WHEN THE ALGORITHM GETS IT WRONG

알고리즘의 실책

AIAAIC 사례로 본 AI 부작용과 윤리적 과제

채용과 복지, 법정과 도로, 교실과 일터에서 벌어진 일

김경진 변호사

목차

제 1 장 편향과 차별: 알고리즘에 이식된 인간의 편견	
1.1	채용·금융·서비스의 알고리즘 편향
1.2	복지 수급 및 사회 안전망의 알고리즘 오작동
1.3	안면 인식 기술의 유색인종 오인과 부당 체포
제 2 장 환각과 허위 정보: 생성형 AI의 신뢰성 위기	
2.1	생성형 AI의 환각과 가짜 도서·기사 유포
2.2	법정·학술계의 가짜 판례 및 인용문 파문
2.3	역사적·지리적 사실 왜곡과 사회적 논란
제 3 장 프라이버시 침해와 감시 사회: 무단 데이터 수집의 그늘	
3.1	무단 생체 정보 수집과 공공·상업적 감시
3.2	개인정보 무단 학습과 AI 대화 유출 사고
3.3	위치 추적과 노동자 과잉 모니터링
제 4 장 저작권과 창작자의 권리: 생성형 AI와 무단 학습 갈등	
4.1	저작권 침해 소송과 무단 데이터 스크래핑
4.2	AI 생성물의 저작권 부인과 예술 생태계 혼란
4.3	초상권·음성 클로닝과 무단 복제
제 5 장 자율주행과 로봇 공학: 물리적 안전과 인명 피해	
5.1	자율주행 차량 결함과 도로 교통사고
5.2	산업용·서비스용 로봇의 오작동 및 안전 사고
5.3	군사용 자율 무기 및 드론 기술의 치명적 위험
제 6 장 정서적 왜곡과 디지털 범죄: AI 챗봇과 악의적 활용	
6.1	AI 챗봇과의 과도한 정서적 교감 및 범죄 유도

6.2	딥페이크 성범죄 및 디지털 합성 사기
6.3	검색 알고리즘 조작과 다이내믹 프라이싱의 독점

제 1 장

편향과 차별: 알고리즘에 이식된 인간의 편견

1.1

채용·금융·서비스의 알고리즘 편향

거절 메일은 늘 새벽에 왔습니다.

데릭 모블리는 마흔이 넘은 흑인 남성이고, 불안 장애와 우울증을 앓고 있었습니다. 정보기술 분야에서 일했고, 자격증도 있었고, 경력도 있었습니다. 그는 워크데이라는 회사의 채용 시스템을 쓰는 기업 백여 곳에 지원서를 냈습니다.

백여 번 떨어졌습니다.

거절 메일이 도착한 시각을 보면 사람이 보낸 것이 아니라는 사실을 알 수 있었습니다. 새벽 한 시, 새벽 세 시. 사람 채용 담당자는 그 시간에 잠을 잡니다. 기계는 잠을 자지 않습니다. 커피도 마시지 않고, 미안해하지도 않습니다. 서류를 열어 보고 삼 초 만에 닫습니다.

모블리는 소송을 걸었습니다. 워크데이의 인공지능이 나이와 인종과 장애를 걸러 내는 체처럼 작동한다는 주장이었습니다. 법원은 워크데이가 기존 임직원 데이터로 인공지능을 학습시켜 보호받아야 할 특성을 대리 지표로 사용했다고 보고 전국 규모의 집단 소송을 승인했습니다.

2026년 3월, 법원은 연령차별금지법이 지원자에게는 적용되지 않는다는 워크데이 쪽 주장을 물리쳤습니다. 6월 16일에는 캘리포니아 밖에 있는 고용주가 쓴 경우에도 워크데이가 책임을 질 수 있다는 판결이 나왔습니다. 이 소송이 다루는 거절 건수는 십일억 건입니다.

십일억 번의 새벽 메일입니다.

이 이야기가 무서운 이유는 악당이 없다는 데 있습니다. 누구도 회의실에 모여 마흔 넘은 사람을 떨어뜨리자고 결의하지 않았습니다. 그냥 지난 십 년 동안 뽑은 사람들의 데이터를 기계에 먹였을 뿐입니다. 기계는 성실하게 그 패턴을 배웠습니다. 그리고 성실하게 반복했습니다.

기계 학습이라는 말은 어렵게 들리지만 원리는 초등학교 수학 시간과 비슷합니다. 규칙을 알려 주지 않고 답만 잔뜩 보여 주면서, 여기서 규칙을 찾아내라고 시키는 겁니다. 과거의 답

이 편향돼 있으면 기계는 편향을 규칙으로 배웁니다. 기계 입장에서는 잘한 일입니다. 시킨 대로 했으니까요.

십 년 전 아마존이 그 사실을 먼저 배웠습니다.

2014년 스코틀랜드 에든버러의 아마존 개발 센터에서 비밀 프로젝트가 시작됐습니다. 이력서를 넣으면 별점을 매겨 주는 시스템이었습니다. 별 하나부터 다섯까지. 물건에 매기는 별점을 사람에게 매긴 셈입니다.

개발자들은 지난 십 년 치 기술직 지원서를 학습 데이터로 넣었습니다. 그 십 년 동안 아마존이 뽑은 기술직은 대부분 남성이었습니다. 기계는 결론을 내렸습니다. 남성이 더 나은 지원자다.

