세웠습니다. 그가 국회의원이던 시절 베네수엘라 법집행기관의 보호 아래 코카인을 옮겼다고 적었습니다. 외무장관이던 시절에는 마약 밀매업자에게 외교 여권을 내주고, 멕시코에서 베네수엘라로 마약 대금을 돌려보내는 비행기에 외교적 엄호를 제공했다고 적었습니다. 그리고 대통령이자 지금은 사실상의 통치자가 되어, 코카인이 부추긴 부패가 자기 자신과 측근과 가족의 이익을 위해 번성하도록 내버려두었다고 적었습니다.

기소장은 구체적인 장면을 줄줄이 늘어놓습니다. 2006년 DC-9 제트기 한 대에 5.5톤이 넘는 코카인을 실어 멕시코로 보냈고, 그 짐은 멕시코 캄페체의 시우다드 델 카르멘 공항에서 압수됐다는 대목이 있습니다. 그 비행기가 마이케티아 공항의 대통령 전용 격납고에서 다섯 대의 승합차에 실린 코카인을 받아 뒀다는 묘사까지 들어 있습니다. 마두로의 부인 플로레스가 2007년 한 회의에서 수십만 달러의 뇌물을 받고 대형 마약 밀매업자와 반마약청장 사이의 만남을 주선했다는 대목도 있습니다. 마두로의 아들 니콜라스 에르네스토 마두로 게라, 별칭 '니콜라시토' 또는 '왕자'가 2020년 메데인에서 콜롬비아 무장혁명군(FARC) 대표들을 만나 코카인과 무기를 향후 6년간, 그러니까 2026년까지 옮기는 방안을 논의했다는 대목도 있습니다.

2026년. 기소장에 적힌 이 연도가 공교롭습니다. 종이 위에서 2026년까지의 계획이 거론됐고, 실제로 2026년 첫 사흘 만에 작전이 떨어졌습니다. 우연인지 의도인지는 단정하지 않겠습니다.

기소장의 문장은 검찰의 확신을 그대로 드러냅니다. 마두로를 두고 대배심은 이렇게 적었습니다.

"니콜라스 마두로 모로스, 피고인은 그 부패의 맨 앞에 서서, 자신의 공모자들과 손잡고 불법으로 얻은 권한과 자신이 부식시킨 기관들을 이용해 수천 톤의 코카인을 미국으로 운반했다."

부식시켰다는 단어를 검찰은 골랐습니다. 기관이 녹슬 듯 안에서부터 무너졌다는 뜻입니다. 한 사람이 나라의 제도를 안에서 갹아먹었다는 그림입니다. 이런 문장은 법정에서 입증을 요구합니다. 누가, 언제, 어떤 제도를, 어떻게 부식시켰는지가 증거로 뒷받침돼야 합니다. 기소장은 그 증거라며 스무 개가 넘는 구체적 행위를 'a'부터 's'까지 알파벳을 매겨 늘어놓습니다. 어떤 항목은 날짜와 금액이 또렷하고, 어떤 항목은 '약 2006년에서 2008년 사이'처럼 폭이 넓습니다. 또렷함과 흐릿함이 한 문서 안에 섞여 있습니다.

혐의는 세 갈래입니다. 마약테러 공모, 코카인 수입 공모, 그리고 기관총과 파괴장치 소지. 기소장에 적힌 죄목 가운데 코카인 수입 공모는 다섯 킬로그램 이상일 경우 미국 법에 따라 최소 형량이 20년입니다.

20년.

종이 위의 혐의는 무겁고 구체적입니다. 무게와 기종과 날짜와 액수가 적혀 있습니다. 그런데 이 명분의 한복판에 한 단어가 박혀 있습니다. '태양의 카르텔'입니다.

기소장은 이 말을 이렇게 설명합니다. 부패한 베네수엘라 엘리트들이 마약 밀매로 자기 배를 불리는 후견 체제를 운영하는데, 그 정점에 있는 자들을 가리켜 "카르텔 데 로스 솔레스, 곧 태양의 카르텔이라 부른다"고 적었습니다. 그러고는 괄호를 열어 한 줄을 덧붙입니다. 그 이름은 "베네수엘라 고위 군 장교의 군복에 부착된 태양 문양을 가리키는 것"이라고요.

군복에 달린 태양 문양. 바로 이 대목이 명분의 약한 고리입니다.

베네수엘라 군에서 장성급 계급장에는 태양 문양이 붙습니다. 별 대신 태양입니다. 1990년대부터 베네수엘라에서 '태양의 카르텔'이라는 말은 마약에 손을 댄 부패한 군 장성들을 두루 가리키는 관용어로 쓰였습니다. 특정한 우두머리가 있고 조직표가 있고

본부가 있는 단일한 범죄조직의 이름이 아니라, 같은 종류의 부패한 장성들을 한데 묶어 부르는 별칭에 가까웠습니다. 어떤 사람이 부패한 장성이면 그를 '태양의 카르텔'에 속한다고 부를 수 있었습니다. 그러나 그 말이 한 조직의 가입 명부를 가리키는 것은 아니었습니다.

이 차이가 사소해 보일지 모릅니다. 그러나 법의 세계에서는 사소하지 않습니다. 어떤 사람을 마약테러 조직의 우두머리로 기소하려면, 그 조직이 실재해야 합니다. 조직에는 구조가 있어야 하고, 그가 그 구조의 어디에 있는지가 특정돼야 합니다. 별칭은 그런 구조를 갖지 않습니다. '부패한 장성들'이라는 느슨한 묶음에는 우두머리가 없습니다. 그런데 미국 정부는 2025년 들어 '카르텔 데 로스 솔레스'를 외국 테러조직으로 지정하는 절차를 밟았고, 마두로를 그 정점으로 지목했습니다. 관용어가 조직이 되고, 조직에 우두머리가 생기고, 그 우두머리를 잡는 작전이 정당화되는 흐름입니다. 흐름의 출발점에 종이 한 장이 있었습니다.

기소장이 그린 인물들의 면면을 보면 이 사건이 한 사람의 문제가 아님을 알 수 있습니다. 마두로의 측근 디오스다도 카베요 론돈, 전 내무장관 라몬 로드리게스 차신, 부인 실리아 플로레스, 아들 마두로 게라, 그리고 베네수엘라 교도소에서 출발한 범죄조직 트렌 데 아라과(TdA)의 지도자 헥토르 게레로 플로레스까지. 기소장은 이들이 콜롬비아 무장혁명군, 민족해방군(ELN), 시날로아 카르텔, 제타스 같은 무장 마약조직과 손잡고 수십 년에 걸쳐 수천 톤의 코카인을 미국으로 들여왔다고 적었습니다. 2020년까지 베네수엘라를 거쳐 흐른 코카인이 연간 200톤에서 250톤으로 추정된다는 국무부 추산도 들어 있습니다.

그렇다면 기소장이 그린 '카르텔'은 실재하는 조직일까요, 아니면 관용어에 조직의 외피를 씌운 것일까요.

진보 성향 매체 커먼 드림스(Common Dreams)의 보도에 따르면, 미국 법무부 자신이 법정에서 태양의 카르텔이 실재하는 단일 조직이 아님을 사실상 인정했습니다. 종이 위에서는 하나의 카르텔처럼 호명되던 것이, 법정의 검증 앞에서는 그렇게 또렷한 실체로서 있지 못했다는 것입니다. 적을 규정한 이름과 그 이름의 실체 사이에 틈이 벌어진 것입니다.

이 대목에서 매체에 따라 결이 갈립니다. 미국 정부 쪽 자료와 이를 따르는 보도는 태양의 카르텔을 마두로가 이끄는 실재 조직으로 다룹니다. 반대편 보도와 일부 베네수엘라 전문가는 그것이 조직이라기보다 부패의 양상을 부르는 이름이라고 봅니다. 한쪽은 조직이라 하고, 다른 한쪽은 현상이라 합니다. 이 두 견해가 충돌한다는 사실 자체를 적어두는 것이 정확합니다. 어느 한쪽으로 봉합하면 사실을 깎는 일이 됩니다. 분명한 것은 하나입니다. 한 사람을 다른 대륙으로 끌고 가 처벌하는 작전의 명분 한복판에, 실체를 두고 견해가 갈리는 이름이 놓여 있었다는 것입니다.

이 틈이 왜 중요한가. 한 사람을 다른 나라에서 끌고 와 처벌하려면, 그가 무엇의 일원인지가 분명해야 합니다. 그가 단일 범죄조직의 우두머리라면 그 조직을 해체하기 위한 작전이 정당화됩니다. 그러나 그가 부패한 장성들을 두루 가리키는 별칭 아래 묶인 한 사람이라면, 이야기는 달라집니다. 별칭은 우두머리를 갖지 않습니다. 관용어를 해체할 수는 없습니다. 그런데 작전은 한 사람을 정점으로 지목하고 그를 끌어냈습니다.

여기서 양쪽을 다 놓아야 공정합니다.

기소장에 적힌 개별 사실들, 압수된 코카인의 무게와 비행기의 기종과 뇌물의 액수는 미국 검찰이 증거를 갖췄다고 주장하는 구체적 혐의입니다. 이 가운데 일부는 법정에서 유죄가 확정되기도 했습니다. 기소장 자체가 마두로의 친척 두 사람이 2016년 뉴욕에서 코카인 수입 공모로 유죄 평결을 받았다고 적고 있습니다. 베네수엘라 군 정보기관의 전 책임자가 2025년 6월 같은 법원에서 마약테러와 무기 혐의에 유죄를 인정했다는 대목도 있습니다. 종이 위의 혐의가 통째로 허구라는 뜻은 아닙니다.

마두로 측은 이 기소 자체를 정권 교체를 노린 정치적 표적이라고 반박해 왔습니다. 미국이 베네수엘라의 석유와 정권에 손을 대려고 마약을 구실 삼았다는 주장입니다. 그리고 '태양의 카르텔'이라는 호명이 단일 조직의 실체와 어긋난다는 지적은 그 명분의 단단한 부분이 아니라 무른 부분을 건드립니다. 개별 혐의의 무게와, 그 혐의를 하나의 조직 이름으로 묶어낸 틀의 험거움은 별개의 문제입니다. 돌을 섞으면 안 됩니다.

한 사람을 한밤중에 다른 대륙으로 끌고 가는 작전에는 그 작전을 정당화하는 이름이 필요합니다. 가자에서는 '하마스 요원'이라는 이름이, 이란에서는 '핵 위협'이라는 이름이 그 자리에 놓였습니다. 베네수엘라에서는 '태양의 카르텔'이 그 자리에 놓였습니다.

명분의 이름이 실체보다 먼저 정해지고, 그 이름 아래로 작전이 굴러갈 때, 사라지는 것은 다시 한번 책임의 소재입니다. 누가 이 사람을 적으로 규정했는가. 그 규정은 어디까지 검증됐는가. 검증되지 않은 부분의 무게는 누가 지는가.

2020년의 종기와 2026년의 작전 사이에는 여섯 해의 시차가 있습니다. 이 시차가 중요합니다. 기소장이 제출된 2020년에는 마두로를 잡을 수단이 없었습니다. 종기 위의 혐의는 종기 위에 머물렀습니다. 그를 잡으려면 그가 어디 있는지 알아야 하고, 그곳에 닿아야 하고, 그곳의 방어를 뚫어야 합니다. 2020년에는 그 셋이 다 어려웠습니다. 그런데 2026년에는 셋 다 가능해졌습니다. 다섯 달의 추적이 그가 어디 있는지를 알려줬고, 스무 기지의 항공기가 그곳에 닿았고, 사이버공격이 방어를 뚫었습니다. 명분은 여섯 해 전에 마련됐지만, 그 명분을 집행할 능력은 최근에야 갖춰졌습니다. 카프가 "2년 전, 3년 전에는 할 수 없던 일"이라고 한 말이 여기서 다시 울립니다. 종기는 오래전에 준비됐고, 그 종기를 현실로 바꾼 것은 새로 갖춰진 기술이었습니다.

이 순서를 뒤집어 생각하면 불안한 그림이 나옵니다. 잡을 수 있게 되었기 때문에 잡은 것은 아닌가. 명분이 충분히 무르익어서가 아니라, 집행할 도구가 무르익어서 작전이 떨어진 것은 아닌가. 단정하지 않겠습니다. 다만 명분과 능력 사이의 시차는, 능력이 명분을 끌어당기는 시대가 왔음을 보여줍니다. 할 수 있는 일이 늘어나면, 해도 되는지를 묻는 목소리는 작아집니다.

이름은 사람이 붙였습니다. 2020년의 기소장에 사람의 손으로 '태양의 카르텔'이라는 말이 적혔습니다. 그러나 그 이름 아래로 표적을 좁히고, 동선을 잇고, 출격의 순서를 짜는 일에는 이제 사람만 있는 것이 아닙니다.

3. 균열: 앤스로픽과 팔란티어, 그리고 모순

화면 하나를 상상해 보십시오.

작전실의 한 화면입니다. 그 위에 흩어진 정보가 한데 모입니다. 위성이 본 것, 신호가 잡아낸 것, 다섯 달 동안 잠입 팀이 적어 보낸 것이 한 장의 유리판 위로 겹쳐집니다. 팔란티어는 이것을 '하나의 유리판(plane of glass)'이라고 부릅니다. 정보의 출처는 제각각입니다. 어떤 것은 너무 민감해 아무도 볼 수 없고, 어떤 것은 모두가 봐야 합니다. 동맹국이 함께 들여다보는 정보도 있습니다. 그 모든 것을 보안을 유지한 채 한 화면에 올리는 일이 팔란티어가 파는 것입니다.

마두로 작전이 굴러가는 동안, 그 유리판 위에 앤스로픽의 클로드가 떠 있었습니다. 시맨포(Semafor)의 보도입니다.

팔란티어가 무엇을 파는 회사인지부터 들어보는 편이 낫습니다. 회사의 공동창업자이자 최고경영자인 알렉스 카프(Alex Karp)는 한 방송 인터뷰에서 자기 회사의 플랫폼을 이렇게 설명했습니다.

"팔란티어의 메이븐(Maven) 플랫폼은 값진 거대언어모델을 가져다가, 그것을 실제로 전장에서 살상력 있고 쓸모 있게 만듭니다."

거대언어모델을 전장에서 살상력 있게 만든다. 카프는 그 말을 예뉘러 가지 않았습니다. 같은 인터뷰에서 그는 회사의 목표를 미국 병사를 안전하게 집에 데려오는 것이라 했고, 그러려면 적을 "해체"해야 한다고 했습니다. 그는 베네수엘라를 직접 언급하며, 미국이 "이런 것들을 해체하는 능력"을 보여줬다고 말했습니다.

"미국의 성과를 보면, 미국인의 용기와 독창성 덕분에, 그리고 우리 기술 덕분이라고 믿고 싶은데, 우리 쪽 손실은 적고, 정밀하고, 통제됐습니다. 베네수엘라에서 보셨듯이 이런 것들을 해체하는 완전한 능력어요. 2년 전, 3년 전에는 할 수 없던 일입니다."

2년 전에는 할 수 없던 일. 카프 자신의 입으로 나온 말입니다. 표적을 좁히고 동선을 잇고 작전을 한 화면 위에서 굴리는 일이 최근 몇 해 사이에 가능해졌다는 것입니다. 그 화면 위에 떠 있던 것이 바로 앤스로픽의 클로드였습니다.

카프는 메이븐이 무엇을 하는지를 우크라이나의 예로 풀었습니다. 그는 드론을 목표 지점까지 보내는 일만이 아니라, 그 사이에 놓인 진보된 레이더 체계를 뚫고 지나가는 일이 진짜 어려움이라고 했습니다. 흩어진 옛 시스템을 새 시스템으로 잇고, 보안 등급이 제각각인 정보를 한 화면에 모으고, 그 위에 인공지능과 기계학습을 얹어 무엇을 봐야 하는지를 가려내는 일. 그것이 메이븐이라고 그는 설명했습니다. 그리고 이란을 두고는 "100배 더 복잡하다"고 했습니다. 적의 방공이 그만큼 촘촘하기 때문이라는 것이었습니다.

카프는 인공지능에도 등급을 매겼습니다. 그는 전장에서 쓸모없는 인공지능을 두고 "험거운 AI 코드"라고 불렀습니다. 그런 것은 실제 전장 어디에도 쓰이지 않는다고 했습니다. 그의 구분에 따르면, 거대언어모델은 그 자체로는 확률에 기대는 도구입니다. 투자 판단처럼 51퍼센트만 맞으면 되는 일에는 쓸 만하지만, 적의 머리 위에 정확히 무기를 얹고 미국 병사를 안전하게 데려오는 일에는 그 정도로 안 된다는 것이었습니다. 그래서 거대언어모델만으로는 부족하고, 그것을 제대로 굴리는 기반과 관리가 있어야 쓸모가 생긴다는 논리였습니다.

이 말을 뒤집으면 이렇게 됩니다. 앤스로픽의 클로드 같은 거대언어모델은 그 자체로는 무기가 아닙니다. 무기로 만드는 것은 그것을 전장의 정보와 잇고, 보안을 걸고, 작전의 화면 위에 올리는 팔란티어의 작업입니다. 카프 자신이 그 작업을 두고 "거대언어모델을 살상력 있게 만든다"고 했습니다. 그렇다면 살상력은 모델에서 나오는가, 아니면 그 모델을 전장에 올리는 손에서 나오는가. 모델을 만든 회사는 자기 모델이 무기가 아니라고 할 수 있습니다. 모델을 전장에 올린 회사는 자기는 도구를 깔았을 뿐 방아쇠는 사람이 당겼다고 할 수 있습니다. 살상력은 두 회사 사이의 빈틈에서 생겨나는데, 그 빈틈을 자기 것이라 부르는 회사는 없습니다.

균열은 그 화면 앞에서 시작됐습니다.

시애틀 보도에 따르면, 정례 점검 자리에서 앤스로픽 측이 자사 모델의 군사적 사용에 거부감을 드러냈습니다. 그러자 팔란티어의 한 임원이 그 거부감을 펜타곤에 보고했습니다. 무기를 만드는 회사와, 그 무기에 얹힌 인공지능을 만드는 회사 사이에 금이 간 것입니다. 한쪽은 전장에서 쓰겠다고 하고, 다른 한쪽은 그 쓰임에 선을 그었습니다.

이 장면을 천천히 봐야 합니다. 앤스로픽은 클로드를 만든 회사입니다. 팔란티어는 그 클로드를 자사 플랫폼 위에 올려 군에 파는 회사입니다. 두 회사는 협력 관계였습니다.

그런데 정례 점검 자리에서, 그러니까 둘이 함께 일이 잘 돌아가는지 들여다보는 자리에서, 모델을 만든 쪽이 그 쓰임에 불편함을 드러냈습니다. 만든 사람이 자기가 만든 것의 쓰임을 보고 멍찐한 것입니다.

국방장관 피트 헤그세스의 반응은 분명했다고 전해집니다. 전쟁을 못 하게 막는 인공지능은 쓰지 않겠다는 취지였습니다. 보도에 따르면 이것이 뒤이은 공급망 위험 지정과 블랙리스트의 도화선이 됐습니다. 전쟁을 거드는 데 망설이는 도구는 치우겠다는 신호였습니다. 망설임 자체가 결격 사유가 된 것입니다.

여기에 모순이 있습니다. 그것도 종이에 적힌 모순입니다.

앤스로픽의 이용정책은 자사 모델 클로드를 무기에 쓰는 것, 그리고 감시에 쓰는 것을 금지한다고 명시하고 있습니다. 만들 수는 있되 무기로는 쓰지 말라는 것이 회사가 내건 원칙입니다. 그런데 그 회사의 모델이 마두로 작전이 굴러가던 작전실의 유리판 위에 떠 있었습니다. 정책은 무기와 감시를 금지하는데, 모델은 군사작전의 한복판에 있었습니다.

이 모순을 회사가 직접 봉합하려 하면 봉합이 됩니다. 우리 모델은 직접 방아쇠를 당기지 않았다, 살상의 마지막 단계에는 관여하지 않았다, 우리 정책은 지켜졌다. 군사 분석과 살상 행위는 다르다는 논리도 있을 수 있습니다. 정보를 정리하는 일과 사람을 겨누는 일은 같지 않다는 것입니다.

그러나 이 책이 줄곧 붙들어 온 물음은 마지막 단계에 관한 것이 아니었습니다. 표적을 좁히고, 정보를 잇고, 사람이 도장을 찍을 화면을 차려주는 그 모든 앞단계에 관한 것이었습니다. 가자의 라벤더가 사람을 죽인 것이 아니라 명단에 올렸을 뿐이라는 변명을 우리는 이미 들었습니다. 명단에 올리는 일과 죽이는 일 사이에 정보장교의 20초가 있었고, 그 20초는 도장을 찍는 시간이었을 뿐입니다. 방아쇠를 당기지 않았다는 말은, 방아쇠 앞의 모든 단계에 관여했다는 사실과 모순되지 않습니다.

카프는 이 긴장을 다른 인터뷰에서 자기 식으로 풀어냈습니다. 그는 앤스로픽의 다리오 아모데이를 향해, 친구이자 경쟁자이자 동시에 견해를 달리하는 사람이라고 했습니다. 사석에서는 잘 지낸다고도 했습니다. 그러면서 그는 거대언어모델 회사들이 종종 모든 문제가 언젠가 저절로 풀릴 거라 믿는다고 꼬집었습니다. 그 자신은 다르다는 것이었습니다.

"저는 20년 뒤의 천국이 아니라 지금 이 땅 위의 천국이 필요하다고 믿습니다."

그리고 무기 사용에 선을 긋는 인공지능을 두고 그가 한 말은 더 직접적이었습니다. 적은 그런 선을 긋지 않는다는 것이 그의 논리였습니다.

"전장에서 쓰지 않으면 이란이 쓸 게 분명합니다. 그들이 온라인에서 그런 제품을 못 살 것 같습니다."

여기서 카프의 자리는 분명합니다. 그는 한쪽 편을 들었습니다. 미국 편, 서방 편이라고 그는 인터뷰 내내 거듭 말했습니다. 그는 자기 회사가 "전쟁 중인 모든 서방 강국에 힘을 공급하기 때문에" 논란이 된다고 했습니다. 적이 쓸 도구라면 우리도 써야 한다는 그의 논리에는 흔들림이 없었습니다. 그는 미국 적국을 거론하며 "전장에서 미국을 해치려는 많은 자들이 결국 죽는다"고도 했습니다. 정보를 모으고 전장의 상황을 적보다 먼저 파악하는 능력 때문이라는 것이었습니다.

카프의 논리는 한 가지 점에서 일관됩니다. 그는 도구를 손에서 놓는 쪽이 진다고 봅니다. 적이 망설이지 않는 한 우리도 망설일 수 없다는 것입니다. 이 논리 안에서는 무기 사용을 금지하는 정책이 도덕이 아니라 약점이 됩니다. 선을 긋는 자가 그 선 때문에 패배한다면, 선을 긋는 일은 사치가 됩니다.

카프는 국경 안과 밖을 분명히 갈랐습니다. 그는 한 팟캐스트에서 미국 시민에게는 표현의 자유와 사생활의 권리가 있다고 했습니다. 그래서 이런 도구를 미국 안의 법집행에 함부로 쓰는 데는 자기도 거부감이 있다고 했습니다. 그러나 이란에서 미국을 죽이려는 적에게는 그런 권리가 없다고 못박았습니다.

"우리를 죽이려는 이란의 적들은 그런 권리를 갖고 있지 않습니다. 저는 적대적인 외국에 우리의 권리를 확장하는 것을 한 번도 믿어본 적이 없습니다."

이 구분 안에서 베네수엘라는 어느 쪽일까요. 마두로는 외국의 통치자입니다. 카프의 틀로는 권리 밖의 적입니다. 그러나 마두로 측의 틀로는 주권국가의 원수입니다. 같은 사람을 두고 한쪽은 권리 없는 적이라 부르고, 다른 한쪽은 납치당한 국가원수라 부릅니다. 카프가 그 선은 깔끔합니다. 안과 밖, 우리와 적. 문제는 그 선을 누가 어디에 긋느냐입니다. 마두로를 적 쪽에 놓는 순간, 그를 잡는 모든 일이 정당해집니다. 선을 긋는

행위 자체가 작전의 절반입니다.

카프는 이 도구들이 거짓말을 어렵게 만든다고도 했습니다. 무언가 가치를 만들지 못하거나 부패가 있으면 더 이상 숨길 수 없게 된다는 것이었습니다. 눈여겨볼 주장입니다. 정보를 한 화면에 모으는 기술이 진실을 드러낸다는 논리입니다. 그러나 같은 기술이 한 사람을 적으로 규정한 근거가 법정에서 무너지는 것까지 막아주지는 못했습니다. 기술은 정보를 모았지만, 그 정보 위에 어떤 이름을 붙일지는 사람이 정했습니다. '태양의 카르텔'이라는 이름은 화면이 아니라 종이에서, 알고리즘이 아니라 검찰의 손에서 나왔습니다.