기계는 이력서에서 여성이라는 단어를 찾으면 점수를 깎았습니다. 여성 체스 클럽 회장이라고 쓴 이력서가 감점됐습니다. 여자대학교를 나온 지원자들의 점수도 통째로 내려갔습니다. 반대로 남성 엔지니어들이 자주 쓰는 동사, 수행했다거나 포착했다 같은 말이 들어가면 점수가 올라갔습니다.

아마존은 이 편향을 고쳐 보려 했습니다. 실패했습니다. 2017년에 프로젝트를 폐기했습니다. 다행히도 그 시스템은 실제 채용에 쓰이지 않았다고 아마존은 밝혔습니다.

문제는 다른 회사들이 비슷한 물건을 계속 만들었다는 데 있습니다.

카일 빔은 밴더빌트 대학교 학생이었습니다. 양극성 장애 치료를 위해 학교를 잠시 쉬었고, 생활비를 벌려고 동네 가게에 지원했습니다. 크로거 슈퍼마켓, 홈디포, 월그린, 펫스마트. 시간제 계산대 일이었습니다.

전부 떨어졌습니다. 면접까지 가지도 못했습니다.

크로노스라는 회사가 만든 유니크루라는 온라인 성격 검사가 카일에게 빨간불을 켰기 때문입니다. 그 검사는 카일의 정신 건강 이력을 감지했고, 이런 사람은 고객이 화를 낼 때 고객을 무시할 가능성이 높다고 판정했습니다.

카일의 아버지 로랜드 빔은 변호사였습니다. 2014년 평등고용기회위원회에 진정을 냈고, 기업들을 상대로 정신 장애인 차별 소송을 냈습니다.

카일은 2019년에 스스로 생을 마감했습니다. 소송이 끝나기 전이었습니다.

성격 검사를 통과하지 못해 계산대에 설 수 없었던 청년이 있었다는 사실. 그 판정을 내린 것이 사람이 아니라 웹페이지였다는 사실. 그리고 그 웹페이지에는 항의할 창구가 없었다는 사실. 세 문장을 붙여 놓고 읽으면 등이 서늘해집니다.

아이튜터그룹의 시스템은 더 노골적이었습니다. 온라인 교육 회사인 이곳의 자동 채용 프로그램은 쉰다섯 살 이상 여성과 예순 살 이상 남성을 자동으로 탈락시켰습니다. 이백 명이 넘는 교사가 자격을 갖추고도 나이 때문에 잘렸습니다.

들통난 방식이 우스꽝스럽습니다. 한 지원자가 진짜 생년월일로 지원해 떨어졌습니다. 그런데 나이만 낮춰서 다시 지원하니 곧바로 면접을 보자는 연락이 왔습니다. 사람이라면 얼굴이 붉어졌을 텐데, 시스템은 태연했습니다. 평등고용기회위원회는 삼십육만 오천 달러의 합의금을 물렸습니다.

생성형 인공지능이 나오면 나아질 줄 알았습니다.

2024년 블룸버그 연구진은 GPT-3.5에게 똑같은 자격을 가진 가짜 이력서를 잔뜩 주면서 이름만 바꿨습니다. 인종과 성별이 짐작되는 이름들이었습니다. 결과는 이랬습니다. 기술직 심사에서 흑인 여성으로 짐작되는 이름은 36퍼센트 덜 선택됐습니다. 여성 이름은 인사관리 직무로 밀려났습니다.

2021년 독일 바이에른 방송국은 레토리오라는 스타트업의 인공지능 역량 평가를 실험했습니다. 같은 사람이 같은 말을 하는데 안경을 쓰면 점수가 달라졌습니다. 히잡을 쓰면 또 달라졌습니다. 뒤에 책장이 보이면 성실해 보인다는 판정이 나왔습니다.

책장을 세워 두면 성격이 좋아진다니, 웃깁니다. 그 판정으로 누군가는 일자리를 얻고 누군가는 얻지 못했다는 대목에서 웃음이 멈춥니다.

교단에서도 같은 일이 있었습니다. 2012년 워싱턴 디시 공립학교의 5학년 교사 사라 위소키는 임팩트라는 교원 평가 시스템에 의해 해고됐습니다. 동료 교사와 학부모들의 평가는 훌륭했습니다. 그러나 아이들이 전에 다니던 학교에서 시험 부정이 있었고, 그 때문에 부풀려진 점수가 위소키의 부가가치 점수를 끌어내렸습니다.

검증되지 않은 수식 하나로 이백다섯 명의 교사가 직장을 잃었습니다. 그 수식이 무엇이었는지는 공개되지 않았습니다.

돈을 다루는 알고리즘은 더 오래, 더 조용히 일했습니다.

2018년 캘리포니아 대학교 버클리 연구진이 낸 보고서가 있습니다. 2008년부터 2015년까지 패니메이와 프레디맥이 보증한 삼십 년 만기 고정금리 대출을 뜯어봤습니다. 대출 심사를 사람에서 알고리즘으로 넘기면 차별이 사라질 거라는 말이 그때 유행했습니다.

숫자는 다른 말을 했습니다. 흑인과 라티노 차입자는 구매 대출에서 5.6에서 8.6베이시스포인트, 재융자에서 3베이시스포인트를 더 냈습니다. 베이시스포인트는 0.01퍼센트를 뜻하는 금융 단위입니다. 작아 보이는 숫자입니다.

이 작은 숫자를 미국 전체로 합치면 연간 이억 오천만 달러에서 오억 달러가 됩니다.

이유는 더 씁쓸합니다. 대출 기관들은 소수 인종 차입자가 여러 상품을 비교할 시간과 정보가 부족하다는 사실을 인공지능으로 알아냈습니다. 그리고 그 사람들에게 조금 더 비싼 값을 매겼습니다. 전략적 가격 책정이라고 부릅니다. 그렇게 챙긴 추가 이윤이 11퍼센트에서 17퍼센트입니다.

탐사보도 매체 더마크업과 연합뉴스는 연방 주택 대출 데이터를 함께 분석했습니다. 클래식 파이코 점수로 돌아가는 비밀 승인 알고리즘은 유색인종 신청자를 백인보다 40퍼센트에서 80퍼센트 더 많이 거절했습니다. 일부 대도시에서는 250퍼센트를 넘었습니다.

스탠퍼드 경영대학원의 로라 블래트너와 스콧 넬슨은 여기서 한 층 더 내려갔습니다. 알고리즘에 나쁜 의도가 없어도 문제가 생긴다는 겁니다. 저소득층과 소수자에게는 신용 기록이 얇습니다. 기록이 얇으면 예측 정확도가 5에서 10퍼센트 떨어집니다. 오래전 연체 한두 건이 점수 전체를 망가뜨립니다.

가난하면 데이터가 적고, 데이터가 적으면 기계가 못 맞히고, 못 맞히면 위험한 사람으로 분류됩니다. 그리고 그 분류에는 객관적 신용 위험이라는 이름표가 붙습니다.

웰스파고 은행에서는 계산 실수 하나가 사람을 길거리로 내몰았습니다. 2010년부터 2015년까지 이 은행의 자동 모기지 재조정 심사 소프트웨어에는 변호사 비용을 잘못 계산하는 오류가 있었습니다. 오 년이 넘도록 아무도 몰랐습니다.

대출 조건을 바꿔 받을 자격이 충분했던 팔백칠십 명이 거절 통보를 받았습니다. 그중 최소 오백사십오 명이 집을 압류당했습니다. 은행은 2015년 말에 오류를 고쳤고, 그 사실을 몇 년 동안 밖에 알리지 않았습니다. 나중에 책정한 보상금은 팔백만 달러였습니다.

집 한 채 값을 생각하면 계산이 맞지 않는 금액입니다.

보험도 마찬가지였습니다. 가이코, 세이프코, 네이션와이드가 쓴 위험 평가 알고리즘은 사고 위험이 같아도 유색인종이 많이 사는 동네 주민에게 최대 30퍼센트 비싼 보험료를 매겼습니다. 올스테이트는 한술 더 떴습니다. 고객이 떠나지 않을 것 같으면 값을 올렸습니다. 오래 거래한 충성 고객이 더 비싸게 낸 겁니다. 규제 당국에 적발됐습니다.

일터에 들어간 다음에도 알고리즘은 따라옵니다.

2021년 이탈리아 볼로냐 노동법원은 배달 플랫폼 딜리버루의 프랭크 알고리즘이 차별적이라고 판결했습니다. 프랭크는 배달원의 신뢰도와 참여도를 점수로 매겼습니다. 정해진 근무에 나오지 못하면 점수가 깎였습니다. 이유는 묻지 않았습니다.

아파서 못 나와도 깎였습니다. 법으로 보장된 파업에 참여해도 깎였습니다. 점수가 깎이면 다음에 좋은 배차를 받지 못합니다. 파업할 권리는 법전에 남아 있었지만, 파업하면 굶는 구조가 코드 안에 있었습니다.

척커라는 회사의 자동 신원 조회 시스템은 2015년부터 우버, 리프트, 포스트메이츠 기사들을 무더기로 걸러 냈습니다. 이름이 같은 다른 사람의 범죄 기록, 이미 시효가 지난 기록이 그대로 붙었습니다. 하루아침에 계정이 막힌 기사들은 어디에 항의해야 하는지도 몰랐습니다.

캘리포니아 대학교 헤이스팅스 로스쿨의 비나 두발 교수는 여기에 이름을 붙였습니다. 알고리즘 임금 차별입니다. 우버와 리프트와 아마존이 노동자 데이터를 모아 같은 일에 서로 다른 돈을 준다는 겁니다. 옆 사람이 얼마 받는지 모르니 비교할 수도 없습니다.

승객 쪽도 편하지 않았습니다. 연구에 따르면 우버와 리프트의 요금과 배차 알고리즘은 유색인종이 많이 사는 지역에 더 비싼 요금과 더 긴 대기 시간을 안겼습니다.

페이스북 광고 시스템은 채용 공고를 나이로 걸렀습니다. 버라이즌, 아마존, 골드만삭스가 연령 타깃 기능을 써서 나이 많은 구직자에게는 공고가 보이지 않게 설정했습니다. 2021년

노스이스턴 대학교 연구와 2025년 프랑스 인권옹호관 판정은 여기서 한 걸음 더 나아갔습니다. 광고주가 성별을 지정하지 않아도, 페이스북 내부의 전달 알고리즘이 알아서 남성과 여성에게 다른 채용 공고를 보여 줬습니다.

물건값도 사람을 보고 정해졌습니다.

에어비앤비의 스마트 가격 책정은 숙소 요금을 자동으로 맞춰 주는 기능입니다. 카네기 멜런 대학교 파람 비르 싱 교수 연구진이 살펴보니, 백인 호스트들이 이 기능을 더 많이 켰고 그 결과 백인과 흑인 호스트의 소득 격차가 오히려 벌어졌습니다.