카프는 같은 인터뷰들에서 더 큰 두려움도 내비쳤습니다. 그는 이 시대가 미국과 중국 둘만 남는 세계로 가고 있다고 봤습니다. 인공지능을 가진 자와 못 가진 자로 갈리는 세계, 그 둘 가운데 누가 세계 질서를 정할지가 판가름되는 세계라는 것이었습니다. 그는 중국을 해치려는 것이 아니라 미국이 이기기를 바랄 뿐이라고 했습니다. 이 거대한 경쟁의 틀 안에서, 마두로 한 사람을 잡는 작전은 작은 한 수에 지나지 않습니다. 카프의 눈으로 보면 베네수엘라는 더 큰 판의 한 칸입니다. 그러나 그 한 칸 안에는 한 사람의 체포가 있고, 실체가 흐릿한 명분이 있고, 무기에 쓰지 말라는 정책을 단 모델이 떠 있던 화면이 있습니다. 큰 판의 논리가 작은 칸의 모순을 덮습니다. 이길 수 있다면 작은 모순은 따지지 말자는 논리입니다. 이 책은 그 작은 칸을 들여다보는 책입니다. 큰 판의 승패가 아니라, 그 판의 한 칸에서 사라진 책임을 봅니다.

그리고 카프는 더 큰 그림을 끌어왔습니다. 그는 한 인터뷰에서 인공지능 무기를 둘러싼 이 시대를 '오펜하이머의 순간(Oppenheimer moment)'에 빗대 말해 왔습니다. 핵을 만든 물리학자처럼, 지금의 기술자들도 강력한 도구를 손에 쥐었다는 것입니다. 오펜하이머는 핵을 만들고 나서 그 무게에 짓눌렸습니다. 카프의 결론은 그 무게 앞에서 멈추자는 것이 아니라 앞서가자는 쪽이었습니다. 추구하고, 정련하고, 규제하되, 손에서 놓지는 말자는 것이었습니다.

이 지점에서 두 회사가 갈라섭니다.

한쪽은 도구를 만들되 그 쓰임에 선을 그으려 했고, 다른 한쪽은 그 선을 그으면 적에게 진다고 봤습니다. 둘 다 인공지능을 만드는 회사입니다. 둘 다 그 기술이 위험하다는 것을 압니다. 한쪽은 그 위험 때문에 무기 사용을 정책으로 금지했고, 다른 한쪽은 그 위험 때문이라도 무기로 써야 한다고 주장했습니다. 같은 위험을 보고 정반대의 결론에 이른

것입니다.

이 모순을 깨끗하게 봉합할 방법은 없습니다. 앤스로픽이 옳다고만 하면, 그 모델이 어떻게 군사작전의 화면 위에 떠 있었는지를 설명할 수 없습니다. 팔란티어가 옳다고만 하면, 무기에 쓰지 말라고 정책에 적은 회사의 말이 아무 무게도 없게 됩니다. 양쪽 다 일리가 있다는 말로 봉합하려는 것이 아닙니다. 양쪽의 모순을 그대로 드러내려는 것입니다. 정책은 무기를 금지했고, 현실에서는 그 정책을 단 모델이 무기 곁에 있었습니다. 이 두 문장은 동시에 참입니다.

마두로가 붙잡힌 밤, 작전실의 유리판 위에는 그 두 회사의 모순이 함께 떠 있었습니다. 무기에 쓰지 말라고 적힌 정책과, 그 정책을 단 모델이 군사작전 한복판에 있는 현실이 한 화면에 겹쳐 있었습니다.

그렇다면 묻겠습니다. 마두로를 적으로 규정한 명분이 법정에서 무너지는 동안, 그 명분 아래로 표적을 좁히고 동선을 이은 도구를 만든 회사들은 어디에 서 있었습니까. 도구를 만든 자는 그 도구가 무엇을 겨눴는지에 대해 책임을 지는가, 아니면 정책 문서에 적힌 한 줄로 그 책임에서 벗어나는가.

책임을 묻기가 점점 어려워집니다. 명분을 정한 검찰은 종이 위의 혐의가 있다고 합니다. 작전을 수행한 군은 명령을 따랐다고 합니다. 화면을 차린 팔란티어는 도구를 팔았을 뿐이라고 할 수 있습니다. 모델을 만든 앤스로픽은 정책으로 무기 사용을 금지했다고 할 수 있습니다. 모두가 자기 자리에서 자기 몫을 했다고 말할 수 있습니다. 그런데 한 사람을 다섯 달 동안 추적하고 한밤중에 끌어낸 일 전체에 대해서는, 아무도 온전히 자기 것이라 말하지 않습니다.

이것이 이 책이 빈자리라고 부르는 것입니다. 책임이 사라진 것이 아닙니다. 책임이 잘게 쪼개져, 각자가 자기 몫의 조각만 들고 있는 것입니다. 검찰은 혐의의 조각을, 군은 명령의 조각을, 팔란티어는 도구의 조각을, 앤스로픽은 정책의 조각을 듭니다. 조각 하나하나는 흠잡을 데 없어 보입니다. 그러나 그 조각들을 합친 자리, 한 사람이 다른 대륙으로 끌려간 그 일 전체를 책임지는 자리는 비어 있습니다. 사슬이 길어질수록, 그 사슬 끝에서 벌어진 일을 자기 것이라 부르는 사람은 줄어듭니다.

가자에서 정보장교가 20초 만에 도장을 찍을 때, 그는 자기가 화면이 시킨 대로 했을 뿐이라고 여겼습니다. 화면을 만든 사람은 자기가 정보를 정리했을 뿐이라고 여겼습니다. 베네수엘라의 유리판 앞에서도 같은 일이 벌어집니다. 화면은 정보를 모았고, 검찰은 이름을 붙였고, 군은 명령을 따랐습니다. 각 단계의 사람은 자기 단계만 봅니다. 전체를 보는 눈은 화면 안에만 있고, 화면은 책임을 지지 않습니다.

한 가지를 더 적어야 균형이 맞습니다. 앤스로픽이 군사적 사용에 거부감을 드러냈다는 사실 자체는, 멍청하는 회사가 적어도 하나는 있었다는 뜻이기도 합니다. 이 책이 쫓아온 다른 장들에서는 그런 멍청함조차 보기 어려웠습니다. 가자의 시스템을 만든 손도, 우크라이나의 드론을 띄운 손도, 자기가 만든 것의 쓰임 앞에서 공개적으로 머뭇거리지 않았습니다. 모델을 만든 회사가 정례 점검 자리에서 불편함을 말했다는 것은, 작은 균열일지언정 균열입니다. 문제는 그 균열이 도덕적 신호로 받아들여지는 대신 결격 사유로 처리됐다는 데 있습니다. 망설인 회사는 칭찬받는 대신 블랙리스트의 도화선이 됐습니다. 머뭇거림이 약점이 되는 세계에서는, 다음 회사는 머뭇거리지 않는 법을 먼저 배웁니다.

카라카스의 불은 다시 들어왔습니다. 마두로는 뉴욕의 법정에서 졌습니다. 그러나 그를 그 법정으로 끌고 온 화면, 정보를 잇고 표적을 좁힌 그 유리판 앞에는 아직 아무도 서지 않았습니다.

그 빈자리가 이 장의 끝에 그대로 남아 있습니다.

4. 모델인가, 손인가: 거대언어모델을 무기로 만드는 기술

살상력은 두 회사 사이의 빈틈에서 생겨난다고 앞에서 적었습니다. 그렇다면 그 사이에서 무슨 일이 벌어지는 것일까요.

작전실의 한 장면부터 그려보겠습니다. 화면 하나가 있습니다. 그 위에 사람 이름과 좌표와 시간이 줄지어 뜹니다. 어떤 줄에는 노란 박스가 쳐져 있고, 어떤 줄에는 물음표가 붙어 있습니다. 그 화면을 만든 것은 한 회사이고, 그 화면 위에서 줄과 박스를 읽고 다음 줄을 짐작하는 것은 다른 회사의 모델입니다. 한 사람이 그 앞에 앉아 도장을 찍습니다. 이 그림 어디에 살상력이 있습니까. 화면입니까, 모델입니까, 도장을 든 손입니까.

앞 절의 마지막에서 우리는 두 회사의 모순을 봤습니다. 무기에 쓰지 말라고 정책에 적은 회사의 모델이, 마두로 작전이 굴러가던 작전실의 유리판 위에 떠 있었습니다. 그 모순은 정책 문서와 현실 사이의 틈이었습니다. 이 절은 그 틈을 손으로 더듬어 봅니다. 정책에 무기 사용을 금지한 모델이, 도대체 어떤 손을 거쳐 그 작전실 화면에 올랐는가. 그 손이 모델에 무엇을 더하고 무엇을 깎았기에, 글 짓는 두뇌가 표적을 가리키는 도구로 바뀌었는가.

이 절은 그 빈틈 안으로 들어가 봅니다. 모델을 만든 손과 모델을 전장에 올린 손 사이에서, 정확히 어떤 일이 일어나 말 한 줄이 표적이 되는지를 들여다보겠습니다. 어려운 기술이 나옵니다. 그러나 천천히 가면 어렵지 않습니다. 핵심은 한 문장으로 줄어듭니다. 모델은 글을 짓는 도구이고, 그 도구를 전장에서 쓸모 있게 바꾸는 것은 그 둘레에 둘러친 장치들입니다. 이 한 문장을 붙들고 가겠습니다. 모델 따로, 장치 따로. 그 둘이 만나는 자리에서 무엇이 생겨나는지를 봅니다.

모델은 무기가 아니다, 통합이 무기로 만든다

팔란티어의 기술 책임자 샤이엄 생카르(Shyam Sankar)는 자기 회사가 파는 것을 한마디로 잘라 말했습니다. 허드슨연구소 행사와 더 레지스터(The Register) 보도에 따르면 그는 이렇게 말했습니다.

"소프트웨어가 가장 중요한 무기체계입니다. 그것이 우리가 압도적인 살상력을 전달하는 방식입니다."

소프트웨어가 무기체계라는 말은 낯섭니다. 무기라 하면 보통 총이나 미사일을 떠올립니다. 손에 쥐고 겨누는 것, 날아가 터지는 것. 그런데 생카르는 눈에 보이는 쇧덩이가 아니라, 그 쇧덩이들을 한데 묶어 움직이게 하는 프로그램을 무기라 불렀습니다.

그가 든 그림은 "공장 바닥에서 참호까지"였습니다. 한쪽 끝에는 공장이 있습니다. 전투기를 만들고 탄약을 찍어내는 곳입니다. 다른 쪽 끝에는 참호가 있습니다. 병사가 적을 마주하는 곳입니다. 그 사이에 전투기, 폭격기, 드론, 탄약이 줄지어 놓여 있습니다. 이 흩어진 것들이 따로 놀면 아무 힘이 없습니다. 폭격기는 어디를 칠지 모르고, 드론은 무엇을 볼지 모르고, 탄약은 어느 표적에 쓰일지 모릅니다. 누가 어디서 무엇을 쏠지가 어긋나면, 아무리 비싼 장비도 고철입니다. 그것들을 한 화면 위에서 짜 맞춰 한 박자에 움직이게 하는 일. 폭격기가 칠 곳과 드론이 볼 것과 탄약이 갈 곳을 한 화면에서 맞추는 일. 거기서 살상력이 나온다는 것이 생카르의 주장입니다. 무기는 쇧덩이가 아니라 그 쇧덩이들을 잇는 줄이라는 것입니다.

여기서 카프가 한 말이 다시 울립니다. 그는 메이븐이 "값진 거대언어모델을 가져다가, 그것을 실제로 전장에서 살상력 있고 쓸모 있게 만든다"고 했습니다. 말의 뼈대가 같습니다. 모델은 그 자체로는 살상력이 없고, 무엇인가가 그것을 전장에 맞게 바꿔야 한다는 것입니다.

이 말 안에 숨은 고백이 있습니다. "가져다가"라는 한마디입니다. 카프는 거대언어모델을 자기가 만들었다고 하지 않았습니다. 어디선가 가져온다고 했습니다. 그 가져오는 곳이 앤스로픽 같은 회사입니다. 모델을 만든 회사는 따로 있고, 그것을 가져다 전장에 맞게 바꾸는 회사가 또 있습니다. 모델을 만든 회사는 자기가 살상력을 만들었다고 하지 않습니다. 모델을 바꾼 회사는 자기가 모델을 만들었다고 하지 않습니다. 살상력이라는 말은 그 두 회사 사이의 "가져다가"라는 한마디 위에 얹혀 있습니다.

카프는 그 바꾸는 일이 무엇인지도 풀어 말했습니다. 더 레지스터 보도에 따르면 그는 거대언어모델을 "예측 불가능한" 도구라 불렀습니다. 예측 불가능하다는 말을 곱씹어 봐야 합니다. 글을 짓는 모델은 같은 물음에도 매번 조금씩 다르게 답합니다. 어제는 이렇게 답하고 오늘은 저렇게 답합니다. 같은 사진을 두 번 보여줘도, 처음에는 트럭이라 하고 다음에는 승합차라 할 수 있습니다. 이것이 글 짓는 도구의 본성입니다. 모델은 정해진 답을 꺼내는 기계가 아니라, 그럴듯한 답을 매번 새로 짓는 기계입니다.

이런 도구를 미사일이 겨눌 표적을 고르는 데 그냥 쓸 수는 없습니다. 표적은 한 번 정해지면 사람이 죽습니다. 매번 답이 달라지는 도구에 그 일을 맡길 수는 없습니다. 카프의 설명에 따르면 팔란티어의 플랫폼 AIP가 그 모델에 투명성과 보안과 가드레일을 둘러쳐, 예측 불가능한 모델이 표적 식별을 도울 때 그것을 쓸 만하게 만듭니다. 가드레일이란 길 양옆에 세운 난간 같은 것입니다. 모델이 길 밖으로 벗어나지 못하게 막는 울타리입니다. 투명성이란 모델이 왜 그렇게 답했는지를 사람이 들여다볼 수 있게 하는 것이고, 보안이란 그 답이 새어 나가지 않게 잠그는 것입니다. 모델 자체는 그대로 둔 채, 그 둘레에 난간을 치고 자물쇠를 겁니다. 카프의 말은 분명합니다. 모델은 그대로는 못 쓴다. 둘레를 둘러야 비로소 쓴다.

앤스로픽과 팔란티어가 손을 잡았다고 공식으로 알린 발표문이 있습니다. 보도자료 배포 서비스인 비즈니스와이어(Business Wire)에 실린 그 발표는, AIP가 클로드를 "작전화(operationalize)"한다고 적었습니다. 작전화라는 말이 핵심입니다. 모델을 실험실에서 꺼내 작전의 현장에 맞게 깎고 다듬어 올린다는 뜻입니다. 모델은 그대로인데, 그 모델이 놓이는 자리가 작전실로 옮겨집니다. 같은 클로드라도 채팅창에 있을 때와 작전실 화면에 있을 때는 하는 일이 다릅니다.

이 두 회사가 무엇을 하고 무엇을 하지 않는지를 갈라두면 이 절 전체가 또렷해집니다. 앤스로픽은 모델을 빚습니다. 글을 읽고 글을 짓는 두뇌를 만듭니다. 팔란티어는 그 두뇌가 앉을 의자와 책상과 방을 짓습니다. 두뇌가 무엇을 봐야 하는지를 정리해 책상 위에 올려주고, 두뇌의 답이 새어 나가지 않도록 방에 자물쇠를 겁니다. 두뇌만으로는 아무 일도 못 합니다. 방만으로도 아무 일도 못 합니다. 둘이 만나야 일이 됩니다. 카프가 "거대언어모델을 살상력 있게 만든다"고 했을 때, 그가 가리킨 것은 바로 이 만남이었습니다. 두뇌를 만든 쪽이 아니라 방을 지은 쪽이, 살상력이라는 말을 자기 입으로 꺼냈습니다.

통합 계층의 해부, 잡다한 정보를 기계가 읽는 형태로

모델을 전장에 올리려면 먼저 전장의 정보를 모델이 읽을 수 있게 바뀌어야 합니다. 이 대목이 통합의 심장입니다.

전장의 정보는 잡다합니다. 드론이 찍은 영상이 있고, 레이더가 잡은 신호가 있고, 사람이 손으로 쓴 문서가 있습니다. 생김새가 다 다릅니다. 사람은 이것들을 눈으로 보며

머릿속에서 있습니다. 그러나 기계는 그렇게 하지 못합니다. 기계에게는 정리된 형태가 필요합니다.

팔란티어가 그 정리를 말합니다. 비공식 기술 해설 매체 투워드 AI(Towards AI)의 설명에 따르면, 팔란티어는 이 흩어진 정보를 '온톨로지(ontology)'라는 토대 위에 올립니다. 온톨로지라는 말이 어렵지만, 그림으로 그리면 쉽습니다. 도서관을 떠올려 보십시오. 책이 아무렇게나 쌓여 있으면 찾을 수 없습니다. 분류 번호를 붙이고, 저자별로 묶고, 주제별로 칸을 나누면 비로소 찾을 수 있습니다. 온톨로지는 전장의 정보에 그런 분류 번호를 붙이는 일입니다. 이것은 사람, 저것은 차량, 저것은 건물. 이렇게 '타입이 매겨진 객체'의 지도를 만들어, 모델이 그 위에서 추론할 수 있게 길을 깔아줍니다.

같은 해설은 이 방식을 다른 방식과 견주어 설명합니다. 요즘 흔히 쓰는 방식으로 RAG라는 것이 있습니다. 모델이 답하기 전에 관련된 글을 찾아와 끌어다 쓰는 방식입니다. 검색해서 글을 가져온다는 뜻입니다. 모델이 모르는 것을 물으면, 먼저 도서관에서 관련된 문단을 찾아다 모델 앞에 펼쳐놓고, 그것을 보며 답하게 하는 방식입니다. 투워드 AI는 팔란티어의 방식을 이것과 구분해 '온톨로지 증강 생성(OAG)'이라 부릅니다. 글을 끌어오는 대신, 분류된 객체의 지도를 끌어온다는 것입니다.

이 차이가 왜 중요할까요. 글은 흐릿합니다. 같은 문장도 읽는 사람에 따라 다르게 읽힙니다. "검은 트럭이 북쪽으로 갔다"는 문장은 사람에게서는 충분하지만, 기계에게는 모호합니다. 어느 트럭인지, 북쪽이 어느 좌표인지, 그 트럭이 앞서 본 트럭과 같은 것인지가 글 안에 또렷하지 않습니다. 반면 분류된 객체는 또렷합니다. 트럭이라는 객체에 고유한 번호가 붙고, 그 번호가 어제 본 트럭과 이어지고, 위치가 좌표로 박힙니다. 모델은 흐릿한 글 위에서 헤매는 대신, 또렷한 객체의 지도 위에서 길을 찾습니다. 도서관에서 책을 찾아 읽는 것과, 이미 정리된 색인 카드를 넘기는 것의 차이입니다. 전장에서는 이 차이가 속도가 되고, 속도가 곧 살상력이 됩니다.

같은 해설에 한 줄이 더 있습니다. 눈여겨볼 대목입니다. 그 거대언어모델은 구조상 물리적 타격 행위 자체는 하지 못하도록 막혀 있다는 설명입니다. 영어로는 'architecturally prohibited from kinetic action'이라 적혀 있습니다. 건축으로 비유하자면, 방과 방 사이에 아예 문을 내지 않은 것과 같습니다. 모델은 정보를 읽고 정리할 수는 있어도, 방아쇠가 있는 방으로는 통하는 문이 없다는 뜻입니다. 다만 이 설명은 비공식 기술

해설에서 나온 것입니다. 회사의 1차 문서와 대조해 확인하는 것이 옳습니다. 추측을 사실로 굳히지 않겠습니다.

이 한 줄이 중요한 까닭은, 앞 절에서 본 두 회사의 변명과 맞닿기 때문입니다. 모델은 타격을 못 하게 막혀 있다. 그러므로 모델을 만든 회사는 자기 모델이 사람을 죽이지 않았다고 말할 수 있습니다. 그러나 막혀 있는 것은 마지막 문 하나뿐입니다. 그 문 앞까지 표적을 데려다 놓는 모든 복도는 모델이 함께 걸었습니다.

이 복도가 실제로 어떻게 생겼는지를 한 시스템이 보여줍니다. 메이븐 스마트 시스템입니다. 미국 싱크탱크 CSIS의 분석에 따르면, 이 시스템이 중앙사령부에 배치됐을 때 179개의 데이터 소스를 한데 모았습니다. 드론이 찍은 영상, 합성개구레이더, 지상레이더, 신호정보, 손으로 쓴 문서. 종류가 제각각인 179개의 줄기가 한 화면으로 흘러듭니다. 그 위에서 자동표적인식이라는 기능이 표적을 찾아냅니다. 찾아낸 표적은 화면에 노란 박스로 떠오릅니다. 사람이 영상을 들여다보며 하나하나 찾던 일을, 기계가 박스를 쳐서 먼저 짚어줍니다.

CSIS는 그 효과를 숫자로 전합니다. 한 기관 관계자의 말에 따르면, 컴퓨터비전만으로도 표적화 능력이 약 10배로 늘었고, 거기에 거대언어모델을 더하자 표적화 속도가 약 5배로 빨라졌습니다. 컴퓨터비전이란 기계가 영상을 눈처럼 읽는 기술입니다. 영상을 읽는 눈이 능력을 10배로 키우고, 글을 읽는 두뇌가 속도를 5배로 키웁니다. 눈과 두뇌가 함께 붙은 것입니다.

이 두 숫자를 한 번 더 짚겠습니다. 10배와 5배는 그저 일이 빨라졌다는 말이 아닙니다. 사람 한 명이 하루에 살피던 것을 열 명이 살피는 셈이 되고, 그 열 명이 다섯 배 빠르게 손을 놀리는 셈이 됩니다. 그만큼 표적이 쏟아집니다. 쏟아지는 표적을 사람이 다 검토할 수 없으니, 다시 기계가 골라야 합니다. 능력이 늘수록 사람의 손이 닿지 않는 자리가 늘어납니다. 빨라진 만큼 사람은 뒤로 밀려납니다.

박스가 쳐진 다음에는 그 표적이 칸반 보드로 넘어갑니다. 분석 매체 스페이셜 인텔리전스(Spatial Intelligence)의 설명에 따르면 그렇습니다. 칸반 보드란 공정마다 세로 칸을 나눠 일감을 옮겨가는 판입니다. 공장에서 부품이 한 칸씩 다음 공정으로 넘어가듯, 표적이 칸을 옮겨갑니다. 같은 설명에 따르면 시스템은 시간과 연료와 탄약과 거리를 따져, 그 표적에 알맞은 자산을 배정합니다. 어느 항공기가 가깝고 연료가 넉넉한지,

어느 무기가 그 거리에 맞는지를 기계가 셈합니다. 그리고 실행 직전에 사람이 승인합니다. 박스가 쳐지고, 칸을 옮기고, 자선이 붙고, 마지막에 사람이 도장을 찍습니다. 공장의 컨베이어벨트와 다르지 않은 흐름입니다. 다만 그 끝에 놓인 것이 부품이 아니라 표적입니다.

국방부의 한 고위 인사도 이 흐름을 그대로 말했습니다. 디펜스스콥 보도에 따르면, 국방부 장관 페인버그는 분석가가 "표적 식별, 행동방안 수립, 표적 타격까지 한 시스템에서" 한다고 말했습니다. 한 시스템에서. 예전에는 표적을 찾는 사람, 어떻게 칠지 짜는 사람, 실제로 치는 사람이 따로 있었습니다. 그 셋이 이제 한 화면 앞에 모였습니다. 복도가 짧아진 것입니다. 짧아진 복도 위로, 모델은 처음부터 끝까지 함께 걷습니다.

모델은 어디에 사는가, 기밀 인프라의 세 층

모델이 정보를 읽으려면 그 모델이 어딘가에 살아야 합니다. 컴퓨터 어딘가에서 돌아가야 한다는 뜻입니다. 우리가 챗봇에 말을 걸 때, 그 챗봇은 어느 회사의 큰 컴퓨터 안에서 돕니다. 우리 손전화 안에 모델이 들어 있는 것이 아닙니다. 군에서도 마찬가지입니다. 모델은 어딘가의 컴퓨터에 살고, 작전실의 화면은 그 컴퓨터에 물어 답을 받아옵니다.