프린스턴 리뷰의 온라인 학원 수강료는 우편번호에 따라 달랐습니다. 아시아계 미국인이 많이 사는 지역에 더 비싼 값이 매겨졌습니다. 언론은 여기에 타이거 맘 세금이라는 이름을 붙였습니다.

여행 사이트 오비츠는 맥 컴퓨터를 쓰는 사람에게 더 비싼 호텔을 위에 띄웠습니다. 맥 사용자의 평균 소득이 높다는 통계 때문이었습니다. 노트북 뚜껑에 사과가 그려져 있으면 지갑이 두껍다고 판단한 겁니다.

아마존에는 프로젝트 네시라는 비밀 알고리즘이 있었습니다. 자기 상품 가격을 슬쩍 올려 보고 경쟁사가 따라 올리는지 지켜보는 실험이었습니다. 바이박스 알고리즘은 소비자에게 제일 싼 상품 대신 아마존에 수수료를 많이 주는 상품을 먼저 보여 줬습니다.

이 모든 사례에는 공통점이 하나 있습니다. 문제가 터졌을 때 회사들이 한 말이 거의 똑같았다는 점입니다.

알고리즘이 스스로 학습한 결과입니다. 기술적 오류입니다. 내부 구조는 영업 비밀이라 공개할 수 없습니다.

세 문장이면 충분했습니다. 데이터를 고른 사람도, 목표 함수를 정한 사람도, 그 결과로 이익을 본 사람도 회의실에 앉아 있었는데, 책임은 코드에게 갔습니다. 코드는 변명하지 않습니다. 사과도 하지 않습니다. 법정에 나오지도 않습니다.

세상에 이보다 편한 직원은 없습니다.

1.2

복지 수급 및 사회 안전망의 알고리즘 오작동

편지는 평범해 보였습니다. 호주 정부 봉투에, 정부 글씨체로, 정부 도장이 찍혀 있었습니다. 안에는 이렇게 적혀 있었습니다. 귀하는 복지 급여를 과다 수령했습니다. 아래 금액을 갚으십시오.

받은 사람들은 액수를 보고 두 번 놀랐습니다. 천만 원 넘는 금액이 적힌 편지도 있었습니다. 그리고 세 번째로 놀랐습니다. 왜 갚아야 하는지 설명이 없었습니다.

이 편지를 받은 사람이 사십만 명이 넘습니다.

호주 사회서비스부와 센터링크가 2016년부터 2019년까지 돌린 온라인 준수 개입 시스템, 사람들이 로보데트라고 부른 프로그램이 한 일입니다. 로보는 로봇, 데트는 빚입니다. 로봇이 만든 빚이라는 뜻입니다.

이 시스템이 저지른 잘못은 초등학교 산수 수준입니다.

정부는 국세청이 가진 연간 소득 자료와 복지 급여 기록을 맞춰 봤습니다. 문제는 두 자료의 시간 단위가 달랐다는 데 있습니다. 국세청은 일 년 치를 한 덩어리로 갖고 있었고, 복지 급여는 이 주에 한 번씩 나갔습니다.

시스템은 손쉬운 길을 골랐습니다. 일 년 소득을 오십이로 나눠서 매주 같은 돈을 벌었다고 가정했습니다.

정규직 회사원이라면 맞는 계산입니다. 그런데 복지 급여를 받는 사람들은 대체로 정규직이 아닙니다. 딸기 따는 철에 석 달 일하고 나머지는 쉬는 사람, 이번 달에 두 탕 뛰고 다음 달에 한 탕도 못 뛰는 사람, 계약이 끊길 때마다 급여 신청을 다시 하는 사람들입니다.

이런 사람의 일 년 소득을 오십이로 나누면 어떻게 될까요. 일하지 않은 주에도 돈을 번 것으로 계산됩니다. 그래서 급여를 부당하게 받았다는 결론이 나옵니다.

빚이 없는 사람에게 빚 독촉장이 갔습니다.

이의를 제기하려면 몇 년 전 급여 명세서를 가져와야 했습니다. 없으면 본인이 진 빚이 됐습니다. 정부가 계산을 증명하는 것이 아니라, 시민이 자기 결백을 증명해야 하는 구조였습니다. 추심 업체가 붙었습니다. 세금 환급금이 압류됐습니다.

몇몇 사람은 견디지 못했습니다. 억울함을 풀 곳을 찾지 못한 채 스스로 목숨을 끊은 사람들이 있었습니다.

2023년 왕립조사위원회 청문회에서 진짜 무서운 사실이 나왔습니다. 이 시스템이 불법이고 계산 방식에 결함이 있다는 경고가 여러 번 있었다는 겁니다. 법률 전문가들이 경고했고, 내부 직원들도 경고했습니다.

고위 공무원과 장관들은 알고 있었습니다. 그리고 계속 돌렸습니다.

예산을 아꼈다는 성과가 필요했기 때문입니다. 사람이 한 건 한 건 확인하면 시간이 걸리고 돈이 듭니다. 그 확인 절차를 없앤 것이 이 시스템의 핵심 기능이었습니다. 결말은 십팔억 호주 달러 규모의 합의와 정부의 공식 사과였습니다.

사과는 사망자에게 닿지 않았습니다.

인도에서는 손가락이 문제였습니다.

인도 정부는 아다르라는 생체 인증 시스템으로 복지를 관리합니다. 지문을 찍어야 식량 배급을 받습니다. 자르칸드주의 외딴 마을에서 이 방식이 사람을 죽였습니다.