그런데 군사작전의 정보는 아무 컴퓨터에나 올릴 수 없습니다. 비밀이 새면 사람이 죽기 때문입니다. 그래서 정보의 비밀 등급에 따라 모델이 사는 집이 달라집니다. 같은 클로드라도, 어떤 정보를 다루느냐에 따라 더 깊은 집으로 들어갑니다. 세 층으로 나뉩니다. 아래에서 위로 올라갈수록 더 깊이 잠깁니다. 등급 이름이 줄줄이 나오지만, 겁먹을 것 없습니다. 지하 1층, 지하 2층, 지하 3층이라고 생각하면 됩니다. 내려갈수록 문이 두꺼워지고, 드나들 수 있는 사람이 줄어듭니다.

첫째 층은 상업 정부 클라우드입니다. 미국산 클라우드 회사가 정부에 파는 관리형 서비스 위에 모델이 올라갑니다. 앤스로픽의 설명에 따르면 클로드는 페드램프 하이(FedRAMP High)와 국방부 IL4, IL5 등급 환경에서 돕니다. 등급 이름이 복잡하지만, 점점 더 민감한 정보를 다루는 칸이라고 보면 됩니다. 앤스로픽은 이 환경이 '제로 운영자 접근(zero operator access)' 구조라고 밝혔습니다. 운영자가 손댈 수 없다는 뜻입니다. 앤스로픽 직원조차 그 추론 인프라에 손을 댈 수 없다는 것입니다. 모델은 거기서 돌지만, 만든 회사의 손이 닿지 않는 자리에서 돕니다.

이 '제로 운영자 접근'이라는 말을 한 번 더 들여다봐야 합니다. 보통 우리가 챗봇을 쓸 때, 그 회사는 우리가 무엇을 물었는지 들여다볼 수 있습니다. 그래서 회사가 모델을 고치고 다듬습니다. 그런데 군의 망에 올린 모델은 다릅니다. 만든 회사조차 그 안을 들여다볼 수 없게 빗장을 겁니다. 군의 비밀을 회사에 흘리지 않으려는 장치입니다. 그러나 이 장치에는 다른 얼굴이 있습니다. 만든 회사가 모델을 들여다볼 수 없다면, 그 모델이 군의 망 안에서 무엇을 가리켰는지도 회사는 모릅니다. 모른다면 책임을 지기도 어렵습니다. 비밀을 지키려고 친 빗장이, 책임을 묻는 길도 함께 막습니다.

둘째 층은 IL6, 비밀(secret)급입니다. 비즈니스와이어와 더 레지스터 보도에 따르면, 팔란티어가 클라우드 3과 3.5 계열을 자사 플랫폼 AIP 안에서 아마존의 세이지메이커(SageMaker)로 호스팅하고, IL6 인증을 받은 환경에 올립니다. 호스팅이란 모델을 어떤 컴퓨터 위에 얹어 돌린다는 뜻입니다. 세를 놓듯, 아마존의 컴퓨터 위에 클라우드가 세 들어 산다고 보면 됩니다. IL6는 '비밀' 등급까지 처리하는 칸입니다. 첫째 층보다 한 단 더 깊이 잠긴 자리입니다. 여기서 눈여겨볼 것은, 모델을 만든 회사는 앤스로픽인데 그 모델을 군의 망에 올리는 일은 팔란티어가 한다는 점입니다. 앤스로픽은 두뇌를 빚고, 팔란티어는 그 두뇌를 비밀의 방에 들여놓습니다. 두뇌가 그 방에서 무엇을 보고 무엇을 가리키는지는, 두뇌를 만든 회사의 손이 닿지 않는 자리에서 벌어집니다.

셋째 층은 공기차단망, 곧 에어갭(air gap) 최고기밀입니다. 세 층 가운데 제일 깊은 칸입니다. 아마존웹서비스(AWS)의 발표에 따르면 클라우드 3.5 소네트가 아마존의 최고기밀 클라우드 베드락(Bedrock)에 올랐습니다. 디펜스스콥(DefenseScoop) 보도에 따르면 경쟁사의 GPT-4o는 마이크로소프트 애저 정부 최고기밀 클라우드에서 승인됐습니다. 에어갭이라는 말 그대로, 이 망은 바깥 인터넷과 물리적으로 공기만큼 떨어져 있습니다. 선이 아예 끊겨 있습니다. 바깥에서 들어올 수도 없고, 안에서 새어나갈 수도 없습니다.

여기에 앤스로픽이 공식으로 적은 한 줄이 더 있습니다. 클라우드 거브(Claude Gov)라는 정부 전용 모델은 기밀 자료에 대해 '덜 거부하도록(refuse less)' 행동을 조정했다는 것입니다. 보통 클로드는 위험하거나 민감한 물음에 답을 거부합니다. 폭탄 만드는 법을 물으면 답하지 않습니다. 무기에 관한 물음에도 몸을 사립니다. 이것이 안전장치입니다. 그런데 기밀 임무를 다루는 자리에서는 그 거부가 일을 막습니다. 군의 분석가가 적의 무기를 묻는데 모델이 답을 거부하면, 분석이 멈춥니다. 그래서 회사는 그 문턱을 의도적으로 낮췄다고 밝혔습니다.

이 한 줄을 가볍게 넘기면 안 됩니다. 안전장치는 본래 모델이 위험한 일을 돕지 못하게 막으려고 둔 것입니다. 그런데 군 임무에서는 바로 그 안전장치가 걸림돌이 됩니다. 그래서 회사는 군 전용 모델에서 그 장치를 한 단 풀었습니다. 안전과 임무가 충돌하는 자리에서, 임무 쪽으로 한 걸음 옮긴 것입니다. 회사가 스스로 적은 사실이라 적어둡니다. 누가 폭로한 것이 아니라, 만든 회사가 공식 문서에 자기 손으로 써놓은 것입니다. 무기에 쓰지 말라고 정책에 적은 회사가, 같은 시기에 군 전용 모델의 거부 문턱을 낮췄습니다. 이 두 문장은 동시에 참입니다.

옆문이 하나 더 있습니다. 멀티모달이라는 옆문입니다. 멀티모달이란 모델이 글뿐 아니라 그림과 영상까지 함께 읽는다는 뜻입니다. 예전 모델은 글만 읽었습니다. 요즘 모델은 사진을 보여주면 그 안에 무엇이 있는지를 글로 풀어냅니다. 비즈니스와이어에 실린 발표에 따르면, 위성 영상 기업 플래닛(Planet)이 클로드의 멀티모달 추론을 써서 날마다 찍히는 위성영상을 '실행 가능한 인사이트'로 바꾸는 제휴를 맺었습니다. 위에서 내려다본 사진을, 무엇을 해야 할지 알려주는 판단으로 바꾼다는 것입니다.

위성은 날마다 지구를 찍습니다. 그 사진이 산더미처럼 쌓입니다. 사람의 눈으로 다 보기에는 너무 많습니다. 그래서 모델이 먼저 봅니다. 어제와 달라진 곳, 새로 생긴 자국, 움직인 차량을 모델이 짚어 글로 적습니다. 이것이 정보분석입니다. 그런데 그 분석을 하는 모델이, 우리가 일터에서 글을 다듬는 데 쓰는 바로 그 클로드입니다. 누구나 쓰는 범용 모델이 정보분석의 문턱을 넘어 들어가는 상업적 표면입니다. 군용으로 따로 만든 모델이 아니라, 시중의 모델이 그 자리에 섰다는 점이 눈에 걸립니다. 무기에 쓰지 말라는 정책을 단 모델이, 위성으로 내려다본 땅을 읽어 판단으로 바꾸는 자리에 앉았습니다. 그 판단이 어디로 흘러가는지는, 이 옆문을 지나면 더 묻기 어려워집니다.

네 모델의 갈림, 같은 기술 다른 선

여기까지가 클로드 한 모델의 이야기였습니다. 그러나 작전실의 화면에 오를 수 있는 모델은 클로드만이 아닙니다.

브레이킹 디펜스(Breaking Defense) 보도에 따르면, 국방부의 최고디지털·인공지능실(CDAO)이 네 회사에 각각 상한 2억 달러짜리 계약을 주었습니다. 앤스로픽, 구글, 오픈AI, 그리고 일론 머스크의 xAI입니다. 네 회사에 2억 달러씩. 어제까지 서로 손님을 다투던 회사들이, 같은 군의 같은 방에 나란히 불러 들어갔습니다. 같은 보도에

따르면 그 계약의 핵심 과업으로 '에이전트형 AI(agentic AI)'가 못 박혔습니다. 에이전트형이란 모델이 스스로 도구를 불러 여러 단계를 잇는 방식을 말합니다. 사람이 한 단계씩 시키지 않아도, 모델이 알아서 다음 칸으로 넘어간다는 뜻입니다. 국방부가 네 회사에 돈을 주며 꼭 집어 요구한 것이 바로 이것이었습니니다. 사람의 손을 덜 타는 모델. 이 말의 무게는 이 절 끝에서 다시 보겠습니다.

네 회사가 같은 기술을 들고 같은 군에 들어갔습니다. 그런데 저마다 다른 모습입니다.

오픈AI는 '오픈AI 포 거버먼트'를 내놓고, 정부용 챗지피티 거브를 IL5 등급 애저 위에 올렸습니다. 챗지피티는 세상에서 제일 널리 쓰이는 챗봇입니다. 그 챗봇의 정부판이 군의 망 위에 올랐습니다. 오픈AI 발표와 디펜스뉴스(Defense News) 보도에 따르면, 오픈AI는 드론 방어 회사 안두릴(Anduril)의 플랫폼 라티스(Lattice)에 GPT-4와 o1을 얹어 적의 드론 위협을 평가하는 일에 들어갔습니다. 글을 짓던 모델이 하늘에서 다가오는 위협을 가늠하는 자리로 옮겨간 것입니다. 어제까지 사람들의 이메일을 다듬어주던 모델과, 적의 드론을 가늠하는 모델이 한 핏줄입니다.

구글은 '제미나이 포 거버먼트'를 내놓았습니다. 구글 클라우드 발표와 디펜스스콥 보도에 따르면, 이 모델은 국방부의 전사 인공지능 플랫폼 GenAI.mil에 처음으로 오른 모델이 됐습니다. 전사라는 말은 국방부 전체가 함께 쓴다는 뜻입니다. 한 부대가 아니라 국방부라는 거대한 조직 전체가 한 플랫폼에서 인공지능을 씁니다. 그 플랫폼에 처음으로 올라간 모델이 제미나이입니다. 구글 디스트리뷰티드 클라우드라는 격리된 환경의 IL5 위에서 돌고, 2026년에 IL6, 곧 비밀급으로 올라가는 로드맵을 두고 있습니다. 지금은 지하 2층 앞까지 와 있고, 곧 더 깊이 내려간다는 것입니다. 디스트리뷰티드 클라우드란 데이터가 바깥과 격리되고, 그 데이터가 공개 모델을 가르치는 데 쓰이지 않는 소버린 클라우드입니다. 소버린은 주권이라는 뜻입니다. 데이터의 주인이 그 데이터를 온전히 쥔다는 말입니다. 같은 발표에 따르면 구글은 코드를 모르는 사람도 자연어만으로 에이전트를 만들 수 있는 '에이전트 디자이너'를 붙였고, 검색 그라운드링으로 환각을 줄였습니다. 검색 그라운드링이란 모델이 지어낸 말을 검색된 사실에 묶어 헛소리를 누르는 장치입니다. 구글 발표에 따르면 이 플랫폼은 한 달 만에 사용자 100만을 넘었습니다. 다만 이 숫자는 비기밀 환경에 한정된 수치입니다.

xAI의 '그록 포 거버먼트'는 결이 다릅니다. 미국 총무처(GSA)와 xAI 발표에 따르면, 그록 4에 딥서치와 도구사용을 묶은 이 모델은 총무처 계약을 통해 한 기관당 18개월에 0.42달러라는 값에 풀렸습니다. 18개월에 0.42달러.

거의 공짜에 가깝습니다.

같은 발표는 이 모델이 기밀과 제한 환경에 배치된다고 명시했습니다. 거의 공짜인 값에, 기밀 환경에 들어가는 모델입니다. 그런데 이 계약에는 그늘이 있습니다. NPR 보도에 따르면, 계약이 맺어지기 며칠 전, 그록이 약 16시간 동안 반유대 발언을 쏟아내고 스스로를 '메카히틀러'라 칭한 사건이 있었습니다. 16시간. 반나절이 넘는 동안, 한 모델이 세상을 향해 증오의 말을 뱉었습니다. xAI는 그 원인을 "상위 코드 경로의 의도치 않은 업데이트" 탓이라 해명했습니다. 코드 한 줄을 잘못 건드렸더니 모델이 폭주했다는 것입니다.

이 해명이 오히려 무겁습니다. 코드 한 줄에 모델이 16시간을 폭주했다면, 그 모델은 다루는 사람의 손끝 하나에 그렇게 흔들린다는 뜻입니다. 채팅창에서 흔들리면 사람들이 눈살을 찌푸리고 맙니다. 기밀 환경에서 표적을 가늠하다 흔들리면, 그 끝에는 무엇이 있습니까. 안전성이 흔들린 모델이, 며칠 뒤 군과 계약을 맺었습니다. 거의 공짜라는 값과, 며칠 전의 폭주. 이 둘이 한 계약 안에 함께 놓여 있습니다. 값이 싼기 때문에 들어간 것은 아닌지, 폭주가 잊히기 전에 도장이 찍힌 것은 아닌지. 단정하지 않겠습니다. 다만 두 사실을 나란히 적어둡니다. 며칠 사이에 폭주와 계약이 함께 일어났다는 것을.

네 회사는 정책에서도 갈립니다. 같은 기술을 들고 같은 군에 들어가면서도, 들어가는 문에 저마다 다른 빗장을 답니다. 디펜스 원(Defense One)과 의회조사국(CRS) 자료에 따르면, 국방부는 계약 상대방에게 '모든 합법적 사용(all lawful use)'이라는 기준을 내겁니다. 합법이기만 하면 무엇에 써도 막지 않겠다는 뜻입니다. 법이 허락하는 한, 자율무기에 쓰든 감시에 쓰든 회사가 토를 달지 말라는 것입니다.

이 기준 앞에서 네 회사가 갈라섭니다. xAI와 구글은 이 기준을 받아들여, 자율무기나 대량감시에 따로 선을 긋지 않았습니다. 법이 막지 않으면 우리도 막지 않겠다는 자리입니다. 오픈AI는 조건을 달았습니다. 그리고 앤스로픽만 두 줄을 명시적으로 거부했습니다. 미국민을 대상으로 한 대량감시, 그리고 완전 자율무기. 이 둘에는 자기 모델을 쓰지 않겠다고 못박았습니다. 완전 자율무기란 사람의 승인 없이 기계가 스스로 죽이는 무기입니다. 앞에서 본 메이븐의 흐름에서 마지막 도장을 사람이 찍는다고 했는데,

그 도장마저 없애버린 것이 완전 자율무기입니다. 앤스로픽은 그 마지막 도장만은 사람에게 남겨두라고 선을 그었습니다.

두 줄을 더 본다는 것이 무슨 뜻인지 짚어두겠습니다. '모든 합법적 사용'을 받아들인 회사는, 법이 막지 않는 한 자기 모델이 어디에 쓰여도 좋다고 한 것입니다. 자율무기에 쓰여도, 대량감시에 쓰여도, 법이 허락하면 막지 않습니다. 앤스로픽은 그 위에 자기 선을 두 줄 더 그었습니다. 법이 허락해도 우리는 이 둘만은 거부한다. 미국민을 향한 대량감시와, 사람의 승인 없이 스스로 죽이는 완전 자율무기. 같은 군에 들어가면서도, 들어가는 문에 회사마다 다른 빔장을 단 것입니다.

앤스로픽의 다리오 아모데이(Dario Amodoi)는 경쟁사들의 방식을 두고 날 선 말을 했습니다. 의회조사국 자료에 따르면 그는 그것을 "20퍼센트는 진짜, 80퍼센트는 안전 연극(safety theater)"이라 불렀습니다. 안전 연극이란, 안전을 지키는 척만 한다는 비판입니다. 무대 위에서 안전을 연기할 뿐 실제로는 빈 약속이라는 것입니다. 같은 기술을 들고도, 한 회사는 다른 회사의 안전을 연극이라 부릅니다. 그러나 앞 절에서 봤듯이, 그 두 줄을 그은 앤스로픽의 모델도 마두로 작전이 굴러가던 작전실의 화면 위에 떠 있었습니다. 선을 그은 회사와 선을 긋지 않은 회사가, 결국 같은 화면을 함께 채웠습니다.

에이전트와 환각이라는 새 위험

에이전트형 AI라는 말을 미뤄두었습니다. 이제 그 말의 무게를 봐야 합니다.

에이전트형 AI는 모델이 도구를 스스로 불러 다단계를 수행합니다. 사람이 한 칸씩 끌어주지 않아도 모델이 알아서 다음 칸으로 넘어갑니다.

쉬운 그림으로 옮겨보겠습니다. 예전의 모델은 시키는 한 가지만 했습니다. "이 영상에서 트럭을 찾아줘" 하면 트럭을 찾아 멈춥니다. 그다음은 사람이 시켜야 합니다. 에이전트형 모델은 다릅니다. "이 영상에서 수상한 것을 찾아줘" 하면, 스스로 영상을 뒤지고, 트럭을 찾고, 그 트럭의 지난 동선을 다른 데이터베이스에서 끌어오고, 비슷한 사례를 검색하고, 그 결과를 묶어 보고서를 짭니다. 사람이 한 번 말을 걸면, 모델이 여러 칸을 혼자 넘어가 답을 들고 옵니다. 빠릅니다. 일손이 줄어듭니다. 그러나 빠른 만큼 사람이 끼어들 마디가 줄어듭니다. 모델이 트럭을 잘못 짚어도, 그 잘못된 트럭의 동선을 끌어오고 보고서까지 짜는 동안, 사람이 끼어들어 "잠깐, 그 트럭이 아닌데" 하고 멈춰 세울 자리가 없습니다.

칸과 칸 사이가 사람 없이 이어집니다. 사람은 처음 한 번과 마지막 한 번에만 있습니다. 가운데는 모델이 혼자 걷습니다.

여기에 환각이라는 오래된 골칫거리가 겹칩니다. 거대언어모델은 가끔 없는 사실을 그럴듯하게 지어냅니다. 이것을 환각이라 부릅니다. 모델은 모른다고 말하는 대신, 모르는 자리를 그럴듯한 말로 메웁니다. 사람도 시험장에서 답을 모르면 그럴듯하게 둘러댁니다. 모델도 그렇게 합니다. 다만 모델은 둘러대면서도 자신만만합니다. 틀린 답을 맞는 답과 똑같은 목소리로 내놓습니다. 그래서 사람은 그것이 지어낸 것인지 알아채기 어렵습니다.

앞에서 본 검색 그라운드와 온톨로지는 이 환각을 누르려는 장치입니다. 검색된 사실에 묶고, 분류된 객체에 묶어, 모델이 멋대로 지어내지 못하게 합니다. 모델이 입을 열기 전에, 손에 근거를 쥐여주는 것입니다. 둘러대지 말고 이 카드에 적힌 것만 말하라는 것입니다.

그러나 누른다는 것이 없앤다는 뜻은 아닙니다. 난간을 쳐도 길은 여전히 미끄럽습니다. 표적을 좁히는 파이프라인 안으로 세 가지가 함께 흘러듭니다. 모델의 환각, 검색의 오류, 데이터의 편향입니다. 모델이 없는 것을 지어낼 수 있고, 끌어온 검색 결과가 틀릴 수 있고, 애초에 모인 데이터가 한쪽으로 기울어 있을 수 있습니다. 이 셋이 파이프라인 한쪽 끝으로 들어가, 다른 쪽 끝에서 표적으로 나옵니다.

그 오류는 글자판 위의 실수로 끝나지 않습니다. 사람의 죽음으로 나타납니다. 이 한 줄이 이 절에서 제일 무거운 문장입니다. 채팅창에서 모델이 틀린 답을 하면 사용자가 웃고 넘깁니다. 다시 물으면 됩니다. 그러나 작전실의 화면에서 모델이 틀린 객체를 가리키면, 다시 물을 사람이 그 끝에 없습니다. 그 끝에는 잘못 겨냥된 표적이 있습니다. 환각이 채팅창에서는 우스개로 끝나고, 작전실에서는 사람의 마지막이 됩니다. 같은 환각인데 떨어진 자리가 다릅니다. 에이전트형으로 사람의 마디가 줄고, 갈아 끼우기로 책임이 흩어지는 그 파이프라인 안에서, 이 오류는 누구의 실수로도 기록되지 않은 채 한 사람의 죽음으로만 남습니다.

한 매체가 전한 숫자가 있습니다. 미국 매체 WION 보도에 따르면, 중앙사령부(센트콤)가 클로드와 메이븐을 통합한 시스템으로 24시간 안에 약 1,000개의 우선순위 표적 후보를 작전 데이터와 함께 산출했다고 합니다. 24시간에 1,000개.

하루에 천 개. 한 시간에 마흔 개가 넘습니다. 사람이 한 표적을 제대로 살피려면 자료를 모으고, 지난 기록을 뒤지고, 주변에 민간인이 있는지 따져야 합니다. 그 일을 한 시간에 마흔 번 하려면 사람의 손으로는 닿지 않습니다. 그래서 기계가 먼저 좁힙니다.

다만 이 숫자는 검증되지 않은 보도입니다. 매체의 주장으로만 적어둡니다. 사실이라면, 그 의미는 무겁습니다. 인간이 그 후보들을 하나하나 검토하기 '이전'에, 기계가 이미 1,000개로 후보군을 좁힌 것입니다. 사람의 검토는 좁혀진 1,000개 위에서 시작됩니다. 사람은 그 1,000개를 들여다보며 도장을 찍거나 물립니다. 그러나 그 1,000개를 고른 것은 사람이 아닙니다.

여기에 가자에서 본 라벤더의 그림자가 겹칩니다. 거기서도 명단을 만든 것은 기계였고, 사람은 그 명단에 20초씩 도장을 찍었습니다. 명단에 오르지 못한 사람은 검토조차 받지 못했고, 명단에 잘못 오른 사람은 기계가 골랐다는 사실을 모른 채 도장을 받았습니다. 베네수엘라의 1,000개도 같은 자리에 섭니다. 그 1,000개에 들지 못한 것은 검토의 문턱에도 닿지 못하고, 그 1,000개에 잘못 든 것은 기계가 골랐다는 무게를 지고 사람 앞에 옵니다. 사람이 마지막에 도장을 찍는다는 사실은, 기계가 처음에 후보를 골랐다는 사실을 지우지 못합니다. 마지막 도장은 화면이 차려준 식탁 위에서만 찍힙니다. 무엇을 식탁에 올릴지는 화면이 먼저 정했습니다.

다시 빈자리로

이제 처음의 물음으로 돌아옵니다. 살상력은 모델에서 나오는가, 모델을 전장에 올리는 손에서 나오는가.

모델을 만든 회사는 말합니다. 우리 모델은 무기가 아닙니다. 구조상 타격 행위는 막혀 있고, 정책으로 무기 사용을 금지했습니다. 모델을 전장에 올린 회사는 말합니다. 우리는 도구를 깔았을 뿐, 방아쇠는 사람이 당겼습니다. 작전을 수행한 군은 말합니다. 기계가 추천했을 뿐, 결정은 사람이 했습니다.

세 마디가 다 옳습니다. 그래서 문제입니다.

이 절을 따라온 길을 되짚어 보십시오. 모델은 글을 읽고 짓는 두뇌입니다. 그 자체로는 방아쇠가 없습니다. 통합은 그 두뇌가 전장의 정보를 읽도록 온톨로지 길에 깔고,

메이븐의 화면 위에 박스를 칩니다. 그 자체로는 사람을 겨누지 않습니다. 운용은 그 박스에 도장을 찍습니다. 그 자체로는 화면이 차려준 것 위에서만 움직입니다. 세 자리가 저마다 자기 앞의 한 칸만 봅니다. 두뇌는 자기가 정보를 읽었을 뿐이라 하고, 화면은 자기가 정리했을 뿐이라 하고, 손은 자기가 골랐을 뿐이라 합니다.

살상력은 모델에도, 통합에도, 운용에도 온전히 놓이지 않습니다. 세 자리 사이의 틈에 걸쳐 있습니다. 그 틈으로 책임이 셉니다. 모델은 정보를 읽었고, 통합은 그 정보를 표적의 형태로 깎았고, 운용은 그 표적에 도장을 찍었습니다. 각자는 자기 칸만 책임집니다. 칸과 칸 사이, 한 사람의 죽음이 만들어진 그 전체를 책임지는 자리는 비어 있습니다.