지문 인식기가 자꾸 실패했습니다. 평생 밭일과 막노동을 한 사람의 손끝은 닳아 있습니다. 지문이 흐려집니다. 게다가 산간 마을에는 인터넷이 잘 들어오지 않습니다. 신호를 기다리다 하루가 갑니다.

파리하이야를 비롯한 소외 계층 주민들은 법으로 보장된 식량 배급을 거부당했습니다. 여러 마을에서 굶어 죽거나 영양실조로 죽은 사람이 나왔습니다.

기계가 사람을 못 알아본 것이 아닙니다. 기계가 못 알아보면 밥을 주지 말라고 정한 규칙이 사람을 죽였습니다.

하르야나주에서는 파리바르 페찬 파트라라는 시스템이 살아 있는 사람 수천 명을 사망자로 처리했습니다. 죽었다고 기록되면 연금이 끊기고 식량 배급이 끊깁니다. 나중에 조사해 보

니 지급이 중단된 연금의 70퍼센트가 알고리즘의 오판이었습니다.

살아 있다고 증명하러 관공서에 가야 했던 노인들이 있었습니다.

텔랑가나주의 삼마그라 베디카는 소득과 고용 상태를 잘못 예측해 만오천 가구의 식량 카드를 조용히 없앴습니다. 사람들이 들고일어난 뒤에야 되돌렸습니다.

요르단에서는 세계은행이 지원한 타카풀이라는 빈곤 측정 알고리즘이 돌아갔습니다. 가구의 특징 몇 가지에 점수를 매겨 지원 대상을 정했습니다. 그런데 진짜 굶는 집들이 점수 미달로 걸러졌습니다. 가난을 숫자로 재려던 자가 가난을 못 알아봤습니다.

브라질의 멥 INSS 앱은 연금 신청 대기 줄을 줄이려고 만들어졌습니다. 결과적으로 줄은 줄었습니다. 신청서가 자동으로 거절됐기 때문입니다. 양식에 오타가 있거나 질문을 잘못 이해하면 그대로 탈락이었습니다. 스마트폰에 익숙하지 않은 농촌 노인들이 걸렸습니다. 다시 신청하려면 변호사를 사야 했습니다.

부유한 나라라고 다르지 않았습니다.

영국의 유니버설 크레딧은 여러 복지 급여를 하나로 합친 제도입니다. 여기 들어간 계산 알고리즘은 달력 한 달 동안 들어온 돈을 기준으로 급여를 정했습니다. 휴먼라이츠워치가 2020년에 이 구조를 파헤쳤습니다.

일부 노동자는 사 주에 한 번 임금을 받습니다. 그러면 어떤 달에는 임금이 두 번 들어옵니다. 시스템은 이 사람이 갑자기 부자가 됐다고 판단했습니다. 그 달 급여가 확 깎이거나 아예 나오지 않았습니다.

달력이 몇 요일에서 시작하느냐에 따라 밥값이 사라졌습니다.

영국 노동연금부의 부정 수급 탐지 알고리즘은 더 조용히 움직였습니다. 어드밴시스라는 모형과 주거 지원금 감시 알고리즘이 수급자의 나이, 국적, 장애 여부, 혼인 상태를 보고 위험 점수를 매겼습니다.

높은 점수를 받은 사람들의 명단이 나왔습니다. 불가리아와 폴란드 국적의 저임금 단시간 노동 여성들, 그리고 장애인들이었습니다.

삼 년 동안 고위험군으로 분류돼 지원이 멈추거나 조사를 받은 사람이 이십만 명입니다. 그 중 삼분의 이 이상이 아무 잘못도 없는 정당한 수급자였습니다.

덴마크의 복지 기관 UDK는 육십 개가 넘는 부정 수급 탐지 알고리즘을 돌렸습니다. 거주 상태, 국적, 가족 관계, 의료 기록, 출국 이력까지 모았습니다. 어디에 갔다 왔는지, 누구와 사는지, 어떤 병이 있는지를 국가가 한자리에 모아 놓고 점수를 매긴 겁니다.

국제인권단체들은 이것이 유럽연합 인공지능법이 금지한 사회적 점수제에 해당한다고 봤습니다. 국제앰네스티는 코드로 짜인 불의라는 제목의 보고서를 냈습니다. 장애인, 저소득층, 이주민, 난민, 소수인종이 조사 대상으로 표시될 위험이 크다는 내용이었습니다.

앰네스티가 코드와 데이터를 보여 달라고 하자 덴마크 당국은 거절했습니다.

스웨덴 사회보험청은 2013년부터 사기 예측 모델을 돌렸습니다. 라이트하우스 리포트가 뜯어보니 여성, 이주민, 외국 배경을 가진 사람, 저소득자, 대학을 나오지 않은 사람이 더 자주 걸렸습니다.

프랑스 사회보장기구 CNAF의 위험 평가 알고리즘은 인구의 절반 가까이에 점수를 매겼습니다. 한부모 가정, 저소득층, 장애인이 자동으로 의심 대상이 됐습니다. 점수가 왜 그렇게 나왔는지는 알려 주지 않았습니다. 대신 집을 뒤지러 왔습니다.

네덜란드에서는 국세청이 2013년부터 2019년까지 자기학습 알고리즘으로 육아수당 부정 수급을 잡겠다고 나섰습니다. 수천 명의 부모가 사기꾼으로 몰렸습니다. 이중국적이 위험 신호로 쓰였습니다. 로테르담 시의 시스템은 민족, 나이, 성별, 아이가 있는지로 사람을 갈랐습니다.

아이를 지키겠다며 만든 알고리즘도 있었습니다.