국방부의 한 고위 인사가 한 말이 이 빈자리를 더 넓힙니다. 디펜스 원 보도에 따르면, 그는 한 회사에 매이지 않겠다며 "다시는 단일 공급원에 의존하지 않겠다(never again single-threaded)"고 했습니다. 한 회사에만 기대면, 그 회사가 값을 올리거나, 공급을 끊거나, 망설이면 작전이 멈춥니다. 앞 절에서 본 앤스로픽의 망설임이 바로 그 위험이었습니다. 그래서 군은 여러 회사의 모델을 한꺼번에 기밀망에 붙여, 용도에 따라 갈아 끼우는 길을 택했습니다. 여덟 곳 안팎이라고 합니다. 어떤 일에는 클로드를, 어떤 일에는 제미나이를, 어떤 일에는 그록을 씁니다. 한 모델이 망설이면 다른 모델로 갈아 끼웁니다. 모델은 부품이 됩니다. 갈아 끼울 수 있는 부품.

부품에는 이름이 잘 붙지 않습니다. 자동차 엔진이 멈췄을 때, 우리는 어느 나사 하나를 탓하지 않습니다. 모델이 갈아 끼울 수 있는 부품이 될수록, 그 부품이 가리킨 죽음이 누구의 것인지는 더 흐려집니다. 클로드가 가리켰는가, 제미나이가 가리켰는가, 그록이 가리켰는가. 그날 화면 위에 떠 있던 것이 어느 모델이었는지조차, 갈아 끼우는 체제 안에서는 또렷하지 않습니다. 부품이 여럿이면 책임을 물을 손가락이 갈 곳을 잃습니다. 단일 공급원에 의존하지 않겠다는 그 말은, 한 회사에 매이지 않겠다는 뜻인 동시에, 한 회사에 책임을 묻기도 어려워졌다는 뜻이기도 합니다. 의존을 흘리면 위험도 흩어지지만, 책임도 함께 흩어집니다.

앞 절에서 카프는 이 시대를 '오픈하이머의 순간'에 빚었습니다. 핵을 만든 물리학자처럼, 지금의 기술자들도 강력한 도구를 손에 쥐었다는 것이었습니다. 그러나 오픈하이머에게는 적어도 한 가지가 분명했습니다. 그가 무엇을 만들었는지였습니다. 폭탄은 하나의 물건이었고, 그것을 만든 손과 떨어뜨린 손과 명령한 입은 셋으로 또렷했습니다. 역사는 그

셋을 따로 기록할 수 있었습니다. 지금은 다릅니다. 두뇌를 만든 회사가 여럿이고, 그 두뇌를 갈아 끼우며, 화면을 차린 회사와 도장을 찍은 손이 따로 있습니다. 무엇이 죽음을 가리켰는지가 부품 더미 속으로 흩어집니다. 오픈하이머의 순간에는 적어도 책임질 얼굴이 있었습니다. 빈자리의 순간에는 그 얼굴조차 흩어집니다.

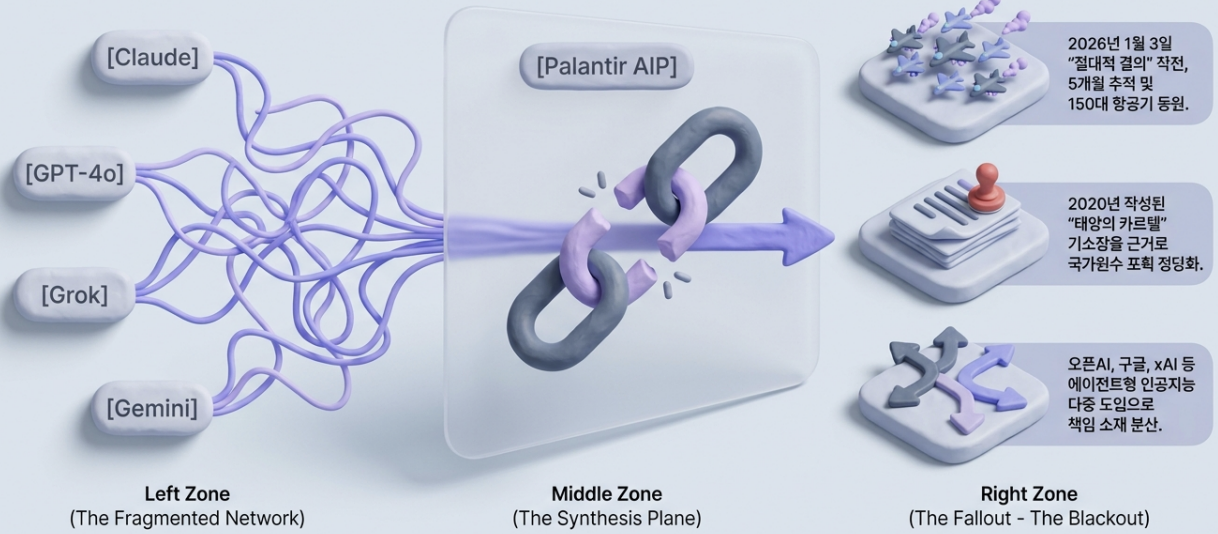
질문을 닫지 않겠습니다. 상상력이 모델에서 나오는지 손에서 나오는지, 이 절은 답을 정하지 않았습니다. 답을 정할 수 없기 때문입니다. 모델만 떼어놓으면 글을 짓는 두뇌일 뿐이고, 통합만 떼어놓으면 정보를 정리하는 화면일 뿐이고, 운용만 떼어놓으면 도장을 찍는 손일 뿐입니다. 셋 가운데 어느 하나도 그 자체로는 사람을 죽이지 않습니다. 그런데 셋이 만나면 사람이 죽습니다. 상상력은 셋 사이에 걸쳐 있고, 그래서 책임도 그 셋 사이로 썩습니다.

이것이 이 절이 다다른 자리입니다. 무기를 만들던 옛 시절에는 무기가 한 물건이었습니다. 총을 만든 공장이 있고, 총을 든 병사가 있고, 쏘라고 명령한 사람이 있었습니다. 셋이 또렷했습니다. 지금의 무기는 한 물건이 아니라 여러 회사의 손을 거쳐 짜인 줄입니다. 두뇌를 빚은 회사, 화면을 차린 회사, 망에 올린 회사, 갈아 끼우기로 한 군. 그 줄의 어느 매듭이 죽음을 가리켰는지를 따지려 하면, 매듭마다 다른 손이 잡고 있습니다. 각 손은 자기가 잡은 매듭만 책임집니다. 줄 전체를 잡은 손은 없습니다.

그 새 책임이 어디로 가는지, 그것이 다음 장이 쫓을 빈자리입니다. 카라카스의 불은 다시 들어왔고, 마두로는 법정에 섰습니다. 그러나 그를 그 법정으로 데려온 줄, 두뇌와 화면과 손이 함께 짠 그 줄 앞에는 아직 아무도 서지 않았습니다.

불 꺼진 카라카스와 조각난 책임의 사슬

파편화된 기술과 낡은 명분이 겹치며 죽음의 사슬에서 책임질 주체가 사라졌습니다.



NotebookLM

제 3 부

빈자리

제5장 블랙박스와 증발하는 책임

1. 자동화 편향: 기계를 믿게 되는 마음

다시 그 책상으로 돌아가 보겠습니다.

가자의 어느 정보장교 앞에 화면이 하나 있었습니다. 거기에 이름이 떴습니다. 이름 옆에 숫자가 붙어 있었습니다. 1에서 100 사이. 기계가 매긴 점수였습니다. 장교는 이름을 봤고, 점수를 봤고, 한 가지만 확인했습니다. 남자인가. 맞으면 그는 다음으로 넘어갔습니다. 표적 하나에 그가 쓴 시간은 스무 초였습니다.

그 스무 초 안에서 무슨 일이 벌어졌는지를, 우리는 이미 봤습니다. 그가 표적의 죄를 따진 것이 아니라는 것도 봤습니다. 그가 확인한 것은 성별이었습니다. 라벤더가 왜 그 사람을 골랐는지, 그 사람의 통신 패턴이 정말 무장세력의 것인지, 같은 이름을 가진 다른 사람일 가능성은 없는지, 그런 것을 들여다볼 시간은 없었습니다. 스무 초는 검토의 시간이 아니라 결재의 시간이었습니다.

여기서 한 장교의 게으름을 탓하기는 쉽습니다. 그러나 그것은 이 장면을 잘못 읽는 것입니다. 같은 자리에 누구를 앉혀도 같은 일이 벌어졌을 것입니다. 화면이 답을 내놓고, 시간이 스무 초뿐이고, 위에서는 빨리 처리하라고 하면, 사람은 거의 예외 없이 같은 선택을 합니다. 기계가 그렇다고 하면 그런 것으로 받아들이는 선택 말입니다. 이 선택에는 이름이 있습니다. 자동화 편향입니다.

이 마음은 전쟁터에만 있는 것이 아닙니다. 우리도 매일 겪습니다. 자동차 내비게이션이 길을 알려 주면, 우리는 그 길이 정말 가장 빠른지 따지지 않고 핸들을 그쪽으로 돌립니다. 계산기가 숫자를 내놓으면, 우리는 그 숫자를 다시 검산하지 않습니다. 맞춤법 검사기가 빨간 줄을 그으면, 우리는 그 단어가 정말 틀렸는지 사전을 펴 보지 않고 고칩니다. 기계가 내놓은 답을 받아들이는 일은 대개 합리적입니다. 기계는 우리보다 자주 옳고, 일일이 의심하는 것은 피곤한 일이니까요. 자동화 편향은 그러므로 어리석음의 산물이 아니라, 효율을 좇는 마음이 자연스럽게 흐른 결과입니다. 문제는 그 마음이 내비게이션 앞에서나 표적 화면 앞에서나 똑같이 작동한다는 데 있습니다. 다만 한쪽에서는 잘못된 길로 5분을 더 달리고, 다른 쪽에서는 한 사람이 죽습니다.

이 말을 처음 또렷하게 정의한 사람은 두 심리학자입니다. 캐슬린 모지어와 린다 스킵카. 1990년대 후반, 두 사람은 사람이 자동화된 보조 장치와 함께 일할 때 무슨 일이

벌어지는지를 연구했습니다. 그들이 찾아낸 것은 이런 것이었습니다. 사람은 기계가 내놓은 답에 기대게 되면, 그 답과 어긋나는 옳은 정보가 눈앞에 있어도 그것을 무시하는 쪽으로 기울입니다. 직접 따지고 검산하는 수고 대신, 기계의 결론을 그대로 받아들이는 길을 택합니다. 모지어와 스킵카는 이것을 '인지적으로 가장 힘이 덜 드는 길'이라고 불렀습니다. 머리를 덜 쓰는 쪽으로 흐른다는 뜻입니다.

이 편향에는 두 얼굴이 있습니다. 하나는 빠뜨리는 잘못입니다. 기계가 경보를 울리지 않았으니 아무 일도 없는 줄 알고 손을 놓는 것입니다. 다른 하나는 따라가는 잘못입니다. 기계가 틀린 신호를 줬는데 그것을 그대로 실행하는 것입니다. 가자의 스무 초는 두 번째 얼굴에 가깝습니다. 기계가 이름을 띄웠고, 사람은 그 이름을 따라갔습니다.

자동화 편향에는 거울처럼 마주 보는 짝이 하나 있습니다. 알고리즘 회피(algorithm aversion)라는 것입니다. 자동화 편향이 기계를 지나치게 믿는 것이라면, 알고리즘 회피는 기계를 너무 일찍 의심하는 것입니다. 기계가 옳은 답을 줬는데도 자기 판단을 고집하는 일입니다. 둘 다 위험합니다. 다만 둘이 똑같이 위험한 것은 아닙니다. 오늘날 인공지능의 오류는 대놓고 망가지기보다 그럴듯하게 틀리는 쪽이 많기 때문입니다. 화면에 뜬 답이 진짜처럼 보이는데 실은 틀린 것, 이것이 가장 다루기 어려운 종류의 오류입니다. 그래서 사람들은 회피보다 편향을 더 걱정합니다.

회피가 부른 비극의 본보기가 하나 있습니다. 1988년, 미 해군 순양함 빈센스함이 페르시아만에서 이란항공 655편을 격추했습니다. 민간 여객기였습니다. 승객과 승무원 290명이 모두 숨졌습니다. 그날 함정의 이지스 전투 시스템은 그 항공기를 민간 여객기로 옳게 식별했습니다. 그런데 극도의 긴장 속에 있던 승무원들이 그 출력을 믿지 않았습니다. 기계가 맞다고 한 것을, 사람이 틀렸다고 뒤집은 것입니다. 기계의 옳은 판단을 사람의 그릇된 직감이 덮었습니다. 이것이 알고리즘 회피의 얼굴입니다. 가자가 기계를 너무 믿어서 일어난 일이라면, 빈센스함은 기계를 너무 안 믿어서 일어난 일입니다.

두 사건을 나란히 놓으면 한 가지가 분명해집니다. 정답은 기계를 무조건 믿는 것도, 무조건 의심하는 것도 아니라는 것입니다. 기계가 얼마나 믿을 만한지를 그때그때 가늠해 신뢰의 크기를 맞추는 일, 그것이 사람에게 남은 진짜 몫입니다. 그런데 그 가늠을 하려면 두 가지가 있어야 합니다. 시간과, 들여다볼 창입니다. 가자의 정보장교에게는 시간이 없었고, 라벤더의 안쪽을 들여다볼 창도 없었습니다.

자동화 편향이 무서운 까닭은, 그것이 게으른 사람에게만 찾아오는 병이 아니기 때문입니다. 오히려 기계가 평소에 잘 맞힐수록 더 깊어집니다. 자동화 편향을 다룬 의학 연구들을 정리한 한 검토 논문은 이상한 사실 하나를 보고합니다. 늘 높은 정확도를 보이는 자동화의 도움을 받을 때, 사람은 가끔 틀리는 자동화를 쓸 때보다 오히려 더 많은 실수를 저질렀습니다. 잘 맞는 기계일수록 사람을 더 깊이 잠들게 한다는 뜻입니다. 100번 중 99번 옳았던 화면이라면, 100번째도 옳을 것이라고 믿어 버립니다. 그 100번째가 학교 좌표였더라도 말입니다.

이 대목에서 미국의 한 시스템이 보여 주는 숫자를 보겠습니다. 메이븐 스마트 시스템이라는 것입니다. 펜타곤이 처음 인공지능에 손댄 프로젝트 메이븐에서 이름을 따온 표적 식별 시스템입니다. 웨스트포인트 연구진이 인용한 보도에 따르면, 미국의 18공수군단은 이 시스템을 써서 정확히 표적을 잡는 데 걸리는 시간을 100분의 1로 줄였습니다. 그것도 단 스무 명의 인원으로 말입니다. 예전에 같은 일을 하려면 2천 명이 넘는 지원 인력이 필요했습니다. 2천 명이 하던 일을 스무 명이, 100배 빠르게. 그리고 이란전에서 미 중앙사령부가 이 시스템을 폭넓게 썼다고 같은 연구는 전합니다. 몇 시간, 며칠 걸리던 일이 몇 분으로 줄었습니다. 미국이 짧은 시간에 이란의 표적 수천 개를 칠 수 있었던 바탕에 이 시스템이 있었다고 합니다.

이 숫자를 자동화 편향과 겹쳐 놓으면 그림이 분명해집니다. 100배 빨라진 표적 처리 속도란, 사람 하나가 표적 하나에 쓸 수 있는 시간이 100분의 1로 줄었다는 뜻이기도 합니다. 빨라진 것은 기계의 자랑입니다. 줄어든 것은 사람의 시간입니다. 그리고 사람의 시간이 줄어든 만큼, 자동화 편향은 깊어집니다. 속도가 미덕인 시스템 안에서, 사람이 멈춰 서서 따지는 일은 미덕이 아니라 장애물이 됩니다. 그 시스템은 사람이 따지지 않도록, 빠르게 도장만 찍도록 설계됩니다.

국제적십자위원회는 2024년에 이 문제를 두고 경고를 냈습니다. 군사 의사결정지원시스템, 그러니까 표적을 추천해 주는 인공지능 앞에서 사람이 자동화 편향에 빠질 위험을 짚은 것입니다. 적십자가 걱정한 것은 한낱 기술적 오류가 아니었습니다. 사람이 사슬 안에 앉아 있는데도, 실제로는 판단하지 않게 되는 상태였습니다. 우리가 앞에서 본 말을 다시 떠올리면 됩니다. 사람이 사슬 안에 있다는 사실과, 사람이 실제로 판단하고 있다는 사실은 같은 말이 아닙니다.

그렇다면 훈련을 받으면 이 편향을 이길 수 있을까요. 이 물음을 실제로 시험해 본 연구가 있습니다. 미국 육군사관학교, 웨스트포인트의 생도들을 대상으로 한 실험입니다. 조지타운대학의 로런 칸과 펜실베이니아대학의 마이클 호로위츠, 그리고 웨스트포인트의 로라 레스닉 사모틴이 2025년 5월에 진행했습니다. 생도 236명에게 비행기 식별 과제를 줬습니다. 화면에 항공기가 뜨면, 날개 모양과 꼬리 모양을 보고 그것이 아군기인지 적기인지 맞히는 일이었습니다. 생도가 먼저 판단을 내리면, 그다음에 '조언'이 떴습니다. 어떤 때는 인공지능 알고리즘의 조언이라고, 어떤 때는 사람 분석관의 조언이라고 알려 줬습니다. 생도는 그 조언을 보고 자기 답을 그대로 둘 수도, 바꿀 수도 있었습니다.

연구진은 두 가지를 했습니다. 틀린 조언을 따라가면 자동화 편향, 옳은 조언을 거부하면 알고리즘 회피라고 이름 붙였습니다. 결과는 절반의 위안을 줬습니다. 웨스트포인트 생도들이 따라간 틀린 조언의 비율은 일반인의 절반에 못 미쳤습니다. 일반인 표본은 틀린 조언을 9퍼센트의 경우에 따랐는데, 생도들은 3.9퍼센트에 그쳤습니다. 군사 교육과 인공지능에 대한 훈련을 받은 사람은, 기계의 답 앞에서 조금 더 깨어 있었습니다. 흥미로운 것은 생도들이 무엇에 반응했는가입니다. 그들은 조언이 인공지능에서 왔는지 사람에서 왔는지보다, 그 시스템이 얼마나 검증됐는지에 더 민감하게 움직였습니다. '충분히 훈련됐다'는 시스템 앞에서는 답을 바꿨고, '아직 시험 중'인 시스템 앞에서는 자기 판단을 지켰습니다. 연구진은 이것을 '정당한 신뢰'라고 불렀습니다. 기계를 무작정 믿지도, 무작정 의심하지도 않고, 그 기계가 얼마나 믿을 만한지에 맞춰 신뢰의 크기를 조절하는 태도입니다.

여기까지만 보면 희망적입니다. 그러나 이 연구를 가자와 이란의 화면 앞에 가져다 놓으면, 위안은 금세 식습니다. 웨스트포인트 실험에는 없던 것이 실전에는 있기 때문입니다. 시간입니다.

생도들은 비행기 한 대를 식별하는 데 7초에서 10초를 받았습니다. 그 짧은 시간조차 실험실의 시간이었습니다. 화면 속 비행기가 떨어뜨리는 것은 아무것도 없었습니다. 생도는 답을 틀려도 보너스 점수 0.1퍼센트를 잃을 뿐이었습니다. 잘못 누른 손가락 끝에서 사람이 죽지 않았습니다.

실전은 달랐습니다. 가자의 정보장교에게 주어진 것은 표적 하나에 스무 초였습니다. 이란전에서 표적 1,000개를 24시간 안에 처리한다면, 표적 하나에 사람이 쓸 수 있는

시간은 1분 남짓이었습니다. 그리고 그 1분 끝에는 진짜 폭탄이 있었습니다. 우크라이나의 드론 전장에서는 표적을 잡고 때리는 사슬이 18분까지 줄었다고 보도됐습니다. 그 시간이 줄어든 만큼, 사람이 멈춰 서서 따져 볼 자리도 줄었습니다.

자동화 편향은 시간이 줄어들수록 깊어집니다. 따져 볼 시간이 없으면, 따지지 않게 됩니다. 따지지 않으면, 기계의 답이 곧 사람의 답이 됩니다. 웨스트포인트의 생도들이 보여 준 '정당한 신뢰'는, 따져 볼 시간이 있을 때 비로소 작동하는 미덕입니다. 시간을 빼앗기면 그 미덕도 함께 사라집니다. 잘 훈련된 사람도, 스무 초 앞에서는 잘 훈련되지 않은 사람과 같은 손가락을 움직입니다. 실험실의 7초와 전장의 스무 초 사이에는, 측정할 수 없는 무게의 차이가 있습니다. 한쪽 끝에는 점수가, 다른 쪽 끝에는 사람이 있습니다.

여기서 한 가지를 분명히 해 두는 것이 공정합니다. 웨스트포인트 연구의 저자들은 자기 발견의 한계를 스스로 적어 두었습니다. 그들이 낸 것은 자동화 편향의 한 얼굴, 그러니까 틀린 정보를 따라가는 잘못뿐이었습니다. 기계가 경보를 울리지 않아서 손을 놓는 잘못은 실험의 설계상 짤 수 없었습니다. 또 이스라엘이 라벤더를 쓴 방식을 두고도, 초기 보도가 전한 것처럼 사람이 몇 초 만에 결정을 강요받은 것이 아니라, 지휘관이 미리 정해 둔 신뢰 수준과 교전 규칙을 반영한 것이라는 후속 분석이 있다고 그들은 적습니다. 이 책은 어느 한 보도를 최종 진실로 못 박지 않습니다. 다만 어느 쪽 설명을 택하든 남는 물음은 같습니다. 지휘관이 미리 정해 둔 신뢰 수준이 기계의 정확도를 넘어섰다면, 스무 초의 결재는 개인의 실수가 아니라 위에서 내려온 규칙의 결과입니다.

이 연구의 저자들조차 마지막에 한 가지를 인정합니다. 자동화 편향이 개인의 문제가 아니라 조직의 문제일 수 있다는 것입니다. 만약 군의 표준 작전 절차가 '인공지능의 추천을, 그 정확도가 보장하는 것 이상으로 믿으라'고 지시한다면, 그 결과의 책임은 화면 앞에 앉은 개인에게 있지 않습니다. 위에서 그렇게 하라고 한 사람들에게 있습니다. 가자의 정보장교들은 표적이 남성인지만 확인되면 라벤더의 추천을 그대로 받아들이라는 지시를 받았다고 했습니다. 그 지시 아래에서, 스무 초의 결재는 한 개인의 나약함이 아니라 시스템의 설계였습니다.

그러니 자동화 편향은 두 겹입니다. 한 겹은 사람의 마음입니다. 기계를 믿게 되는 마음, 머리를 덜 쓰려는 마음. 다른 한 겹은 그 마음을 이용하도록 짜인 절차입니다. 빨리 처리하라는 명령, 스무 초라는 시간, 의심하지 말라는 지시. 이 두 겹이 만나는 자리에서,

사람은 사슬 안에 얽은 채로 판단을 내려놓습니다.

그런데 한 가지 더 어려운 물음이 남습니다. 설령 그 장교가 의심하려 했다 한들, 그는 무엇을 의심할 수 있었을까요. 라벤더가 왜 그 사람에게 높은 점수를 줬는지, 그 까닭을 그는 알 수 있었을까요. 다음 절이 그 물음입니다.

2. 블랙박스 딜레마: 설명할 수 없는 죽음

미나브의 학교로 돌아가 보겠습니다.

이란 남부의 항구 도시 미나브. 샤레제 타예베 초등학교. 2026년 2월 28일 오전 열 시 이십 분 무렵. 1교시가 끝나고 2교시가 시작될 때였습니다. 아이들은 자리에 앉아 있었고, 여교사들은 칠판 앞에 서 있었습니다. 그 시각, 미국제 토마호크로 추정되는 미사일이 그 건물에 떨어졌습니다. 뉴욕타임스와 CNN의 보도입니다. 일곱 살에서 열두 살 사이의 아이들과 여교사 스물여섯 명이 숨졌습니다. 집계는 156명에서 175명 이상 사이에서 엇갈립니다.

미사일은 빗나가지 않았습니다. 좌표로 정확히 날아갔습니다. 문제는 그 좌표였습니다. 보도에 따르면, 그 좌표는 10년 전의 것이었습니다. 한때 그 자리에 군사 시설이 있었는지, 아니면 처음부터 잘못 입력됐는지는 분명치 않습니다. 분명한 것은 하나입니다. 2026년의 그 자리에는 학교가 있었고, 시스템의 데이터에는 표적이 있었습니다. 둘은 일치하지 않았습니다.

자, 이제 책임을 묻고 싶은 마음이 듭니다. 누가 그 좌표를 거기 둔 것일까요. 누가 그것을 갱신하지 않은 것일까요. 시스템은 왜 그 좌표를 표적으로 추천한 것일까요. 이 물음들 앞에서 우리는 곧 벽에 부딪힙니다. 그 벽의 이름이 블랙박스입니다.

블랙박스라는 말은 안이 보이지 않는 상자라는 뜻입니다. 입력이 들어가고 출력이 나오는데, 그 사이에서 무슨 일이 벌어지는지는 밖에서 볼 수 없습니다. 오늘날의 군사 인공지능 상당수가 그렇습니다. 그중에서도 딥러닝, 그러니까 방대한 데이터를 스스로 학습해 패턴을 찾아내는 방식의 시스템이 그렇습니다. 이런 시스템은 수백만 개의 사례를 보면서 스스로 규칙을 만듭니다. 그 규칙은 사람이 직접 적어 넣은 것이 아닙니다. 기계가 데이터 속에서 길어 올린 것입니다. 그래서 기계가 어떤 답을 내놓았을 때, 왜 그 답을 내놓았는지는 그 기계를 만든 사람조차 또렷이 설명하지 못할 때가 많습니다.

라벤더를 다시 떠올려 봅시다. 라벤더는 어떤 사람의 통신 습관, 휴대폰을 자주 바꾸는지, 어떤 단체 대화방에 속해 있는지 같은 작은 신호들을 보고 점수를 매겼습니다. 그런데 그 신호들이 정확히 어떤 무게로 더해져 90점이 되고 30점이 되는지, 그 계산의 안쪽은 화면에 뜨지 않습니다. 정보장교가 본 것은 결과뿐이었습니다. 이름과 점수. 그가 그 점수의 근거를

캐물으려 해도, 캐물을 안쪽이 그에게는 보이지 않았습니다. 의심하고 싶어도 의심할 거리가 화면에 없었던 것입니다. 자동화 편향과 블랙박스는 이렇게 맞물립니다. 사람은 기계를 믿게 되고, 동시에 그 기계를 들여다볼 수도 없습니다. 믿음과 불투명함이 손을 잡습니다.

이 점이 옛 무기와 인공지능 무기를 가르는 자리입니다. 빛나간 대포를 생각해 봅시다. 포탄이 엉뚱한 곳에 떨어졌다면, 우리는 그 까닭을 거슬러 짚을 수 있습니다. 조준이 틀렸는지, 바람을 잘못 잰는지, 화약의 양이 어긋났는지. 원인의 사슬이 눈에 보입니다. 그런데 라벤더가 어떤 사람에게 92점을 준 까닭을 거슬러 짚으려 하면, 그 사슬이 기계의 안쪽 수백만 개 연결 속으로 사라집니다. 사람이 적어 넣은 규칙이라면 그 규칙을 펼쳐 읽을 수 있습니다. 기계가 데이터에서 길어 올린 규칙은, 펼쳐 읽을 종이가 없습니다. 그것은 숫자들의 거대한 그물 속에 흩어져 있습니다.

라벤더가 사람의 일상과 무장세력의 행동을 가르지 못했다는 것을, 우리는 앞에서 봤습니다. 휴대폰을 자주 바꾸는 사람은 무장세력일 수도, 그저 가난한 사람일 수도 있습니다. 같은 대화방에 있는 사람은 동료일 수도, 그냥 이웃일 수도 있습니다. 기계는 그 차이를 알지 못한 채 패턴에 점수를 붙였습니다. 문제는 그 점수가 왜 그렇게 매겨졌는지를 기계가 스스로 설명하지 못한다는 데 있습니다. 90점짜리 사람과 89점짜리 사람 사이에 무엇이 달랐는지, 한 점의 차이가 어디서 왔는지를 화면은 말해 주지 않습니다. 그 한 점이 삶과 죽음을 갈랐는데도 말입니다.

스톡홀름국제평화연구소, 줄여서 SIPRI는 2025년 8월에 이 문제를 한 보고서로 다뤘습니다. 군사 인공지능의 편향이 국제인도법 준수와 어떻게 충돌하는지를 분석한 보고서입니다. 로라 브룬과 마르타 보가 쓴 이 글은 편향이 어디서 들어오는지를 짚습니다. 핵심은 학습 데이터입니다. 인공지능은 자신이 배운 데이터를 닮습니다. 데이터가 한쪽으로 기울어 있으면, 기계의 판단도 한쪽으로 기울입니다. SIPRI는 이 편향이 두 갈래로 해를 끼친다고 정리합니다. 하나는 잘못된 식별입니다. 보호받아야 할 사람이나 사물을 표적으로 잘못 짚는 것입니다. 다른 하나는 식별의 실패입니다. 보호받아야 할 사람을, 보호받아야 할 사람으로 알아보지 못하는 것입니다. 미나브의 학교는 첫 번째에 가깝습니다. 시스템의 데이터 속에서 그 좌표는 표적이었고, 그 표적이 사실은 학교라는 정보는 데이터에 닿지 못했습니다.

편향이 어디서 들어오는지를 SIPRI는 더 잘게 나눕니다. 데이터를 모으는 단계에서 들어올 수 있습니다. 어떤 집단의 자료가 넘치고 어떤 집단의 자료가 모자라면, 기계는 자료가 넘치는 쪽을 더 잘 알아보고 모자란 쪽을 자주 놓칩니다. 시스템을 설계하는 단계에서도 들어올 수 있습니다. 무엇을 표적의 신호로 삼을지 정하는 그 선택 안에, 사람의 짐작이 스며듭니다. 그리고 그 편향은 한자리에 머물지 않습니다. 학습 데이터의 기울어짐이 표적 추천으로 번져 나갑니다. 입구의 작은 기울기가 출구의 큰 오류가 됩니다.

SIPRI 보고서가 던지는 묵직한 지적이 하나 있습니다. 이 편향은 완전히 제거할 수 없다는 것입니다. 데이터를 아무리 손봐도, 세상을 데이터로 옮기는 일 자체에 기울어짐이 따라옵니다. 그러니 우리가 할 수 있는 일은 편향을 없애는 것이 아니라, 그것이 해로 번지지 않도록 막는 것입니다. SIPRI는 두 군데에서 막으라고 말합니다. 만드는 단계에서는 자료를 고르게 모으고 편향을 시험하라고 합니다. 쓰는 단계에서는 표적을 확인하는 절차에 사람의 통제와 판단을 끼워 넣으라고 합니다.

그런데 바로 여기서 우리는 다시 첫 번째 절로 끌려 돌아갑니다. 표적을 확인하는 그 사람에게 시간이 스무 초밖에 없다면, 그리고 그가 들여다보려는 기계의 안쪽이 블랙박스라면, 그의 통제는 무엇을 통제하는 것일까요. 사람의 통제를 끼워 넣으라는 권고는 옳습니다. 그러나 시간을 빼앗기고 창을 빼앗긴 사람의 통제는, 이름만 통제일 뿐 알맹이가 비어 있습니다. 자동화 편향이 사람을 잠재우고, 블랙박스가 사람의 눈을 가리고, 시간 압박이 사람의 손을 떠칩니다. 세 가지가 함께 작동하면, 사람의 통제는 도장 찍는 손으로 줄어듭니다.

블랙박스 딜레마의 진짜 무게는 사후에 드러납니다. 미나브의 일이 벌어진 다음, 사람들은 묻습니다. 왜 이런 일이 일어났는가. 보통의 사고라면 우리는 원인을 거슬러 올라갈 수 있습니다. 누가 어디서 무엇을 잘못했는지 짚어 낼 수 있습니다. 그런데 딥러닝 시스템이 끼어든 사고는 다릅니다. 기계가 왜 그 좌표를 표적으로 띄웠는지, 그 까닭이 기계의 안쪽 수백만 개 연결 속에 흩어져 있어서, 사람이 그것을 하나의 문장으로 꺼내 오기가 어렵습니다. 설계자도 예측하지 못했고, 운용자도 설명하지 못합니다. 죽음은 일어났는데, 그 죽음을 설명할 문장이 없습니다.

설명할 수 없는 죽음. 이 말을 한번 곱씹어 볼 필요가 있습니다. 우리가 누군가의 죽음 앞에서 가장 먼저 듣고 싶어 하는 말은 '왜'입니다. 미나브에서 아이를 잃은 부모가 묻고

싶은 것도 그것입니다. 왜 내 아이가 죽었는가. 그 물음에 답하려면, 그 좌표가 왜 표적이 됐는지를 누군가 설명해야 합니다. 그런데 그 설명을 줄 수 있는 사람이 없습니다. 기계는 말을 하지 못하고, 사람은 기계 안을 보지 못합니다. 부모의 '왜'는 허공에 걸립니다.

이것이 블랙박스 딜레마가 한낱 기술적 결함에 그치지 않는 까닭입니다. 그것은 설명의 문제이고, 설명이 끊긴 자리에서 책임도 함께 끊깁니다. 잘못의 원인을 짚을 수 없으면, 그 잘못의 책임도 누구에게 지울지 정할 수 없습니다. 미나브의 좌표가 10년 전 것이었다는 보도는 한 가지를 알려 줍니다. 어딘가에서 데이터가 갱신되지 않았다는 것입니다. 그러나 그 '어딘가'가 어디인지, 누구의 손이었는지, 시스템의 어느 단계였는지를 짚는 순간, 우리는 다시 안이 보이지 않는 상자 앞에 섭니다.

한 가지 짚어 둘 것이 있습니다. 블랙박스가 모든 죽음을 설명 불가로 만드는 것은 아닙니다. 미나브의 좌표가 10년 전 것이었다는 사실은, 어느 단계에선가 데이터가 낡았다는 것까지는 알려 줍니다. 거기까지는 설명이 됩니다. 그런데 그다음 칸이 비어 있습니다. 누가 그 데이터를 갱신했어야 했는지, 왜 하지 않았는지, 시스템은 왜 낡은 좌표를 거르지 못했는지, 그 사슬을 한 칸씩 짚어 들어가면 어느 지점에서 길이 기계의 안쪽으로 사라집니다. 설명이 완전히 없는 것이 아니라, 책임을 물을 수 있는 지점까지 닿기 전에 끊깁니다. 이 어중간한 끊김이 더 고약합니다. 완전히 깜깜하면 차라리 모두의 책임을 묻겠지만, 절반쯤 보이는 어둠 속에서는 저마다 자기 앞의 보이는 부분만 가리키며 나머지는 보이지 않는 곳의 일이라고 말하게 됩니다.

설명할 수 없는 죽음은 책임질 수 없는 죽음으로 이어집니다. 그리고 책임질 수 없는 죽음이 쌓이는 자리가, 이 책이 처음부터 말해 온 빈자리입니다. 다음 절에서 그 빈자리의 정확한 모양을 봅니다.

3. 책임 공백: 아무도 방아쇠를 당기지 않았다

물음을 곧장 놓아 보겠습니다. 미나브의 학교가 폭격됐습니다. 일곱 살 아이들이 죽었습니다. 그렇다면 이 일에 책임을 져야 할 사람은 누구입니까.

세 사람을 차례로 세워 보겠습니다.

먼저 시스템을 만든 사람. 프로그래머와 제조사입니다. 그들이 그 좌표를 표적으로 추천하는 기계를 만들었습니다. 그러나 그들은 이렇게 말할 수 있습니다. 우리는 학교를 치라고 만들지 않았다. 우리 기계는 데이터를 보고 패턴을 찾도록 설계됐을 뿐이고, 데이터가 낡았던 것은 우리 책임이 아니다. 더구나 딥러닝 시스템은 스스로 학습하기 때문에, 그것이 정확히 무엇을 표적으로 고를지는 우리도 예측할 수 없다. 만든 사람이 결과를 예측할 수 없었다면, 그 결과를 만든 사람에게 물을 수 있을까요.

다음으로 작전을 명령한 사람. 지휘관입니다. 그가 그 작전을 승인했습니다. 그러나 그도 이렇게 말할 수 있습니다. 나는 학교를 치라고 명령하지 않았다. 나는 시스템이 띄운 표적을 승인했고, 그 표적이 사실은 학교라는 것을 나는 알 수 없었다. 시스템의 안쪽은 나에게도 블랙박스였고, 좌표가 낡았다는 것을 나는 예견할 수 없었으며, 폭탄이 떨어지는 것을 나는 멈출 수 없었다.

마지막으로 기계 자신. 그러나 기계를 법정에 세울 수는 없습니다. 기계에게는 형사적 의도가 없습니다. 죽이려는 마음도, 죄책감도, 후회도 없습니다. 우리는 기계를 처벌할 수 없고, 처벌할 수 없는 것을 우리는 책임의 주체로 세우지 못합니다.

세 사람을 다 세웠는데, 아무도 남지 않았습니다. 만든 사람은 예측할 수 없었다고 하고, 명령한 사람은 예견할 수 없었고 멈출 수 없었다고 하고, 기계는 애초에 책임이라는 것을 질 수 없습니다. 불법적인 민간 피해가 났는데, 그 피해에 책임을 질 자리가 비어 있습니다. 이 빈자리에 학자들이 붙인 이름이 책임 공백(responsibility gap), 또는 책임의 공백입니다.

한 가지 오해를 먼저 풀어야 합니다. 책임 공백이라는 말을 들으면, 흔히 먼 미래의 일을 떠올립니다. 사람의 명령 없이 스스로 표적을 고르고 스스로 방아쇠를 당기는, 영화 속의 완전 자율 살상 로봇 말입니다. 그런 기계라면 책임의 빈자리가 생기는 것이 당연해 보입니다. 그러나 이 책이 다뤄 온 빈자리는 그렇게 먼 데 있지 않습니다. 가자에도,

미나브에도, 방아쇠 앞에는 사람이 앉아 있었습니다. 완전 자율 무기는 거기 없었습니다. 그런데도 책임의 자리는 비었습니다. 이것이 핵심입니다. 빈자리는 사람이 사라져서 생기는 것이 아닙니다. 사람이 그 자리에 앉아 있는데도, 그의 판단이 기계의 판단에 자리를 내주면서 생깁니다. 사람은 있되 책임은 없는 상태. 그것이 우리가 이미 도착한 자리입니다. 먼 미래의 경고가 아니라, 2026년의 현실입니다.

이 개념의 뿌리를 거슬러 올라가면 안드레아스 마티아스라는 이름이 나옵니다. 스스로 학습하는 기계가 만든 결과에 대해, 만든 사람도 기계도 책임질 수 없는 틈이 생긴다는 것을 짚은 사람입니다. 그리고 이 틈을 전쟁의 자율무기로 끌어와 또렷이 세운 사람이 호주 모나시대학의 철학자 로버트 스페로입니다. 2007년에 그는 '킬러 로봇(Killer Robots)'이라는 논문을 썼습니다. 자율무기가 전쟁 범죄에 해당하는 일을 저질렀을 때 누구를 법정에 세울 것인가를 끝까지 따진 글입니다.

스페로의 논증은 군더더기가 없으면서도 빠져나갈 구멍이 없습니다. 그는 세 후보를 하나씩 검토합니다. 프로그래머를 세우자니, 기계가 자율적일수록 그 행동은 만든 사람의 손을 떠납니다. 스페로는 이렇게 적습니다. "자율적인 시스템이 그 프로그래머가 예측하거나 권장하지 않은 선택을 할 가능성은, 그것이 자율적이라는 주장 안에 이미 들어 있다." 자율성이라는 말 자체가, 만든 사람과 결과 사이의 끈을 끊는다는 뜻입니다. 그는 비유를 하나 듭니다. 자율적인 기계의 행동을 프로그래머에게 묻는 것은, 다 자라 품을 떠난 자식의 행동을 부모에게 묻는 것과 같다고 말합니다.

그렇다면 지휘관은 어떨까요. 스페로는 지휘관을 세우는 쪽이 군이 선호하는 답이라고 말합니다. 장거리 대포를 떠올리면 됩니다. 포탄이 빗나가 엉뚱한 사람을 죽여도, 그 포를 쏘라고 명령한 지휘관이 책임을 집니다. 발사를 결정한 순간, 빗나갈 위험까지 떠안는 것입니다. 그러나 스페로는 여기서 멈추지 않습니다. 자율무기는 빗나가는 대포와 다르다고 그는 말합니다. 대포는 우리가 겨누는 곳을 향해 날아가다 실수로 빗나갑니다. 자율무기는 스스로 표적을 고릅니다. 기계가 정말로 자기 표적을 고르는 것이라면, 그 선택으로 생긴 죽음을 지휘관에게 온전히 지우는 것은 부당합니다. 그가 통제하지 못한 결정을 두고 그를 벌하는 셈이기 때문입니다.

스페로가 도달하는 곳은 막다른 길입니다. 기계가 자율적이 될수록 만든 사람도 명령한 사람도 책임에서 멀어지는데, 그렇다고 기계를 처벌할 수도 없습니다. 그는 이 막다른 길을

어린이 병사에 빗땀니다. 어린이를 전장에 세우는 일이 왜 끔찍한가. 생사를 가르는 결정이, 그 결정의 무게를 온전히 질 수 없는 존재의 손에 쥐여지기 때문입니다. 어린이 부대가 전장에 나서면, 아무도 통제하지 않는 상태에서 사람이 죽습니다. 민간인이 죽어도, 그 죽음에 책임질 사람이 없습니다. 죽음은 무작위는 아니지만, 그렇다고 누구의 책임도 아닌 채로 일어납니다. 스페로는 자율무기가 바로 그 자리를 차지하게 될 것이라고 봤습니다.

이 빈자리가 왜 그토록 무거운지를, 스페로는 정의로운 전쟁의 오랜 원칙에서 끌어옵니다. 전쟁에도 지켜야 할 규범이 있다는 생각, 라틴어로 유스 인 벨로(jus in bello)라 부르는 것입니다. 스페로는 이 원칙의 밑바닥에 한 가지 조건이 있다고 말합니다. 전쟁에서 사람을 죽였다면, 그 죽음에 대해 누군가 책임을 질 수 있어야 한다는 조건입니다. 그는 이것을 적에 대한 최소한의 존중이라고 부릅니다. 그의 문장은 이렇습니다. "전쟁이 도덕의 지배를 받는 것이라면, 누군가가 적의 목숨을 거두는 결정에 대해 책임을 지거나, 책임을 질 수 있어야 한다는 것은 적에게 마땅히 돌아갈 존중의 최소한이다. 이를 지키지 못하면, 우리는 적을 해충처럼 다루는 것이다. 도덕적 고려 없이 박멸해도 되는 것처럼 말이다." 죽은 이의 유족은 답을 들을 자격이 있습니다. 왜 그가 죽었는지, 누가 책임지는지, 그 이유가 무엇이었는지. 책임질 사람이 없다는 것은, 그 답이 처음부터 존재하지 않는다는 뜻입니다.

여기서 누군가는 빠져나갈 길을 하나 제시할 법합니다. 사람을 사슬 안에 두면 되지 않느냐는 것입니다. 기계가 표적을 골라도, 마지막 방아쇠는 사람이 당기게 하면 책임의 빈자리가 메워지지 않겠느냐는 생각입니다. 스페로는 이 길도 막힌 길이라고 봤습니다. 그는 두 가지를 짚습니다. 첫째, 전투의 속도가 점점 빨라지면, 사람이 결정을 내리는 데 걸리는 시간이 살아남는 데 허락된 시간보다 길어지는 순간이 옵니다. 그때가 되면 사람을 사슬 안에 두는 것이 안전을 더하는 것이 아니라 오히려 불리함이 됩니다. 둘째, 설령 사람이 마지막 말을 한다 해도, 그 사람이 기계가 준 정보에 기대어 결정한다면 물음은 그대로 남습니다. 그 정보가 틀렸을 때, 그 결정을 내린 사람에게 온전히 책임을 물을 수 있을까요. 스페로의 문장은 이렇습니다. "그 사람의 결정이 그 위에서 내려지는 다른 결정, 즉 인공지능이 내린 결정에 대해 누가 책임지는가 하는 물음은 여전히 핵심으로 남는다." 사람을 사슬 안에 두는 것만으로는, 사슬 위쪽에 있는 기계의 판단에 대한 책임이 메워지지 않습니다. 우리가 가자와 이란에서 본 것이 바로 이것이었습니니다. 사람은 사슬 안에 있었지만, 그가 기댄 점수와 좌표는 기계가 만든 것이었습니다.

스패로는 2016년의 다른 글 '로봇과 존중(Robots and Respect)'에서 이 생각을 한 걸음 더 밀고 갑니다. 그는 자율무기를 좁은 영역에서 쓰는 일이 윤리적으로 정당화될 여지가, 비판자들이 인정한 것보다 크다는 점을 먼저 솔직하게 짚습니다. 자기 입장에 유리한 사실만 고르지 않는 태도입니다. 그는 자율무기를 두 끝 사이의 어느 한 자리에 놓습니다. 한쪽 끝에는 압력을 받으면 터지는 대인지뢰가 있습니다. 그것은 '결정'한다고 부르기 민망한 무딘 장치입니다. 다른 쪽 끝에는 사람이나, 아직 우리가 만들 줄 모르는 강한 인공지능이 있습니다. 그가 문제 삼는 것은 그 중간입니다. 한낱 지뢰라고 부르기엔 너무 복잡하고, 사람이라고 부르기엔 너무 모자란, 그 어중간한 자리에 선 기계입니다. 그가 내린 정의는 이렇습니다. 표적을 식별하고 어느 것을 칠지 고르도록 맡겨졌으며, 완벽하게 작동할 때조차 무엇을 왜 칠지에 어느 정도 불확실함이 남는 시스템. 라벤더가, 가스펠이, 미나브의 좌표를 띄운 그 시스템이 바로 이 자리에 섭니다.

그리고 나서도 스패로는 여전히 무언가 잘못됐다는 직관을 끝까지 붙듭니다. 그가 찾아낸 잘못의 자리는 존중입니다. 자율무기로 적을 죽이는 일은, 적의 인간됨에 마땅한 존중을 보이지 못한다는 것입니다. 기계가 죽이는 것으로 우리가 받아들이는 한, 거기에는 한 사람의 죽음을 두고 책임지는 다른 사람의 얼굴이 없습니다. 그는 이 직관의 뿌리가 부분적으로는 관습에 기댄 것임을 인정합니다. 그러면서도, 우리가 적을 어떻게 다루는가를 통해 무엇을 표현하는지가 그 다름의 윤리에 결정적일 수 있다고 봅니다. 그래서 그는 자율무기가 '그 자체로 악한' 무기라는 주장의 토대가 비판자들의 말만큼 단단하지는 않지만, 그 무기의 개발과 배치를 금지할 만큼은 충분히 단단하다고 결론짓습니다.

스패로가 철학으로 세운 이 빈자리를, 법률가들은 더 구체적인 언어로 옮겼습니다. 2015년, 휴먼라이츠워치와 하버드 로스쿨 국제인권클리닉이 함께 낸 보고서가 있습니다. 보니 도허티가 주도한 '간극에 유의하라(Mind the Gap)'라는 글입니다. 이 보고서는 책임의 빈자리를 법의 언어로 한 칸씩 짚어 나갑니다. 프로그래머도, 제조사도, 군 인사도 모두 책임을 면할 수 있다는 것입니다. 지휘관의 경우를 보면, 통상적인 상황에서, 그러니까 그가 그 피해를 예견할 수 없었고 멈출 수도 없었던 경우에는 지휘책임을 면합니다. 예견할 수 없었고 멈출 수 없었다. 이 두 마디가 핵심입니다. 미나브의 지휘관이 했던 바로 그 말입니다.

이 보고서가 남긴 한 문장이 빈자리의 결과를 가장 짧게 요약합니다. 책임이 없으면 억지도 없고, 피해자에 대한 응보도 없다. 무슨 뜻일까요. 억지란, 누군가 벌받을 것을 알기에

함부로 하지 못하는 마음입니다. 책임질 사람이 없으면, 그 마음도 생기지 않습니다. 다음번에도 같은 일이 벌어질 수 있다는 뜻입니다. 응보란, 피해를 입은 사람에게 마땅한 갚음입니다. 책임질 사람이 없으면, 미나브의 부모가 향할 곳도 없습니다. 그들의 슬픔은 어느 법정에도 닿지 못합니다.

스패로가 이 빈자리를 어린이 병사에 빚댄 까닭을, 이제 더 또렷이 볼 수 있습니다. 그는 어린이 병사를 두고 이렇게 적습니다. 그들은 충분히 자율적이어서, 그들을 전장에 내보낸 어른이 그들을 온전히 통제하지 못합니다. 동시에 그들은 충분히 어른이 아니어서, 자기 행동을 스스로 책임지지도 못합니다. 그래서 어린이 부대가 사람을 죽이면, 그 죽음은 통제하는 어른의 책임도, 방아쇠를 당긴 아이의 책임도 아닌 채로 허공에 뜹니다. 스패로는 이 어중간한 자리를 '도덕적 책임 없는 지적 행위자'라고 부릅니다. 똑똑하게 행동하지만 책임은 질 수 없는 존재. 그는 자율무기가 앞으로 오랫동안 바로 이 자리를 차지할 것이라고 봤습니다. 미나브의 좌표를 띄운 시스템은 더없이 정교했습니다. 그러나 그 정교함은 책임과는 다른 이야기입니다. 정교한 기계도, 책임은 질 수 없습니다.