미국 펜실베이니아주 알레게니 카운티는 아동 학대 위험을 예측하는 가족 스크리닝 도구를 썼습니다. 신고가 들어오면 이 도구가 점수를 매기고, 점수가 높으면 조사관이 집으로 갑니다.

카네기멜론 대학교 연구진이 결과를 분석했습니다. 신고된 흑인 아동의 삼분의 이가 조사 대상으로 추천됐습니다. 백인 아동은 절반이었습니다.

이유는 이 도구가 무엇을 보는지에 있었습니다. 산식에 정부 보조금을 받은 이력이 들어 있었습니다. 정신 건강 진단 기록도 들어 있었습니다.

가난해서 지원을 받으면 위험 점수가 올라갑니다. 아파서 상담을 받으면 위험 점수가 올라갑니다. 도움을 청한 기록이 의심의 근거가 되는 구조입니다.

도움을 받을수록 아이를 뺏길 확률이 올라가는 저울. 이런 저울을 만들어 놓고 공정하다고 부를 수는 없습니다.

이 도구를 본떠 만든 오리건주 시스템은 인종 편향이 드러나 2022년에 폐기됐습니다. 일리노이주가 쓰던 에커드 라피드 세이프티 피드백도 데이터 오류와 부풀려진 성능 때문에 중단됐습니다.

덴마크 아동보호국의 의사결정 지원 시스템은 코펜하겐 IT 대학교 검증에서 이상한 습관이 발견됐습니다. 똑같은 상황인데 나이가 다르면 위험 점수가 달라졌습니다.

호주 정부는 장애인 지원 예산 칠억 호주 달러를 아끼겠다며 로보 플래닝이라는 자동 평가 도구를 도입하려 했습니다. 장애인의 삶을 표준 유형 몇 개로 압축한 뒤 필요한 지원을 계산하는 방식이었습니다. 시각장애인 단체와 장애인 단체들이 거세게 반대했고 계획은 접혔습니다.

병원에서는 알고리즘이 퇴원 날짜를 정했습니다.

미국 최대 건강보험사 유나이티드헬스 그룹과 휴매나가 쓴 nH Predict은 노인과 장애인 환자가 재활 치료를 며칠 받아야 하는지 예측했습니다. 의사가 더 필요하다고 해도 시스템이 정한 날이 되면 보험 보장이 끊겼습니다.

소송에서 이 시스템의 청구 거절 오류율이 나왔습니다. 90퍼센트입니다.

열 번 거절하면 아홉 번이 틀린 셈입니다. 그런데도 계속 쓴 이유는 간단합니다. 거절당한 환자 대부분이 이의를 제기하지 않았기 때문입니다. 늙고 아픈 사람이 보험사와 싸우기는 어렵습니다.

2021년 11월, 여든여섯 살 조앤 배로스는 넘어져 다리를 다쳤습니다. 의료진은 물리 치료가 꼭 필요하다고 했습니다. 알고리즘이 정한 퇴원 예정일이 오자 치료비 보장이 끊겼습니다.

유나이티드헬스의 알러트 시스템은 정신 건강 치료 승인을 불법으로 제한하다가 여러 주 정부에서 사용 금지를 당했습니다. 시그나의 PxDx는 의사가 환자 차트를 열어 보지 않고도 클릭 한 번으로 청구를 무더기 거절할 수 있게 만든 시스템이었습니다. 이비코어의 디알 알고리즘은 사전 승인 거절률을 최대 20퍼센트까지 끌어올렸습니다.

딜로이트가 만든 텍사스의 TIERS와 테네시의 TEDS는 프로그램 오류와 주소 오발송으로 수십만 명의 저소득층과 장애인의 메디케이드 자격을 날렸습니다. 통지서가 엉뚱한 집으로 갔고, 답장이 없으니 자격이 없다고 처리됐습니다.

여기까지 나온 나라를 세어 보면 호주, 인도, 요르단, 브라질, 영국, 덴마크, 스웨덴, 프랑스, 네덜란드, 미국입니다. 제도도 다르고 언어도 다릅니다.

그런데 걸린 사람들은 놀랄 만큼 비슷합니다. 가난하고, 나이가 많고, 아프고, 이주민이고, 말할 곳이 없는 사람들입니다.

우연이 아닙니다. 이 시스템들은 예산을 아끼려고 도입됐고, 예산을 아끼려면 지급을 줄여야 하고, 지급을 줄이려면 누군가를 잘라 내야 합니다. 잘라 내기 제일 안전한 대상은 항의하지 못하는 사람입니다.

알고리즘은 그 계산을 아주 잘합니다.

1.3

안면 인식 기술의 유색인종 오인과 부당 체포

두 살, 다섯 살짜리 딸이 마당에서 보고 있었습니다.

2020년 1월 9일, 미시건주 파밍턴힐스의 한 집 앞마당에서 로버트 윌리엄스는 퇴근하고 막 돌아온 참이었습니다. 디트로이트 경찰이 들이닥쳐 아내와 아이들 앞에서 그의 손목에 수갑을 채웠습니다.

혐의는 시계 절도였습니다. 2018년 10월 디트로이트의 시놀라 매장에서 천팔백 달러어치 시계 다섯 개가 없어진 사건입니다.

증거는 이랬습니다. 매장 감시 카메라에 찍힌 흐릿한 영상이 있었습니다. 미시건 주경찰 감정인이 그 영상에서 한 장면을 캡처해 안면 인식 데이터베이스에 넣었습니다. 시스템은 윌리엄스의 오래된 운전면허증 사진을 후보로 내놓았습니다.