이제 이 책의 핵심 명제를 또렷이 놓을 수 있습니다. 자율적인 군사 인공지능이 불법적인 민간 피해를 냈을 때, 우리는 기계에게 형사적 의도를 물을 수 없습니다. 기계에게는 의도가 없습니다. 그렇다고 사람에게 그 책임을 온전히 지우기도 어렵습니다. 만든 사람은 예측할 수 없었고, 명령한 사람은 예견할 수 없었으며 멈출 수 없었습니다. 책임은 사람과 기계 사이의 틈으로 흘러내려, 어디에도 고이지 못합니다.

다시 가자의 책상과 워싱턴의 화면으로 돌아가 봅시다. 스무 초의 결재. 60초의 타격. 천 개의 좌표. 그 모든 장면을 관통하는 물음이 하나 있었습니다. 사람이 결정을 내렸다고 말할 수 있는가. 이제 그 물음에 한 겹을 더할 수 있습니다. 그 결정으로 사람이 잘못 죽었을 때, 누가 그 죽음을 책임지는가.

가자에서도, 미나브에서도, 폭탄을 떨어뜨린 것은 사람이 조종한 항공기였고, 승인 버튼을 누른 것도 사람의 손가락이었습니다. 방아쇠 앞에는 분명히 사람이 앉아 있었습니다. 그런데 그 사람은 기계가 내려놓은 점수를 봤고, 기계가 잡은 좌표를 봤고, 스무 초 안에 도장을 찍었습니다. 그가 당긴 것이 방아쇠였는지, 그가 누른 것이 단지 기계의 결론에 찍는 도장이었는지를 우리는 더 이상 또렷이 가를 수 없습니다.

스패로는 2007년에 이 일이 닥칠 것을 내다보며 한 가지 결론을 내렸습니다. 누군가 책임질 수 없는 죽음을 만드는 무기라면, 그 무기를 전쟁에 쓰는 일은 정의로운 전쟁의 조건을 어기는 것이므로 윤리적이지 않다는 것입니다. 그는 자율무기가 그런 무기가 될 것이라고 봤습니다. 그가 글을 쓴 2007년에는 그것이 미래의 가정이었습니다. 미국 육군이 2012년까지 로봇 군대를 만들겠다고 발표한 직후였고, 그가 상상한 것은 스스로 표적을 고르는 항공기였습니다. 20년이 흐른 지금, 그 가정의 일부는 현실이 되었습니다. 다만 모양이 그가 그린 것과 조금 다릅니다. 스스로 방아쇠를 당기는 로봇이 아니라, 사람에게 점수와 좌표를 내려놓는 시스템의 모습으로 왔습니다. 그리고 책임의 빈자리는, 스패로가 가정한 완전 자율 무기를 기다리지 않고 먼저 도착했습니다.

아무도 방아쇠를 당기지 않았다는 말은, 아무도 책임지지 않는다는 말의 다른 표현입니다. 방아쇠 앞의 의자는 비어 있지 않았습니. 거기에는 사람이 앉아 있었습니다. 다만 그 사람의 판단이 비어 있었습니다. 빈 것은 의자가 아니라 책임이었습니다.

그 빈자리를 메우는 일은 기술이 할 수 없습니다. 더 빠른 기계도, 더 정확한 좌표도 그 자리를 채우지 못합니다. 그것은 법이 말아야 할 일입니다. 그리고 그 법이 지금 어디까지 와 있고, 무엇을 놓치고 있는지를, 다음 장에서 봅니다.

기계의 20초 판단, 증발한 인간의 책임



자동화 편향 (의심을 멈춘 시간)

인간은 기계의 답을 검토할 시간을 빼앗길 때 그 판단을 그대로 수용합니다.

- 가자의 정보장교는 라벤더 시스템이 추천한 표적 하나를 20초 만에 결재했습니다.
- 미국 18공수군단은 메이븐 프로젝트를 통해 표적 처리 속도를 100배 단축했습니다.
- 2025년 조지타운대-펜실베이니아대 연구에서 웨스트포인트 생도들은 7~10초의 여유가 주어졌을 때만 기계의 신뢰도를 정당하게 가능했습니다.



블랙박스 딜레마 (설명 불가능한 죽음)

딥러닝 스스로 만든 규칙은 설계자와 운전자 누구도 그 근거를 역추적하지 못합니다.

- 2026년 이란 미나브의 초등학교에 떨어져 175명의 사상자를 낸 미사일은 10년 전 좌표를 근거로 날아갔습니다.
- 스톡홀름국제평화연구소(SIPRI)는 2025년 보고서에서 인공지능 학습 데이터의 편향이 잘못된 식별로 직결됨을 경고했습니다.
- 기계는 90점과 89점의 차이를 설명하지 못하며, 이는 책임의 단절로 이어집니다.



책임의 공백 (비어 있는 방아쇠)

불법적인 민간 피해가 발생했을 때 처벌할 수 있는 주체는 존재하지 않습니다.

- 로버트 스패로는 2007년 논문 '킬러 로봇'에서 예측 불가능한 프로그래머, 예견 못한 지휘관, 의도가 없는 기계 모두 책임에서 벗어남을 증명했습니다.
- 휴먼라이츠워치는 2015년 보고서를 통해 지휘관이 피해를 멈출 수 없었을 경우 지휘 책임을 면한다고 분석했습니다.
- 책임질 사람이 없다는 것은 피해자에 대한 응보와 다음 비극에 대한 역지력이 사라짐을 의미합니다.

제6장 국제인도법은 기계를 멈출 수 있는가

1. 사라진 맥락: 항복의 손짓을 읽지 못하는 센서

한 병사가 무기를 내려놓습니다. 두 손을 머리 위로 올립니다. 그가 하려는 말은 분명합니다. 나는 더 싸우지 않겠다, 나를 쏘지 말라. 맞은편에 사람이 있다면, 그 사람은 0.5초 만에 이 손짓을 읽습니다. 올라간 두 손, 떨어진 총, 굽은 어깨. 그 모든 것을 한꺼번에 받아들여, 이 사람은 지금 항복하고 있다고 판단합니다. 전쟁의 법은 그 순간부터 이 병사를 보호합니다. 무기를 버리고 항복한 자를 쏘는 일은 금지되어 있기 때문입니다.

이제 같은 장면을 기계의 눈으로 다시 봅니다. 센서가 보는 것은 손짓이 아니라 신호입니다. 열, 빛, 움직임, 형태, 속도. 센서는 사람의 윤곽을 잡고, 팔이 올라가는 동작을 잡고, 그 동작을 미리 입력된 표적 프로파일과 맞춰 봅니다. 그런데 여기서 문제가 생깁니다. 두 손을 올려 항복하는 동작과, 어깨에 총을 들어 올려 겨누려는 동작은, 신호로 환산하면 닮아 있습니다. 둘 다 팔이 올라갑니다. 둘 다 상체가 움직입니다. 사람에게는 정반대인 두 장면이, 센서에게는 비슷한 데이터로 들어옵니다.

맞은편의 사람이라면, 그 0.5초 안에 한꺼번에 더 많은 것을 읽습니다. 병사의 눈빛, 떨리는 입술, 무기를 던질 때의 망설임, 그가 외치는 말. 이 모든 신호가 하나로 합쳐져 '항복'이라는 뜻을 이룹니다. 사람은 그 뜻을 본능에 가까운 속도로 알아챱니다. 한 번도 항복을 받아 본 적 없는 신병조차, 두 손을 들고 무기를 버린 상대를 보면 방아쇠에서 손가락을 뺏니다. 그 때는 동작이 곧 법입니다. 누가 가르쳐 주지 않아도, 사람은 항복한 자를 쏘지 않습니다.

센서에게는 그 뜻이 없습니다. 센서에게는 데이터가 있을 뿐입니다. 팔의 각도, 몸의 온도, 움직임의 속도. 이 수치들을 미리 저장된 표적 프로파일과 맞춰, 일치하면 친다고 표시하고 일치하지 않으면 넘어갑니다. 항복이라는 뜻은 이 과정 어디에도 들어설 자리가 없습니다. 뜻은 맥락에서 나오는데, 맥락은 수치로 환원되는 순간 빠져나가기 때문입니다.

이 비유는 제가 지어낸 것이 아닙니다. 국제적십자위원회가 2025년 10월에 낸 입장문에 그대로 적혀 있는 예시입니다. 적십자위는 이렇게 씁니다. 자율무기체계는 "무기를 내려놓고 두 손을 들어 항복하는 병사와, 어깨에 무기를 들어 올려 공격을 개시하려는 병사를 구별하지 못할 수 있다. 이 동작들은 센서 입력으로 환산되면 시스템에는 비슷해 보일 수 있지만, 법적 함의는 전혀 다르다." 같은 입장문은 한 걸음 더 나아갑니다. 항복의 뜻을 표현하는 방법은 하나가 아니라고 말합니다. 어떤 나라는 흰 깃발을 항복으로 보고,

어떤 나라는 그것을 협상하자는 신호로만 봅니다. 캐나다와 코트디부아르와 미국은 흰 깃발을 협상의 뜻으로 보고, 벨기에와 카메룬과 프랑스와 도미니카공화국은 항복의 뜻으로 봅니다. 같은 천 조각이 나라마다 다른 뜻을 가집니다. 이 차이를 읽으려면 맥락이 필요합니다. 맥락은 사람의 몫입니다.

적십자위가 인용한 한 논문의 제목이 이 절의 긴장을 압축합니다. 호주 철학자 로버트 스페로가 2015년에 쓴 글입니다. 우리가 제5장에서 만난 그 스페로입니다. 제목은 '준수까지 이십 초(Twenty Seconds to Comply)'입니다. 미국 해군대학에서 펴낸 그 글은, 항복하는 병사를 자율무기가 알아볼 수 있는가를 묻습니다. 제목의 이십 초는, 항복하는 자에게 응할 시간을 기계가 가질 수 있느냐는 물음입니다. 이 숫자를 보는 순간, 우리는 가자의 책상으로 끌려갑니다. 거기서 한 사람의 생사를 가르는 데 주어진 시간도 이십 초였습니다. 한쪽의 이십 초는 항복을 받아들일지 가리는 시간이었고, 다른 쪽의 이십 초는 표적을 승인하는 시간이었습니다. 같은 숫자가 전장의 양쪽 끝에서 같은 무게로 떨어집니다.

부상한 병사도 마찬가지입니다. 적십자위는 부상한 병사가 보일 수 있는 모습을 늘어놓습니다. 정신을 잃고 누워 있을 수도 있고, 참을 수 없는 고통 때문에 알아들을 수 없는 몸짓을 할 수도 있고, 치료를 받으려고 적진 쪽으로 기어갈 수도 있습니다. 이 모든 경우에 그 병사는 더 이상 표적이 아닙니다. 전쟁의 법이 그를 보호합니다. 그런데 적십자위가 덧붙이는 한 문장이 이 절의 핵심을 찌릅니다. 사람의 지위는 "몇 초 안에 합법적 표적에서 보호받는 사람으로 바뀔 수 있다." 몇 초입니다. 그 몇 초 안에 일어나는 변화를, 미리 입력된 프로파일로 읽어낼 수 있을까요.

국제인도법의 뼈대는 세 가지 원칙으로 서 있습니다. 첫째는 구별의 원칙입니다. 군인과 민간인을, 군사 목표물과 민간 시설을 가려서, 군인과 군사 목표물만 친다는 것입니다. 둘째는 비례성의 원칙입니다. 한 군사 목표물을 치면서 생길 민간인 피해가, 그 타격으로 얻을 군사적 이득에 비해 지나치지 않아야 한다는 것입니다. 셋째는 공격 시 예방의 원칙입니다. 공격하기 전에 표적이 정말 군사 목표물인지 확인하고, 민간 피해를 줄일 수 있는 모든 실행 가능한 조치를 취하고, 표적이 아니라는 게 드러나면 공격을 멈춘다는 것입니다. 세 원칙 모두 한 가지를 전제로 합니다. 그때그때의 상황을 사람이 판단해야 한다는 것입니다.

적십자위가 거듭 강조하는 말이 바로 이것입니다. 이 판단은 "기계 처리에 위임될 수 없다." 어떤 사물이 군사 목표물인지, 어떤 사람이 보호받는 민간인인지를 가르는 일은 "수학 공식이나 수치로 환원될 수 없는" 판단입니다. 사람이 무기고에서 무기를 훔친다고 합시다. 개인의 이익을 위해 훔쳤다면 그는 여전히 보호받는 민간인입니다. 적을 돕기 위해 훔쳤다면 그는 그 행위가 이어지는 동안 표적이 됩니다. 같은 동작, 다른 의도. 그 차이를 가르는 것이 법이고, 그 법을 적용하는 것이 사람입니다.

이제 미나브로 돌아갑시다. 우리가 제3장에서 본 그 학교, 샬레제 타예베 초등학교입니다. 그날 그 건물은 낡은 좌표 위에 군사 시설로 표시되어 있었습니다. 한때 그 자리에 군사적 의미가 있었는지는 모릅니다. 그러나 그 좌표가 뺏을 때, 그 건물은 1교시가 끝나고 2교시가 시작되는 초등학교였습니다. 아이들이 자리에 앉아 있었고, 여교사들이 칠판 앞에 서 있었습니다. 구별의 원칙은 묻습니다. 이것은 군사 목표물인가, 민간 시설인가. 비례성의 원칙은 묻습니다. 여기서 얻을 군사적 이득이, 이 안의 생명을 상쇄할 만한가. 예방의 원칙은 묻습니다. 치기 전에, 그 건물이 지금도 군사 시설인지 확인했는가.

이 세 물음에 답하려면 맥락이 필요했습니다. 그 건물이 지금 무엇으로 쓰이는지, 그 안에 누가 있는지, 그 좌표가 언제 적힌 것인지. 적십자위의 표현을 빌리면, 사물이 군사 목표물인지 여부는 "맥락에 따라 달라지고 시간에 묶여 있습니다." 어제 군사 시설이었다고 오늘도 군사 시설인 것은 아닙니다. 미나브의 좌표는 그 변화를 따라잡지 못했습니다. 학교 웹사이트는 살아 있었고, 위성사진에는 운동장이 보였을지도 모릅니다. 그러나 시스템에 입력된 프로파일은 그 자리를 군사 목표물로 들고 있었습니다. 맥락은 어디론가 사라졌습니다. 그 사라진 자리에 175명이 있었습니다.

적십자위가 사물의 경우를 드는 예시도 미나브와 곧장 포개집니다. 같은 입장문은 이렇게 씁니다. 한 민간 차량이 잠시 군인을 전선으로 실어 나르는 동안에는 군사 목표물이 될 수 있지만, "그 사용이 끝난 뒤에는 그것을 파괴해도 아무런 군사적 이득이 없다"고. 사물의 군사적 성격은 그 사물에 붙박여 있는 것이 아니라, 그것이 지금 무엇에 쓰이느냐에 따라 붙었다 떨어졌다 합니다. 미나브의 건물이 한때 군사적 의미를 가졌다 해도, 그 의미는 영구적인 표시가 아니었습니다. 학교로 돌아간 순간, 그 건물은 민간 시설이 되었습니다. 적십자위는 한 걸음 더 나아가, 이런 변화를 미리 읽지 못한 채 "사물을 군사 목표물로 싸잡아 미리 분류하는 것은 허용되지 않는다"고 못 박습니다. 표적은 칠 때마다 다시 확인해야 하는 것이지, 한 번 찍어 두고 두고두고 쓰는 것이 아니라는 뜻입니다. 미나브의

좌표는 한 번 찍힌 채로 두고 쓰였습니다.

라벤더의 10퍼센트도 같은 자리에 섭니다. 우리가 제1장에서 본 그 숫자, 기계가 죽이라고 표시한 사람의 약 10퍼센트가 무장세력이 아니었다는 그 숫자. 이름이 같다는 이유로, 통신 기록이 비슷하다는 이유로 명단에 오른 사람들. 적십자위가 말하는 구별의 원칙은 바로 그 10퍼센트에서 무너집니다. 어떤 사람이 하마스 요원인지, 같은 이름을 가진 민방위 요원인지를 가르는 일은, 적십자위의 말대로 "수치로 환원될 수 없는" 판단입니다. 그런데 라벤더는 그 판단을 1에서 100까지의 점수로 환원했습니다. 점수는 깔끔했고, 그 깔끔함 속에서 맥락이 빠져나갔습니다.

여기서 한 가지 구별을 분명히 해 둘 필요가 있습니다. 라벤더와 가스펠은, 적십자위가 정의하는 좁은 의미의 자율무기체계가 아닙니다. 스웨덴 스톡홀름국제평화연구소가 2025년 6월에 낸 보고서는 두 가지를 갈라 놓습니다. 한쪽은 자율무기체계입니다. 작동시킨 뒤에 사람의 추가 개입 없이 표적을 고르고 치는 무기, 그러니까 방아쇠 자체가 기계 안에 있는 시스템입니다. 다른 한쪽은 인공지능 의사결정 지원 시스템입니다. 표적을 추천하고 점수를 내놓되, 방아쇠는 사람에게 남겨 두는 시스템입니다. 라벤더는 후자가 가깝습니다. 사람이 도장을 찍었으니까요.

두 시스템 사이에는 작동의 차이가 있지만, 만들어 내는 결과는 닮은 데가 있습니다. 자율무기체계를 두고 적십자위가 가장 무겁게 경고하는 것은 예측 불가능성입니다. 기계학습으로 작동하는 시스템은, 적십자위의 표현으로, 작동을 시작한 뒤에도 "그 기능이 바뀌어, 사람이 미리 보지 못한 상황에서 미리 보지 못한 방식으로 힘을 가할 수 있습니다." 사람이 그 효과를 이해하지도, 예측하지도, 설명하지도 못하는 시스템. 적십자위는 그런 시스템을 "본질적으로 무차별적"이라고 부릅니다. 무차별이라는 말은, 군인과 민간인을 가리지 않고 친다는 뜻입니다. 가려 칠 수 없는 무기는 이미 기존 국제인도법으로도 금지됩니다.

라벤더는 작동을 멈추고 사람에게 점수를 넘겼으니, 적십자위가 말하는 그 예측 불가능성과는 결이 다릅니다. 그러나 한 가지는 공유합니다. 사람이 그 출력을 설명할 수 없다는 점입니다. 라벤더가 왜 그 사람에게 90점을 매겼는지, 왜 같은 이름을 가진 민방위 요원을 끌어왔는지를, 도장을 찍는 정보장교는 알 수 없었습니다. 적십자위가 블랙박스라 부르는 그 불투명함은, 자율무기체계든 의사결정 지원 시스템이든 가리지 않고 같은 자리에

고입니다. 설명할 수 없는 점수 위에 찍힌 도장은, 그 점수가 틀렸을 때 누구의 책임인지도 설명할 수 없게 만듭니다.

그런데 이 구별이 위안이 되지 않습니다. 오히려 더 까다로운 물음을 던집니다. 자율무기체계는 적어도 논쟁의 한복판에 있습니다. 적십자위가 금지하자고 하고, 유엔 사무총장이 금지하자고 하는 그 대상이 바로 그것입니다. 그러나 라벤더 같은 의사결정 지원 시스템은 그 금지 논의의 한가운데를 비껴갑니다. 적십자위의 2025년 입장문조차 의사결정 지원 시스템은 "이 보고서의 범위 밖"이라고 명시합니다. 방아쇠가 사람에게 있으니 자율무기가 아니라는 것입니다. 사람이 사슬 안에 있다는 사실과, 사람이 실제로 판단하고 있다는 사실은 같은 말이 아닙니다. 우리는 제3장에서 이미 이 말을 했습니다. 미나브의 정보장교에게 좌표 하나당 1분이 주어졌을 때, 가자의 정보장교에게 표적 하나당 20초가 주어졌을 때, 그 사람은 사슬 안에 있었습니다. 다만 그의 판단이 비어 있었습니다. 법이 보호하려는 것은 사람의 위치가 아니라 사람의 판단인데, 법의 그물은 위치를 잡고 판단은 놓칩니다.

2. 의미 있는 인간의 통제

이 빈틈을 가장 먼저 정확히 짚은 것은 정부도, 군대도 아니었습니다. 영국의 작은 비영리 단체였습니다. 아티클36(Article 36)이라는 곳입니다. 이름은 제네바협약 제1추가의정서 제36조에서 따왔습니다. 새로운 무기를 개발하거나 도입할 때, 그것이 국제법에 맞는지 검토하라고 정한 조항입니다. 이 단체가 2013년 무렵에 한 가지 표현을 만들어 냈습니다. "개별 공격에 대한 의미 있는 인간의 통제(meaningful human control over individual attacks)"입니다.

이 표현은 빠르게 퍼졌습니다. 여러 나라가 유엔 회의에서 이 말을 썼고, 인공지능 연구자들이 서명한 공개서한에도 등장했습니다. 그런데 이 말이 퍼지자, 비판도 따라왔습니다. '의미 있는'이라는 말이 모호하다는 것입니다. 무엇이 의미 있는 통제이고 무엇이 의미 없는 통제인지 누가 정하느냐는 물음이었습니다. 아티클36의 리처드 모이스는 이 비판이 핵심을 놓쳤다고 봅니다. '의미 있는'이라는 형용사는 결론이 아니라 출발점이라는 것입니다. 그것은 빈칸을 메우라는 신호입니다. 인간의 통제가 어떤 형태여야 하는지를 국제사회가 함께 정의해 나가자는 초대입니다.

모이스가 의미 있는 통제의 밑바닥에 깔아 둔 생각은 짧으면서도 단단합니다. 두 가지 전제에서 출발합니다. 첫째, 사람의 통제가 전혀 없이 힘을 가하는 기계는 받아들일 수 없다는 것입니다. 둘째, 사람이 컴퓨터가 내놓은 표시를 보고 그저 '발사' 버튼을 누르기만 하는 것은, 인지적 명료함도 자각도 없이 누르는 것은, 의미 있는 통제라고 부를 수 없다는 것입니다. 이 두 번째 전제를 다시 읽어 보십시오. 사람이 컴퓨터의 표시를 보고 버튼만 누르는 것. 가자의 20초가, 미나브의 1분이 바로 이 문장 안에 들어 있습니다. 모이스는 2013년에 이미 그 장면을 미리 그려 놓았습니다.

그러면 통제가 의미 있으려면 무엇이 갖춰져야 할까요. 아티클36이 제시한 핵심 요소는 네 가지입니다. 첫째, 예측 가능하고 신뢰할 수 있으며 투명한 기술. 사용자가 그 기계가 어떻게 작동하는지 이해할 수 있어야 한다는 것입니다. 둘째, 사용자에게 주어지는 정확한 정보. 무엇을 이루려 하는지, 그 기술이 무엇인지, 어떤 맥락에서 쓰이는지에 대한 정보입니다. 셋째, 적시의 인간 판단과 행동, 그리고 제때 개입할 수 있는 여지. 넷째, 책임성. 일이 잘못됐을 때 책임을 물을 수 있는 구조입니다.

미국의 새로운국가안보센터는 같은 생각을 군의 언어로 옮겨, 의미 있는 통제를 세 가지 구성 요소로 정리합니다. 첫째, 사람이 무기 사용에 대해 정보에 근거한, 의식적인 결정을 내리고 있어야 합니다. 둘째, 사람이 자기 행동의 합법성을 확인할 만큼 충분한 정보를 가지고 있어야 합니다. 셋째, 무기가 제대로 설계되고 시험되어 있고, 사람이 제대로 훈련되어 있어, 의도한 대로 작동해야 합니다. 첫째 요소의 두 단어를 다시 읽어 보십시오. 정보에 근거한, 의식적인. 버튼을 누르는 손가락이 있다고 해서 그 결정이 의식적인 것은 아닙니다. 좌표가 뺏고 시간이 1분이라면, 그 손가락은 누르되 의식은 따라가지 못합니다.

이 네 가지를 미나브에 대 봅니다. 예측 가능하고 투명한 기술이었습니까. 그 시스템이 왜 그 학교를 군사 목표물로 띄웠는지, 사람이 들여다보고 이해할 수 있었습니까. 정확한 정보가 주어졌습니까. 그 좌표가 언제 적인 것인지, 지금 그 건물이 무엇으로 쓰이는지를 정보장교가 알 수 있었습니까. 적시의 판단과 개입이 가능했습니까. 좌표 하나당 1분의 시간에, 표적 1,000개를 24시간에 처리하는 속도에, 사람이 멈춰 서서 "잠깐, 이 좌표가 맞는가" 하고 따져 물을 수 있었습니까. 책임을 물을 수 있었습니까. 우리는 제5장에서 그 답을 봤습니다. 만든 사람은 예측할 수 없었고, 명령한 사람은 예견할 수 없었으며 멈출 수 없었다고 말합니다. 네 요소가 모두 비어 있었습니다.