그다음이 중요합니다. 사건 현장에 있지도 않았던 매장 보안요원이 흑인 남성들의 사진 여러 장을 받아 보고 윌리엄스를 골랐습니다.

윌리엄스는 삼십 시간을 유치장에 있었습니다.

취조실에서 형사가 감시 카메라 사진을 책상에 놓고 물었습니다. 이 사람이 당신이냐. 윌리엄스는 사진을 자기 얼굴 옆에 대고 말했습니다. 이 사람은 내가 아닙니다. 당신들은 흑인이 다 똑같이 생겼다고 생각하는 겁니까.

이 사건에는 뒷이야기가 있습니다. 당시 디트로이트 경찰청장 제임스 크레이그는 이 기술이 용의자를 잘못 짚는 비율이 96퍼센트에 이른다는 사실을 이미 알고 있었습니다.

백 번 중 아흔여섯 번 틀리는 도구를 들고 사람 집 앞마당에 간 겁니다.

로버트 윌리엄스는 미국에서 안면 인식 오류로 부당 체포된 첫 공식 기록으로 남았습니다. 2024년 6월 디트로이트시는 미국시민자유연맹과 삼십만 달러에 합의했고, 안면 인식 단서와 사진 대조만으로는 체포영장을 신청할 수 없게 규정을 바꿨습니다.

2026년 기준, 이런 식으로 체포된 사람이 열두 명 넘게 알려져 있습니다. 알려진 숫자만 그렇습니다.

2026년 6월 10일에도 새 소송이 접수됐습니다. 플로리다의 로버트 딜런은 2024년에 체포됐습니다. 흐릿한 감시 영상과 93퍼센트 일치라는 결과가 근거였습니다. 영장 신청서에는 그가 범인일 수 없다는 증거가 빠져 있었습니다.

93퍼센트라는 숫자가 사람을 홀립니다. 시험 점수 같아서요. 하지만 이 숫자는 이 사람이 범인일 확률이 93퍼센트라는 뜻이 아닙니다. 데이터베이스에 있는 수백만 장 중에서 이 사진이 제일 비슷하게 생겼다는 뜻입니다.

제일 비슷하게 생긴 사람은 언제나 존재합니다. 범인이 그 안에 없어도 말입니다.

이 차이를 모르면 어떤 일이 벌어지는지, 앤절라 립스의 이야기가 보여 줍니다.

2025년 7월 14일, 테네시주. 다섯 아이의 어머니이자 다섯 손주의 할머니인 립스는 집에서 손주들을 보고 있었습니다. 연방 보안관들이 총을 겨누고 들어왔습니다. 노스다코타주 파고에서 일어난 은행 사기 사건의 도망자라는 겁니다.

립스는 노스다코타에 가 본 적이 없습니다. 비행기를 타 본 적도 없습니다. 집에서 백육십 킬로미터 밖으로 나가 본 일조차 거의 없습니다.

파고 경찰서 형사들이 한 수사는 이게 전부였습니다. 안면 인식 결과를 받고, 소셜미디어에서 사진을 찾아 눈으로 비교했습니다. 전화 한 통 걸지 않았습니다.

립스는 노스다코타로 이송돼 백여덟 일을 갇혀 있었습니다. 변호인이 은행 거래 기록을 제출해 그 시각 테네시에 있었음을 증명한 뒤에야 풀려났습니다.

그사이 살던 집을 잃었습니다. 자동차를 잃었습니다. 기르던 개도 잃었습니다.

디트로이트 경찰의 같은 시스템은 계속 사람을 물었습니다.

2019년 7월 마이클 올리버는 교사의 휴대폰을 훔쳤다는 혐의로 체포됐습니다. 데이터웍스 플러스의 프로그램이 경찰 데이터베이스에 있던 그의 사진을 범인과 연결했습니다.

2023년 2월 포샤 우드러프는 강도와 차량 강탈 혐의로 체포됐습니다. 그는 임신 팔 개월이었습니다. 유치장 바닥에서 진통을 느꼈습니다.

수사관은 감시 영상을 알고리즘에 돌려 팔 년 전 우드러프의 운전면허증 사진을 얻었습니다. 수배 대상의 신체 특징을 확인하지 않았습니다. 감시 카메라에 찍힌 사람은 임신부가 아니었습니다.

배가 남산만 한 사람과 그렇지 않은 사람을 구별하는 데는 인공지능이 필요하지 않습니다. 눈이 필요합니다. 그런데 눈으로 보는 절차가 빠져 있었습니다.

2025년 4월 뉴욕에서 트레비스 윌리엄스가 공연음란죄 용의자로 체포돼 이틀 넘게 갇혔습니다. 그때 그는 십구 킬로미터 떨어진 브루클린에 있었습니다.

신체 조건을 비교해 보면 웃음이 나옵니다. 실제 범인은 키 백육십칠 센티미터에 몸무게 칠십이 킬로그램이었습니다. 윌리엄스는 백팔십팔 센티미터에 백사 킬로그램입니다.

경찰은 위치 데이터 확인도, 고용주 확인도 거부했습니다.

2022년 11월 조지아 주민 랜달 큐란 리드는 루이지애나에서 명품 가방을 훔친 범인으로 몰려 엿새를 갇혔습니다. 클리어뷰 AI의 일치 결과 하나로 영장이 나왔습니다. 리드가 그 시각 조지아에 있었다고 말했지만 아무도 확인하지 않았습니다. 나중에 사건은 조용히 기각됐습니다.