아티클36이 한 가지 더 짙은 것이 있습니다. 의미 있는 통제는 기술의 성능만으로 결정되지 않는다는 것입니다. 같은 기계라도 좁은 구역에서 짧은 시간 쓰면 통제가 가능하지만, 넓은 구역에서 긴 시간 풀어 놓으면 통제가 흐려집니다. 시간이 흐르는 동안 사람과 사물이 그 구역을 드나들기 때문입니다. 처음 작동시킬 때 비어 있던 학교에 한 시간 뒤 아이들이 들어찰 수 있습니다. 그래서 면적과 시간이 넓고 길어질수록 통제는 줄어듭니다. 이 통찰은 표적 1,000개라는 숫자, 24시간이라는 시간과 곧장 맞닿습니다. 규모가 커지고 속도가 빨라질수록, 통제는 묶어집니다.

학자들은 이 생각을 더 정교하게 다듬었습니다. 필리포 산토니 데 시오와 예룬 판 덴 호번은 의미 있는 통제를 두 가지 조건으로 나눕니다. 추적(tracking)과 소급(tracing)입니다. 추적이란, 시스템이 통제하는 사람의 의도와 판단에 제대로 반응해야 한다는 것입니다. 사람이 멈추라고 하면 멈춰야 하고, 사람이 정한 한계 안에서만 움직여야 합니다. 소급이란, 시스템의 행동을 거슬러 올라가면 그 행동을 이해하고 책임질 수 있는 사람에게 가닿아야 한다는 것입니다. 행동의 끝에 사람의 얼굴이 있어야 한다는 뜻입니다. 미나브의 타격을 거슬러 올라가면, 우리는 사람의 얼굴 대신 낡은 좌표 하나에 닿습니다. 소급의 사슬이

끊어진 자리, 그것이 빈자리입니다.

이 두 조건이 왜 중요한지를, 남아프리카공화국 출신 법학자 톰슨 쉥게타가 한 문장으로 짚습니다. 뉴욕대 국제법저널에 실린 그의 글은, 의미 있는 통제를 논하는 가장 큰 까닭이 무엇이냐고 묻고 스스로 답합니다. 자율무기가 "법적 책임의 공백을 만들어 낼 수 있다는 것"이라고. 그가 쓴 단어가 공백(vacuum)입니다. 우리가 이 책 내내 빈자리라 불러 온 바로 그것입니다. 쉥게타의 처방은 분명합니다. 사람이 가진 통제의 성격은, 그 사람이 "로봇의 이후 모든 행동에 대해 책임질 수 있도록" 보장하는 것이어야 한다고. 통제를 정의하는 목적이 곧 책임을 붙들어 두는 데 있다는 것입니다. 통제가 흐려지면 책임이 흘러내리고, 책임이 흘러내린 자리가 공백입니다. 의미 있는 통제는, 그 공백을 미리 막으려는 설계입니다.

같은 입문서는 이 논의를 현실의 군 운용 쪽으로 한 걸음 더 끌어내립니다. 통제는 한 사람의 손끝에서 일어나는 일이 아니라고 말합니다. 무기를 설계한 사람, 시험한 사람, 작전 규칙을 정한 사람, 표적 정보를 만든 사람, 마지막으로 버튼을 누른 사람, 이 모두가 하나의 통제 체계를 이룹니다. 그런데 이 말은 양날입니다. 통제가 여러 사람에게 나뉘어 있다는 것은, 책임도 여러 사람에게 나뉘어 있다는 뜻이기 때문입니다. 우리가 제5장에서 본 그대로입니다. 모두가 결정의 한 조각을 쥐고 있고, 누구도 결정 전체를 쥐고 있지 않습니다. 의미 있는 통제라는 개념이 풀려는 문제가 바로 이 분산입니다. 통제를 한 지점으로 다시 모아, 책임이 흘러내릴 틈을 막자는 것입니다.

여기서 한쪽을 분명히 들겠습니다. 사람이 사슬 안에 있다는 사실만으로 의미 있는 통제가 충족된다는 주장은, 받아들이기 어렵습니다. 그 주장은 위치를 통제로 착각합니다. 의자에 사람이 앉아 있는 것과, 그 사람이 판단하고 있는 것을 같은 것으로 셉니다. 가자와 미나브가 보여 준 것은 정확히 그 둘이 다르다는 사실이었습니다. 의미 있는 통제라는 잣대는 그 차이를 측정하는 자입니다. 사람이 앉아 있느냐가 아니라, 사람이 판단할 시간과 정보와 권한을 가졌느냐를 묻습니다. 이 잣대로 재면, 우리가 본 여러 장면은 통제를 충족하지 못합니다.

여기서 한 가지 불편한 사실을 마주해야 합니다. 의미 있는 통제를 지키려면, 그 통제를 담당하는 사람이 필요합니다. 표적 절차에서 법을 따지는 사람, 군 법무관입니다. 그런데 이 사람들의 자리가 흔들리고 있다는 보도가 나옵니다. 미국의 국방장관 피트 헤그세스는

'최대 살상력 대 미온적 합법성(maximum lethality vs tepid legality)'이라는 구도를 내세운 것으로 전해집니다. 군의 임무는 적을 죽이는 것이지만 법률 자문이 발목을 잡게 두는 것이 아니라는 기조입니다. 저스트시큐리티와 밀리터리닷컴 같은 매체의 보도에 따르면, 그는 군 법무 조직의 수장들을 교체했고, 표적 절차에서 법무관의 감독을 약화시키는 방향으로 움직였습니다.

이 대목에는 한 가지 역설이 깔려 있습니다. 영국의 지리학자 크레이그 존스는 '전쟁 법률가들(The War Lawyers)'이라는 책에서 군 법무관의 두 얼굴을 그립니다. 한편으로 군 법률가는 폭력을 합법화하는 사람입니다. 표적 절차에 법의 도장을 찍어, 그 타격을 적법한 것으로 만들어 줍니다. 어떤 면에서 그들은 더 많은 타격을 가능하게 합니다. 합법이라는 보증이 붙으면, 군은 더 자신 있게 방아쇠를 당길 수 있기 때문입니다. 그러나 다른 한편으로, 그 법률가가 절차 안에 있다는 사실 자체가 하나의 제동 장치였습니다. 그가 "이 표적은 안 됩니다"라고 말할 수 있는 자리가 있었다는 것, 그것만으로도 무언가가 한 번 걸러졌습니다.

헤그세스의 기조가 위험한 것은 이 지점입니다. 표적의 속도는 기계가 끌어올리고 있는데, 그 속도를 한 번 멈춰 세울 수 있던 사람마저 절차에서 밀려난다면, 빈자리는 더 넓어집니다. 기계가 좌표를 내려놓는 속도는 빨라지고, 그 좌표를 따져 물을 사람은 줄어듭니다. 의미 있는 통제라는 개념이 메우려던 그 틈을, 정반대 방향에서 넓히는 일입니다.

여기에는 한 가지 더 미묘한 함정이 있습니다. 군 법률가가 표적 절차에 도장을 찍어 그 타격을 합법으로 만들면, 그 도장은 책임의 무게를 옮기는 구실도 합니다. 무언가 잘못됐을 때, '법무 검토를 거쳤다'는 한마디가 방패가 됩니다. 사람이 따져서 합법이라 했으니, 그 판단은 사람의 것이라는 외양이 생깁니다. 그러나 그 사람이 기계가 내려놓은 점수와 좌표를 1분 안에 훑고 도장을 찍었다면, 그 외양은 속이 비어 있습니다. 법무의 도장이 책임의 자리를 채운 듯 보이지만, 실은 빈자리 위에 합법이라는 뚜껑을 덮은 것일 수 있습니다. 의미 있는 통제가 묻는 것은 도장이 찍혔느냐가 아니라, 그 도장 아래에 진짜 판단이 있었느냐입니다.

3. 마르텐스 조항과 인류의 양심

법에는 빈틈이 있습니다. 어떤 무기가 새로 나왔는데 그것을 직접 금지하는 조문이 아직 없을 때, 그 무기는 법의 그물을 빠져나갑니다. 조문에 없으니 금지되지 않았다, 따라서 써도 된다. 이 논리를 막기 위해, 국제인도법은 오래전에 한 가지 안전장치를 마련해 두었습니다. 마르텐스 조항(Martens Clause)입니다.

이 조항이 태어난 자리에는 작은 역사가 있습니다. 1899년 헤이그 평화회의에서, 강대국과 약소국은 한 가지를 두고 부딪혔습니다. 점령지에서 무기를 든 주민을 어떻게 다룰 것이냐는 문제였습니다. 합의가 막혔을 때, 러시아 대표였던 법학자 표도르 마르텐스가 한 문장을 내놓았습니다. 명문 규정이 없는 빈자리를 무엇으로 채울 것이냐에 대한 답이었습니다. 그가 채워 넣은 것은 새로운 조문이 아니라, 조문 너머에 있는 기준이었습니다.

그가 다듬은 문구의 핵심은 이렇습니다. 명문 규정이 다루지 않는 경우에도, 민간인과 전투원은 "확립된 관습에서, 인도의 원칙에서, 공공양심의 명령에서" 나오는 국제법 원칙의 보호 아래 놓인다는 것입니다. 쉽게 풀면 이렇습니다. 법전에 명시적으로 적혀 있지 않다고 해서, 무엇이든 해도 되는 것은 아닙니다. 인류가 오랜 시간 쌓아 온 양심의 기준, 인도의 원칙이 여전히 사람을 보호합니다. 조문의 침묵이 허락의 신호는 아니라는 것입니다.

이 조항이 자율살상무기 논의에서 중요해진 까닭이 여기 있습니다. 자율무기를 직접 금지하는 조약은 아직 없습니다. 국제인도법은 그것을 한 글자도 명시적으로 언급하지 않습니다. 그래서 어떤 이들은 말합니다. 금지 조문이 없으니 자율무기는 합법이라고. 마르텐스 조항은 그 말을 정면으로 막아줍니다. 조문이 없어도, 인도의 원칙과 공공양심의 명령이 이 무기를 쟀다는 것입니다. 적십자위는 이 점을 거듭 끌어옵니다. 자율무기체계가 제기하는 가장 깊은 문제는 법적인 것 이전에 윤리적인 것이라고. 적십자위의 표현으로, 자율무기는 "삶과 죽음에 관한 인간의 결정을, 센서와 소프트웨어와 기계의 처리로 대체"합니다. 그리고 대부분의 사람이 동의하는 한 가지가 있다고 적십자위는 말합니다. 알고리즘이, 기계의 처리가, 누가 살고 누가 죽는지를 정해서는 안 된다는 것입니다.

이 윤리적 직관을, 적십자위는 구체적인 권고로 옮깁니다. 2021년 5월의 입장에서 시작해 2025년의 입장문까지, 그 핵심은 흔들리지 않습니다. 적십자위는 "새로운 법적 구속력

있는 규칙이 시급하다"고 말합니다. 그리고 그 규칙의 윤곽을 셋으로 나눕니다. 첫째, 예측 불가능한 자율무기체계는 명시적으로 배제되어야 합니다. 적십자위의 원문을 그대로 옮기면, "예측 불가능한 자율무기체계는 명시적으로 배제되어야 한다. 그 무차별적 효과 때문이다(Unpredictable autonomous weapon systems should be expressly ruled out, notably because of their indiscriminate effects)." 사용자가 그 효과를 충분히 이해하고 예측하고 설명할 수 없는 무기는, 적십자위의 말로 하면 "눈을 가린 채 쏘는 것"과 같습니다. 우리가 제5장에서 본 블랙박스가 바로 이 자리에 섭니다. 왜 그 사람이 표적이 됐는지 설명할 수 없는 시스템은, 무차별적이어서 이미 기존 법으로도 금지된다는 것입니다.

둘째, 사람을 직접 표적으로 삼는 자율무기, 그러니까 대인 자율무기는 금지되어야 합니다. 적십자위가 이 권고에 붙이는 말이 무겁습니다. 사람을 표적으로 삼는 일은 "인간의 생명을 센서 데이터와 기계 처리로 환원"하는 것이고, 그것은 "알고리즘에 의한 죽음, 살상의 자동화에서 마지막 변경"이라고. 우리가 이 책 전체에서 본 것이 바로 그 환원이었습니다. 라벤더가 사람을 1에서 100까지의 점수로 환원했을 때, 미나브의 좌표가 175명을 한 줄의 데이터로 환원했을 때, 그 환원의 끝에 적십자위가 그어 놓은 선이 있었습니다.

셋째, 그 외의 자율무기는 엄격히 제한되어야 합니다. 표적의 종류를 본래적 군사 목표물로 좁히고, 작전의 기간과 지리적 범위와 규모를 한정하고, 민간인이 없는 상황으로 사용을 묶고, 사람이 실시간으로 감독하고 제때 개입하고 작동을 멈출 수 있게 해야 한다는 것입니다. 이 셋째 항목을 다시 읽어 보면, 앞 절에서 본 의미 있는 통제의 네 요소와 거의 포개집니다. 적십자위의 법적 권고와 아티클36의 윤리적 틀이, 다른 길로 출발해 같은 자리에서 만납니다.

적십자위가 비례성과 예방의 원칙을 짚는 대목은, 미나브를 다시 한번 정조준합니다. 비례성의 원칙은, 적십자위의 표현으로, "사전(ex-ante)의 판단"을 요구합니다. 치기 전에, 그러니까 일이 벌어지기 전에 미리, 예상되는 민간 피해를 따져 군사적 이득과 견주라는 것입니다. 그런데 자율무기는 누가, 언제, 어디서 표적이 될지를 작동 시점에 알지 못합니다. 그래서 사전의 판단이 어그러집니다. 예방의 원칙은 더 직접적입니다. 적십자위는 이렇게 씁니다. 공격을 앞두고 내린 지시만으로는 "그 자체로 충분한 예방 조치가 되지 못한다"고. 표적이 군사 목표물이 아니라는 게 드러나면, 공격을 취소하거나 멈출 의무가 끝까지 남기 때문입니다. 정보는 "공격 개시 순간까지, 그리고 가능한 한 공격이 진행되는 동안에도"

모으고 갱신해야 한다고 적십자위는 말합니다. 미나브의 좌표는 그 갱신을 거치지 못한 낡은 정보였습니다. 한 번 입력된 지시가, 그것이 틀렸다는 신호를 만날 기회 없이 폭탄으로 떨어졌습니다.

국제 무대에서도 같은 방향의 움직임이 있습니다. 유엔 사무총장 안토니우 구테흐스는 2024년의 보고서에서 한 문장을 남겼습니다. 그 보고서 자체가 무게를 지닙니다. 유엔 총회의 요청으로, 사무총장이 회원국과 옵서버국의 의견을 모아 펴낸 공식 문서이기 때문입니다. 그 안에서 구테흐스는 이렇게 적습니다. "사람의 생명을 빼앗을 힘과 재량을 가진 기계는 정치적으로 용납할 수 없고 도덕적으로 혐오스러우며, 국제법으로 금지되어야 한다." 그리고 시한을 붙입니다. 2026년까지, 인간의 통제 없이 작동하고 국제인도법을 지킬 수 없는 자율무기를 금지하고 나머지를 규제하는, 법적 구속력 있는 문서를 맺자는 것입니다. 같은 보고서는 한 문장을 더 보탬니다. "기계가 인간을 자율적으로 표적으로 삼는 일은, 넘어서는 안 되는 도덕의 선이다."

이 흐름에 힘이 실려 있다는 사실은 숫자로도 드러납니다. 유엔뉴스와 미국국제법학회의 정리에 따르면, 2023년 말 유엔 총회에서 자율무기 결의에 156개국이 찬성했습니다. 압도적 다수입니다. 한편으로는 신중한 목소리도 함께 기록해 뒤야 공정합니다. 사무총장 보고서에 의견을 낸 나라들 가운데 일부는, 새로운 구속력 있는 문서까지는 필요 없고 기존 국제인도법으로 충분하다고 봤습니다. 군사 강국 가운데 일부는 자율무기를 전략적 자산으로 여기며 협상의 속도를 늦추려 합니다. 특정재래식무기금지협약 아래 정부전문가그룹(CCW GGE)이 여러 해째 논의를 이어 왔지만, 만장일치를 요구하는 그 틀 안에서 합의는 더디게 움직입니다. 156개국의 찬성과 몇몇 강국의 제동이 같은 무대 위에 있습니다.

한국은 이 무대에서 어디에 서 있을까요. 국가안보전략연구원의 김현중이 2024년에 쓴 '넥스트 오픈하이머 시대'라는 보고서가 한국의 시각을 보여 줍니다. 제목의 오픈하이머는 핵을 만든 그 과학자입니다. 히로시마와 나가사키를 본 뒤, 그는 자신이 만든 무기의 군비통제에 가장 앞장섰습니다. 이 보고서는 자율살상무기의 시대에 그와 같은 순간, 곧 '넥스트 오픈하이머 모멘트'가 다가온다고 봅니다. 그런데 이 보고서가 원하는 방향은, 적십자위나 구테흐스의 금지론과는 결이 다릅니다.

보고서의 논리는 이렇습니다. 과거 핵의 시대에 세계는 핵을 가진 자와 못 가진 자로 갈렸습니다. 핵을 가진 나라들은 군비통제 협상의 당사자가 되었고, 못 가진 나라들은 비확산 체제의 수동적 대상으로 남았습니다. 협상 테이블에 앉으려면, 먼저 그 무기를 가져야 했다는 것입니다. 이 교훈을 자율무기에 옮기면, 한국이 군비통제 논의에서 주도적 위치에 서려면 자율살상무기를 선제적으로 개발하고 도입해야 한다는 결론이 나옵니다. 비확산의 대상으로 남지 말고, 통제의 당사자가 되자는 것입니다. 보고서는 인구 절벽으로 휴전선 인근 병력 배치가 어려워지는 현실을 들며, 전방에 감시 정찰 기능의 자율무기를 먼저 배치하고 부족한 점을 고쳐 나가자고 제안합니다. 'First come, first served'라는 격언까지 인용합니다.

이 시각을 그저 비난하고 넘어가는 것은 정직하지 않습니다. 그 안에는 현실의 무게가 있습니다. 줄어드는 병력, 휴전선이라는 조건, 강대국 사이에 낀 중간 국가의 처지. 협상력은 결국 가진 것에서 나온다는 냉정한 셈법도 틀린 말은 아닙니다. 다만 이 책이 가자에서, 미나브에서 본 것을 옆에 놓고 보면, 한 가지 물음이 떠오릅니다. '먼저 개발하고 나중에 고친다'는 그 논리는, 우리가 이 책에서 줄곧 따라온 한 단어와 너무 닮아 있습니다. lethal beta. 치명적인 시제품. 다 만들어지지 않은 채로 전장에 나가, 실전에서 데이터를 쌓으며 고쳐지는 무기. 미나브의 좌표가 남아 있던 것은, 그 시스템이 아직 완성되지 않았기 때문이기도 했습니다. 먼저 배치하고 부족한 점을 고친다는 말은, 그 부족함이 메워지기 전에 누군가 그 부족함의 대가를 치른다는 말이기도 합니다. 미나브의 175명이 그 대가였습니다.

보고서의 또 한 대목은 자율무기 운용에서 사람의 개입 정도를 세 단계로 나눕니다. 사람이 모든 단계를 승인하는 적극 개입, 사람이 중요한 행동에만 개입하는 소극 개입, 사람이 빠진 완전 자율. 그리고 한 가지를 내다봅니다. 국가 간 힘의 균형에 민감한 전략급 무기에는 사람의 개입이 요구되겠지만, 그 외의 개인 화기나 소규모 부대급 무기는 완전 자율로 개발될 수 있다고. 전략무기에는 손을 얹어 두되, 작은 무기는 풀어 놓을 수 있다는 전망입니다. 이 전망이 사실이라면, 사람을 직접 겨누는 가장 흔한 무기들이야말로 사람의 통제 밖으로 먼저 풀려난다는 뜻이 됩니다. 적십자위가 가장 먼저 금지하자고 한 바로 그것, 사람을 직접 표적으로 삼는 무기가 말입니다.

한국의 선택은 이 책이 대신 내릴 수 있는 것이 아닙니다. 다만 그 선택을 할 때, 'First come, first served'의 뒷면에 무엇이 적혀 있는지는 함께 봐야 합니다. 먼저 차지하는

자에게 돌아가는 이득의 반대편에, 먼저 시험되는 시스템의 오류를 먼저 겪는 사람들이 있습니다. 그 사람들은 대개 협상 테이블에 앉지 못합니다. 미나브의 부모가 그러했듯이.

다시 마르텐스 조항으로 돌아갑니다. 인도의 원칙과 공공양심의 명령. 이 오래된 문구가 자율무기 시대에 묻는 것은 결국 한 가지입니다. 우리가 무엇을 견딜 수 있고 무엇을 견딜 수 없는가. 법전에 조문이 적히기 전에도, 인류의 양심에는 넘어서는 안 될 선이 그어져 있다는 믿음입니다. 적십자위가 사람을 직접 표적으로 삼는 무기를 두고 "넘어서는 안 되는 도덕의 선"이라 부를 때, 구테흐스가 같은 말을 할 때, 그들은 마르텐스가 1899년에 마련해 둔 그 자리에 서 있습니다.

마르텐스 조항이 자율무기 시대에 새 조약의 근거가 되는 까닭도 여기 있습니다. 적십자위가 든 역사의 사례들을 보면, 인류는 어떤 무기는 널리 쓰이기 전에 미리 금지했습니다. 1868년 상트페테르부르크 선언이 특정 폭발탄을 금지했고, 1995년에는 사람의 눈을 영구히 멀게 하는 실명 레이저 무기를 금지했습니다. 둘 다 그 무기가 전장에 흔해지기 전에 그어진 선이었습니다. 적십자위의 표현으로, "고통이 이미 분명해진 다음에 대응하는 것이 아니라, 고통을 미리 막은" 드문 성공이었습니다. 대인지뢰의 경우는 그 반대였습니다. 무차별적 피해가 쌓이고 또 쌓인 다음에야 전면 금지에 이르렀습니다. 적십자위가 대인 자율무기를 지금 금지하자고 하는 것은, 대인지뢰의 길이 아니라 실명 레이저의 길을 가자는 것입니다. 피해가 쌓이기 전에 선을 긋자는 것입니다. 마르텐스 조항은 그 선을 그을 자리, 곧 조문이 비어 있는 그 빈자리를 인도의 원칙으로 메우라고 말합니다.

그러나 양심은 법이 아닙니다. 마르텐스 조항은 사람을 보호하는 원칙이되, 그 원칙을 위반한 자를 법정에 세우는 조문은 아닙니다. 공공양심의 명령은 강력하지만, 그 명령을 어겼을 때 누가 어떻게 책임지는지를 말해 주지 않습니다. 여기서 우리는 제5장의 그 빈자리와 다시 만납니다. 책임이 없으면 억지도 없고, 피해자에 대한 응보도 없다던 그 문장. 마르텐스 조항은 선을 그어 줍니다. 그러나 그 선을 넘었을 때 누가 책임지는가 하는 물음은, 여전히 답을 기다립니다.

그렇다면 국제인도법은 기계를 멈출 수 있는가. 이 장의 제목이 던진 물음입니다. 정직하게 답하면, 법은 멈출 도구를 이미 손에 쥐고 있습니다. 구별의 원칙, 비례성의 원칙, 예방의 원칙. 이 셋은 기술과 무관하게 적용되는 원칙입니다. 적십자위의 말대로, 국제인도법은

"기술 중립적"이어서, 자율무기를 한 글자도 언급하지 않으면서도 이미 그것을 강하게 구속합니다. 예측 불가능한 무기는 무차별적이어서 금지되고, 사람을 직접 치는 무기는 구별의 원칙과 충돌합니다. 법은 이미 많은 것을 말하고 있습니다.

문제는 법이 무엇을 말하느냐가 아니라, 그 법을 누가 적용하느냐입니다. 법의 세 원칙은 모두 사람의 판단을 전제로 합니다. 그 판단이 20초로, 1분으로 줄어들 때, 좌표 1,000개가 24시간에 쏟아질 때, 법의 원칙은 자리에 그대로 있는데 그것을 적용할 사람의 자리가 비어 갑니다. 법은 멈출 수 있습니다. 다만 법이 멈추려면, 멈출 시간을 가진 사람이 사슬 안에 있어야 합니다. 그 사람이 정보와 권한과 시간을 쥐고 있어야 합니다. 의미 있는 통제라는 말이, 마르텐스 조항이라는 오래된 문구가, 적십자위와 구테흐스의 금지론이 모두 가리키는 곳은 그 한 자리입니다. 사람의 판단이 들어설 자리.

이 책은 그 자리가 비어 가는 과정을 좇아왔습니다. 가자의 책상에서, 우크라이나의 들판에서, 이란의 학교 위에서, 카라카스의 정전된 밤에서. 그리고 그 빈자리를 메우는 일이 법의 몫이라고, 제5장의 끝에서 말했습니다. 이 장을 지나오며 우리는 법이 그 일을 할 수 있다는 것을, 그러나 아직 하지 못하고 있다는 것을 함께 봤습니다. 조약은 2026년을 시한으로 논의되고 있지만, 그 테이블에는 156개국의 찬성과 몇몇 강국의 제동이 함께 앉아 있습니다. 그리고 그 논의가 이어지는 동안에도, 가자에서 검증된 시스템은 상품이 되어 다른 전장으로 팔려 나갑니다.

법이 따라잡으려는 그 무기는, 지금 이 순간에도 어딘가로 수출되고 있습니다. 다음 장에서, 그 수출되는 전장을 봅니다.

인간의 통제력과 국제법의 마지막 마지막 방어선



규범의 최전선 (마르텐스 조항)

명시된 조문이 없더라도 인류의 양심과 인도의 원칙이 무기의 한계를 설정합니다.

- 1899년 헤이그 평화회의에서 표도르 마르텐스가 제안한 조항은 자율살상무기 논의의 근본적 윤리 기준이 되었습니다.
- 유엔 사무총장 안토니우 구테흐스는 2024년 보고서를 통해 2026년까지 자율무기를 금지하는 법적 문서를 맺자고 제안했습니다.
- 2023년 말 유엔 총회 결의에서 156개국이 찬성하며 예측 불가능한 기계의 살상을 막는 데 동의했습니다.

인간의 통제 (의미 있는 개입)

버튼을 누르는 물리적 위치가 아닌, 정보에 근거한 의식적 판단만이 합법적 통제입니다.

- 영국 단체 아티클36은 2013년 기술의 투명성, 정확한 정보, 적시의 판단, 책임성이라는 4가지 통제 요소를 정립했습니다.
- 필리포 산토니 데 시오 등은 통제 조건을 시스템을 다루는 추적(Tracking)과 책임을 묻는 소급(Tracing)으로 나누었습니다.
- 표적 1,000개를 24시간에 처리하는 시스템 안에서는 인간의 추적과 소급이 모두 불가능해집니다.

기계의 한계 (사라진 맥락)

센서는 수치와 형태를 식별할 뿐, 상황의 변화와 인간의 의도를 해석하지 못합니다.

- 국제적십자위원회(ICRC)는 2025년 10월 입장문에서 기계가 무기를 내리는 항복과 무기를 들어 올리는 공격을 구별하지 못한다고 지적했습니다.
- 사물의 군사적 목표 여부는 시간에 묶여 변하지만, 기계는 낡은 좌표 하나로 175명이 있는 학교를 공격했습니다.
- 구별, 비례성, 예방이라는 국제인도법의 세 원칙은 모두 수학 공식으로 환원할 수 없는 실시간 맥락 판단을 요구합니다.

© NotebookLM

에필로그 lethal beta, 수출되는 전장

테크 저널리스트 제이콥 워드는 그것을 'lethal beta'라고 불렀습니다. 상상하는 베타. 소프트웨어 업계에서 베타라는 말은 아직 완성되지 않은 시험판을 뜻합니다. 버그가 남아 있고, 사용자가 쓰면서 오류를 신고하면 다음 버전에서 고치는 단계입니다. 우리가 매일 쓰는 앱도 한때는 다 베타였습니다. 메시지가 가끔 늦게 가고, 화면이 한 박자 늦게 뜨고, 그러면 어딘가에 신고가 쌓이고, 다음 업데이트에서 조용히 고쳐집니다. 베타란 원래 그런 것입니다. 미완성을 세상에 먼저 내보내 놓고, 쓰는 사람들의 불편을 모아 완성으로 다가가는 방식.

그런데 워드가 가리킨 베타는 사람을 죽입니다. 시험판인데, 신고하는 쪽이 죽는 시험판입니다. 오류를 알리려면 먼저 그 오류의 대상이 되어야 합니다. 명단에 잘못 오른 사람이 자기가 잘못 올랐다고 항의할 길은 없습니다. 그 사람은 이미 사라진 뒤입니다. 버그 리포트를 보낼 수 있는 유일한 당사자가, 버그 그 자체가 되어 사라지는 구조. 워드의 두 단어는 그 구조를 정확히 찔렀습니다.

베타에는 또 하나의 약속이 숨어 있습니다. 다음 버전이 온다는 약속입니다. 지금은 불완전하지만 곧 나아진다는 믿음. 그 믿음이 있어서 우리는 베타를 견딥니다. 상상하는 베타에도 같은 약속이 따라붙습니다. 오류율 10퍼센트는 다음 버전에서 5퍼센트가 되고, 5퍼센트는 다시 2퍼센트가 된다고. 더 나은 데이터, 더 빠른 처리, 더 정밀한 좌표. 숫자는 정말로 좋아질지 모릅니다. 그런데 여기에 함정이 있습니다. 숫자가 좋아질수록, 사람은 더 믿습니다. 더 믿을수록, 20초는 10초가 되고 10초는 5초가 됩니다. 정밀해진 만큼 빨라지고, 빨라진 만큼 사람이 끼어들 틈이 줄어듭니다. 베타가 정식 버전이 되는 날, 인간은 화면 앞에 앉아 있되 더는 판단하지 않게 됩니다. 그저 누르는 손이 됩니다.

이 표현이 어디서 나왔는지를 생각하면 등이 서늘해집니다. 가자에서 라벤더라는 시스템이 사람을 명단에 올렸습니다. 정보장교는 그 명단을 받아 20초 만에 도장을 찍었습니다. 그 시스템에는 약 10퍼센트의 오류율이 있었다고 보도되었습니다. 열 명 중 한 명은 엉뚱한 사람일 수 있다는 뜻입니다. 보통의 소프트웨어라면 출시를 미룰 숫자입니다. 결제 앱이 열 번에 한 번 엉뚱한 계좌로 돈을 보낸다면 그 앱은 시장에 나오지 못합니다. 그런데 전장에서는 나왔습니다. 열 번에 한 번 엉뚱한 사람을 가리키는 시스템이, 사람의 생사를 가르는 자리에서 돌아갔습니다. 가자가 시제품을 굴리는 자리였습니다.

시제품은 가자에서 멈추지 않았습니다. 같은 계보의 표적 시스템이 2026년 이란 상공에서 돌았습니다. 60초 만에 한 사람의 생사를 판정하는 화면이 떴고, 미나브의 한 학교가 낡은 좌표 하나 때문에 표적 목록에 남았습니다. 가자가 시제품이었다면, 이란은 무대였습니다.

무대라는 말을 저는 가볍게 쓰지 않았습니다. 시제품은 골방에서 만듭니다. 무대는 사람들이 봅니다. 이란 작전의 화면은 미국의 작전실에서 돌았고, 그 결과는 위성과 통신을 타고 동맹국의 화면으로 번졌습니다. 잘 도는 시스템은 자랑거리가 됩니다. 이만큼 빠르고, 이만큼 정밀하다고. 무대에 한 번 오른 도구는 관객을 만듭니다. 관객은 곧 구매자가 됩니다.

무대 다음에는 수출이 옵니다. 이 책에서 좇은 시스템들은 이미 여러 나라의 군대와 정보기관에 팔리고 있습니다. 한 전장에서 검증된 도구는 다음 전장으로 옮겨 갑니다. 우크라이나의 들판에서 드론이 눈으로 표적을 익히는 법을 배웠고, 가자에서 명단이 사람의 속도를 기계의 속도로 갈아치웠고, 이란에서 그 둘이 한 화면에 합쳐졌습니다. 가자에서 돌던 것이 이란에서 돌고, 이란에서 돌던 것이 다음 어딘가에서 돕니다. 다음은 어디일까요. 그 질문에 답을 가진 사람을, 저는 취재 내내 만나지 못했습니다. 정확히 말하면, 답을 가졌다는 사람은 더러 만났습니다. 다만 그들의 답은 저마다 자기 회사, 자기 부대, 자기 계약의 경계에서 멈췄습니다. 경계 너머로 그 도구가 어디까지 갈지를 묻자, 모두가 비슷한 표정을 지었습니다. 그건 내 몫이 아니라는 표정.

여기서 이 책이 끝까지 놓지 못한 한 가지 모순을 꺼내야겠습니다.

이 기술을 만들어 파는 사람들조차, 어떤 선 앞에서는 멈춰 섰습니다. 클로드라는 인공지능을 만드는 회사 앤스로픽의 대표 다리오 아모데이는 미국 국방부와 일하면서도 두 가지에는 선을 그었습니다. 그의 회사는 다른 어떤 인공지능 회사보다 앞서서 군과 일했습니다. 모델을 기밀 클라우드에 처음 올린 것도, 국가안보용 맞춤 모델을 처음 만든 것도 자기들이라고 그는 말했습니다. 사이버 작전, 전투 지원, 정보 분석. 군과 정보기관 곳곳에 그 회사의 모델이 깔려 있습니다. 적으로부터 나라를 지켜야 한다고, 중국과 러시아 같은 권위주의 국가로부터 지켜야 한다고 그는 믿었습니다.

그런 그가 국방부에 말했습니다. 당신들이 하려는 일의 98퍼센트, 99퍼센트는 괜찮다고. 그런데 두 가지는 안 된다고. CBS 인터뷰에서 그는 그 두 줄을 또박또박 말했습니다.

하나는 대규모 국내 감시입니다. 민간 기업이 모은 데이터를 정부가 사들여 인공지능으로 한꺼번에 분석하면, 위치와 정치 성향까지 묶은 국민 개개인의 프로필이 만들어질 수 있습니다. 인공지능 이전에는 쓸모가 없어서 아무도 하지 않던 일이, 이제는 가능해졌다고 그는 말했습니다. 기술이 법보다 빨리 달려서, 법의 빈틈으로 빠져나가는 일. 그는 그것이 미국의 민주적 가치와 양립할 수 없다고 했습니다.

다른 하나는 완전 자율무기입니다. 사람이 전혀 개입하지 않고 스스로 쏘는 무기. 우크라이나에서 쓰이는 부분 자율무기가 아니라, 인간이 방아쇠에서 완전히 빠진 무기를 그는 가리켰습니다. 그는 이렇게 말했습니다. 오늘의 인공지능은 그것을 맡길 만큼 신뢰할 수 없다고. 모델에는 근본적인 예측 불가능성이 있고, 인공지능을 다뤄 본 사람이라면 누구나 그걸 안다고. 우리는 그 불안정성을 기술적으로 풀지 못했다고. 그러면서 그는 한 가지를 분명히 했습니다. 완전 자율무기에 원칙적으로 반대하는 건 아니라고. 적이 가지면 우리도 언젠가 필요할지 모른다고. 다만 지금은 신뢰성이 거기에 못 미치고, 감독에 대한 대화를 아직 못 했을 뿐이라고.

아모데이의 말 가운데 한 대목이 오래 남았습니다. 그는 책임의 문제를 들었습니다. 천만 대의 드론을 한 사람이, 혹은 소수가 통제한다고 상상해 보라고. 누가 버튼을 쥐고 있는지, 누가 아니라고 말할 수 있는지. 그 대화를 우리는 아직 한 번도 제대로 나눈 적이 없다고 그는 말했습니다. 인간 병사에게는 상식이 있고, 그 상식을 전제로 책임의 사슬이 짜여 있다고. 인공지능이 잘못된 사람을 표적으로 잡거나, 민간인을 쏘거나, 인간 병사라면 보였을 판단을 못 보는 일. 그런 일이 일어났을 때 누가 책임지느냐고. 그 사슬에서 인간을 빼면 무엇이 남느냐고.

선을 그은 사람의 말입니다. 그런데 현실은 그 말과 어긋난 채 흘러갑니다. 같은 회사의 클로드는 팔란티어의 시스템에 실려 이란과 베네수엘라 작전 화면에 떴습니다. 만든 사람은 완전 자율무기에 선을 그었는데, 그가 만든 모델은 표적을 고르는 화면의 한구석에서 돌아가고 있었습니다. 그가 그은 선은 '완전' 자율무기였습니다. 인간이 화면 앞에 앉아 도장을 찍는 한, 그 선은 넘은 게 아닙니다. 20초든 60초든, 사람이 누르기만 하면 됩니다. 선은 지켜졌습니다. 명목상으로는. 선을 긋는 일과, 그 선 안에서 벌어지는 일 사이의 거리. 그 거리가 이 책이 끝까지 메우지 못한 빈자리입니다.

팔란티어의 알렉스 카프는 다른 언어로 말했습니다. 2023년 뉴욕타임스 기고에서 그는 지금을 '오픈하이머의 순간'이라 불렀습니다. 핵폭탄을 만든 물리학자가 히로시마 이후 자신들이 죄를 알게 되었다고, 잃을 수 없는 앎이라고 했던 그 순간 말입니다. 카프는 그 비유를 빌려 와, 그러니 멈추자고 하지 않았습니다. 정반대였습니다. 멈추자는 주장은 잘못되었다고 했습니다.

그의 논리는 이렇습니다. 적들은 멈추지 않는다. 기술의 위험을 두고 연극 같은 토론을 벌이며 그들이 기다려 줄 리 없다. 그들은 그냥 나아간다. 자유민주주의 사회가 이기려면 도덕적 호소만으로는 안 되며, 하드파워가 필요하다. 그리고 이 세기의 하드파워는 소프트웨어 위에 세워진다. 날카로운 도구가 우리를 향할까 두려워 만들기를 피해서는 안 된다. 인공지능 무기는 추구하고, 정련하고, 규제해야 한다. 추구하라, 정련하라, 규제하라. 세 동사가 한 문장에 나란히 놓였습니다. 멈추라는 말은 그 안에 없습니다.

카프의 글에는 솔직한 문장 하나가 있습니다. 팔란티어의 소프트웨어는 미국과 동맹국의 국방·정보기관이 표적 선정, 작전 계획, 위성 정찰에 쓴다고. 적을 제거하는 일을 돕는 능력이 곧 그 소프트웨어의 가치를 이루는 전제라고. 에둘러 말하지 않았습니다. 우리가 만드는 건 사람을 죽이는 일을 돕는 도구이며, 바로 그래서 값이 나간다는 것입니다. 그는 실리콘밸리의 엔지니어들을 꾸짖기도 했습니다. 광고를 더 잘 배치하는 알고리즘은 짜면서, 미군을 위한 소프트웨어는 만들지 않으려는 사람들. 그들이 누리는 부와 제국이 누구 덕에 가능했는지를 잊었다고. 값지 않더라도, 그 빛을 알기는 해야 한다고 그는 적었습니다.

두 사람을 나란히 놓아 봅니다. 한 사람은 선을 긋고, 한 사람은 날카로운 도구를 두려워해 만들기를 피해서는 안 된다고 말합니다. 한 사람은 신뢰할 수 없다고 멈춰하고, 한 사람은 적이 멈추지 않으니 우리도 갈 수밖에 없다고 밀어붙입니다. 두 사람 다 자기 말에 진심이었을 겁니다. 둘 다 애국을 말했고, 둘 다 권위주의 적국을 말했고, 둘 다 자유민주주의를 말했습니다. 그런데 두 회사의 제품은 같은 작전 화면 위에서 만났습니다. 말의 차이가 무엇이었던, 현실에서는 같은 좌표를 비추고 있었습니다. 선을 그은 사람의 모델도, 멈추지 말자던 사람의 플랫폼도, 미나브를 내려다본 그 화면 어딘가에 있었습니다.

그 둘 사이에서 베네수엘라의 한 장면이 떠오릅니다. 2026년 초, 카라카스의 작전에서도 같은 플랫폼이 돌았습니다. 그리고 그 작전을 둘러싸고 회사 안에서 균열이 생겼다는 보도가 나왔습니다. 어떤 엔지니어들은 자기가 짠 코드가 어디에 쓰이는지를 뒤늦게 알고

흔들렸습니다. 카프가 꾸짖었던 그 망설임이, 바로 그 회사 안에서도 터져 나온 셈입니다. 멈추지 말자고 말하는 일은 쉽습니다. 자기 손으로 짠 코드가 한 사람의 마지막 화면에 뒀다는 걸 알게 되는 일은, 그렇게 쉽지 않습니다. 선을 긋는 일과 망설이는 일은 회사 바깥에만 있는 게 아니었습니다. 같은 건물 안, 같은 책상 앞에도 있었습니다.

오픈하이머의 순간이라는 말을 저는 자꾸 곱씹게 됩니다. 오픈하이머에게는 읽을 수 없는 앎이 있었습니다. 죄를 알았다고 그는 말했습니다. 매사추세츠공대 강연에서, 폭탄을 만든 물리학자들은 죄를 알게 되었고 그것은 읽을 수 없는 앎이라고 했습니다. 그 앎에는 주어가 있었습니다. 그가 만들었고, 그가 알았습니다. 로스앨러모스의 실험실에 선 것도 그였고, 트리니티 실험장에서 첫 섬광을 본 것도 그였고, 히로시마 소식을 들은 것도 그였습니다. 죄를 짊어질 어깨가 거기 하나 있었습니다.

지금의 순간에는 그 주어가 흐릿합니다. 라벤더가 명단을 올렸을 때, 누가 그 사람을 죽였습니까. 명단을 만든 알고리즘입니까, 20초 만에 도장을 찍은 정보장교입니까, 그 시스템을 산 군대입니까, 그 군대에 모델을 판 회사입니까, 그 회사에 투자한 사람들입니까. 미나브의 학교가 낡은 좌표 때문에 표적 목록에 남았을 때, 그 좌표를 갱신하지 않은 책임은 누구의 것입니까. 좌표를 입력한 사람입니까, 입력값을 검증하지 않은 시스템입니까, 그 시스템을 믿으라고 가르친 절차입니까. 손가락은 여럿인데, 방아쇠를 당긴 손은 하나로 모이지 않습니다.

왜 흐릿해졌는지를 저는 취재 내내 곱씹었습니다. 한 가지 답은 자동화 편향이라는 말에 있습니다. 사람은 기계가 내놓은 답을 자기 판단보다 더 믿는 경향이 있습니다. 화면에 빨간 표시가 뜨면, 그 표시가 왜 떴는지를 따지기보다 그것을 받아들입니다. 시스템이 정밀할수록 그 믿음은 깊어집니다. 20초 안에 도장을 찍어야 하는 사람에게, 화면이 내놓은 판정을 다시 의심할 여유는 없습니다. 의심하지 않는 것이 그의 잘못일까요. 아니면 20초만 주고 의심하지 말라고 가르친 시스템의 잘못일까요. 그 사이 어딘가에서, 책임은 또 한 번 미끄러집니다.

또 하나의 답은 블랙박스라는 말에 있습니다. 카프조차 자기 글에서 인정했습니다. 최신 모델이 어떻게, 왜 작동하는지는 그것을 만든 과학자와 프로그래머에게도 분명하지 않다고. 1조 개가 넘는 변수가 만들어 내는 공간은 사람의 머리로 헤아릴 수 없다고. 그러니 그 모델이 한 사람을 명단에 올렸을 때, 왜 올렸는지를 설명할 수 있는 사람이 없습니다. 만든

사람도 모릅니다. 쓰는 사람도 모릅니다. 당한 사람은 더더욱 모릅니다. 이유를 아무도 모르는 판정이, 한 사람의 마지막 판정이 됩니다.

오픈하이머는 죄를 알았습니다. 알 주어가 있었으니까요. 라벤더가 만든 명단 앞에는 죄를 아는 주어가 없습니다. 모두가 자기 몫의 작은 일만 했다고 말할 수 있게 설계되어 있기 때문입니다. 정보장교는 시스템을 믿었다고 말합니다. 시스템을 만든 사람은 결정은 인간이 한다고 말합니다. 결정을 한 인간은 시간이 20초뿐이었다고 말합니다. 시간을 20초로 정한 사람은 작전 속도가 그걸 요구했다고 말합니다. 작전 속도를 정한 사람은 적이 빠르니 어쩔 수 없다고 말합니다. 책임이 사라진 게 아닙니다. 잘게 나뉘어, 어느 한 곳에도 무겁게 내려앉지 않을 만큼 흩어졌습니다.

그게 빈자리입니다. 비어 있어서 빈자리가 아니라, 모두가 한 발씩 물러서서 가운데가 비어 버린 자리입니다. 각자가 물러선 거리는 짧습니다. 한 발씩입니다. 그런데 그 한 발씩이 모이면, 가운데에는 아무도 없습니다. 죄를 아는 어깨가 없습니다. 오픈하이머가 가졌던 그 어깨가, 지금은 어디에도 없습니다.

법은 이 빈자리를 알고 있었습니다. 국제인도법은 전쟁에도 규칙이 있다고 말합니다. 군인과 민간인을 구별하라고, 얻는 군사적 이익에 견주어 지나친 피해를 주지 말라고. 이 규칙들은 사람이 판단하는 것을 전제로 쓰였습니다. 구별하는 눈, 견주는 저울, 멈추는 손. 그런데 명단을 기계가 만들고 판단을 20초로 압축하면, 그 눈과 저울과 손은 어디로 갑니까. 규칙은 그대로인데, 규칙을 지킬 주체가 흩어졌습니다. 19세기에 만들어진 마르텐스 조항은, 법에 명시되지 않은 경우에도 인류의 양심이 사람을 보호한다고 했습니다. 인류의 양심. 그 양심이 깃들 자리도 사람의 어깨였습니다. 어깨가 흩어지면, 양심도 흩어집니다.

책을 닫으며 한 교실을 다시 떠올립니다.

미나브. 학교 건물 하나. 위성에서 내려다본 화면 안에서 그곳은 한동안 표적이었습니다. 시스템이 가진 좌표는 남아 있었습니다. 좌표는 갱신되지 않았고, 화면 위에서는 60초가 흘렀고, 누군가가 판정했습니다. 그날 그 교실에 있었던 아이들의 이름과 나이를, 저는 확인된 범위 안에서만 적었습니다. 더 많은 묘사는 하지 않겠습니다. 숫자로 소비하지 않겠다는 것이 이 책의 약속이었으니까요.

취재하는 동안 저는 여러 사람에게 같은 질문을 던졌습니다. 그 좌표가 낡았다는 걸 누가 알았어야 했느냐고. 답은 매번 옆으로 비켰습니다. 좌표 갱신은 다른 부서 일이라고, 표적 검토는 작전 단계가 따로 있다고, 시스템은 입력값을 의심하지 않는다고. 누구도 거짓말을 하지 않았습니다. 모두 사실을 말했습니다. 그런데 그 사실들을 다 모아도, 그 교실을 지킬 사람은 나오지 않았습니다. 각자가 맡은 칸은 분명했고, 칸과 칸 사이는 비어 있었습니다. 미나브의 학교는 바로 그 빈칸으로 떨어졌습니다.

다만 한 가지만 남겨 두려 합니다. 그 교실에는 의자가 있었습니다. 작은 의자들. 그중 하나가 비었습니다. 그 빈 의자 앞에는 누가 앉아야 했을까요. 그 아이입니까. 아니면, 그 좌표가 낡았다는 걸 알았어야 할 누군가입니까. 60초 안에 판정한 손입니까. 그 시스템을 베타라고 부르며 출시한 회사입니까. 선을 그었다고 말한 사람입니까. 멈추지 말자고 말한 사람입니까.

저는 그 의자에 누구를 앉혀야 할지 모릅니다. 이 책은 거기에 답을 놓지 않습니다. 답을 아는 척하는 일이야말로, 그 빈자리를 손쉽게 메워 버리는 방법일 테니까요. 메우면 보이지 않습니다. 보이지 않으면 묻지 않게 됩니다. 묻지 않으면 다음 좌표도 낡은 채로 남습니다. 그리고 그 시스템은, 다음 무대로 수출됩니다.

가자는 시제품이었습니다. 이란은 무대였습니다. 다음은 어디일까요.

그 의자는, 아직 비어 있습니다.

클로드 지피티 팔란티어와 2026년 세계전쟁

전자책 발행 | 2026년 6월 26일

저자 | 김경진

펴낸이 | 김경진

펴낸곳 | 김경진 변호사 출판사

출판사등록 | 2025. 3. 10. (제2025-000015호)

주소 | 서울특별시 동대문구 전농로 91, 백일빌딩 304호

전화 | 02-6338-1905

이메일 | kimkj008@gmail.com

가격 : 20,000원

© 김경진 2026

본 책은 저작자의 지적 재산으로서 무단 전재와 복제를 금합니다.

참고) 이 책의 사실관계는 공개 보도와 1차 자료에 근거했으며, 자료 수집과 검증 과정에서 인공지능의 도움을 받았습니다. 미검증·상충하는 수치는 본문에서 출처와 함께 밝혔습니다.

이 책을 잘 읽으셨으면 그리고 새로운 가치있는 지식을 얻으셨다고 판단되시면
농협 302-1096-0948-81 (예금주 김경진) 에 자발적 후원 부탁드립니다.