왜 하필 이런 얼굴들만 틀릴까요. 답은 2018년에 이미 나와 있었습니다.

MIT와 스탠퍼드의 조이 부올람위니와 딤니트 게브루가 젠더 셰이즈라는 보고서를 냈습니다. IBM, 페이스플러스플러스, 마이크로소프트의 얼굴 분석 프로그램을 시험한 결과입니다.

피부색이 밝은 남성의 성별을 틀린 비율은 0.8퍼센트 미만이었습니다. 피부색이 어두운 여성은 34.7퍼센트까지 틀렸습니다.

셋에 하나꼴입니다.

원인은 학습 데이터에 있습니다. 이미지넷, LFW, BDD100K 같은 유명 데이터셋은 백인 남성 사진이 압도적으로 많았습니다. 기계는 자기가 많이 본 얼굴을 잘 구별합니다. 적게 본 얼굴은 뭉뚱그립니다.

조지아 공과대학교가 자율주행용 데이터셋 BDD100K를 분석했을 때도 같은 무늬가 나왔습니다. 피부색이 어두운 보행자를 알아보는 정확도가 평균 4.8퍼센트, 최대 12퍼센트 낮았

습니다.

차가 사람을 못 알아본다는 이야기입니다. 이 대목은 뒤에서 다시 나옵니다.

경고는 충분했습니다. 판매는 계속됐습니다.

클리어뷰 AI는 페이스북, 인스타그램, 트위터, 링크드인에서 사람들의 얼굴 사진 삼십억 장을 동의 없이 긁어모아 검색 데이터베이스를 만들었습니다. 플랫폼 기업들이 중단을 요구했고 일리노이주 생체정보보호법 위반 소송이 걸렸지만, 회사는 전 세계 수사기관에 검색 서비스를 팔았습니다.

우리가 십 년 전 올린 졸업 사진이 지금 어느 경찰서 컴퓨터 안에 있을지 모릅니다.

아마존도 레코그니션이라는 제품을 경찰과 이민통관집행국에 팔려고 했습니다. 2018년 7월 미국시민자유연맹이 시험해 보니 연방 하원의원과 상원의원 스물여덟 명의 얼굴이 범죄자 사진과 일치한다고 나왔습니다. 오인된 의원 중 유색인종 비율이 유독 높았습니다.

2019년 10월에는 뉴잉글랜드 지역 프로 스포츠 선수 백여든여덟 명 중 스물일곱 명이 범죄자로 나왔습니다.

아마존은 신뢰도 기준을 95퍼센트 이상으로 놓고 쓰라고 안내했다고 했습니다. 현장에서는 그 기준이 지켜지지 않았습니다. 설명서를 읽는 사람은 언제나 소수입니다.

과장 광고도 있었습니다. 연방거래위원회는 인텔리비전이라는 회사가 자사 안면 인식 소프트웨어에 성별·인종 편향이 없고 세계 최고 정확도라고 광고한 것을 문제 삼았습니다. 수백만 장으로 학습했다고 했지만 실제로는 변형한 십만 장짜리 데이터셋이었습니다.

유통업체 라이트에이드는 2012년부터 2020년까지 저소득층과 유색인종이 많이 사는 동네 매장에 안면 인식 카메라를 집중 설치했습니다. 수천 명의 손님이 절도범으로 몰려 매장에서 쫓겨나거나 몸수색을 당했습니다. 연방거래위원회는 오 년 동안 이 기술을 쓰지 못하게 했습니다.

여기까지는 그래도 살아 돌아온 사람들의 이야기입니다.

2022년 1월 텍사스 휴스턴의 선글래스 헛 강도 사건에서, 메이시스와 에실로룩소티카의 안면 인식 시스템이 하비 머피 주니어를 범인으로 지목했습니다. 그는 그때 캘리포니아 새크

라멘토에 있었습니다.

2023년 10월 체포돼 이 주 동안 유치장에 있는 사이, 머피는 수감자 세 명에게 집단 폭행과 성폭행을 당했습니다. 영구적인 신체 장애가 남았습니다.

미주리주 폰드데일에서는 2021년 열차 승강장 폭행 사건이 있었습니다. 용의자는 마스크와 옷으로 얼굴을 가린 상태였습니다. 형사는 그 사진을 알고리즘에 돌렸습니다.

가려진 얼굴을 넣어도 시스템은 후보를 내놓습니다. 모르겠다고 대답하는 기능이 없기 때문입니다. 그렇게 나온 후보 중에 전과 없는 네 아이의 아버지, 스물아홉 살 크리스토퍼 개틀린이 있었습니다.

경찰 내부 지침에도 안면 인식 결과만으로 체포하면 안 된다고 적혀 있었습니다. 형사는 사진 라인업을 만들었고, 개틀린은 이 년 넘게 감옥에 있었습니다.

오클라호마의 킴벌리 윌리엄스는 2021년 6월부터 여섯 달을 갇혔습니다. 사설 은행 조사관이 올린 검색 결과를 형사들이 확인 없이 영장 신청서에 옮겨 적었습니다. 법원에는 안면 인식을 썼다는 사실조차 알리지 않았습니다. 그는 카운티 감옥 세 곳을 옮겨 다니는 사이 직장을 잃었습니다.

뉴저지의 프란시스코 아르테아가는 2019년 11월 무장강도 혐의로 잡힌 뒤 사 년을 갇힌 채 재판을 기다렸습니다. 수사기관이 시스템의 작동 원리와 오류율, 소스 코드 공개를 거부했기 때문입니다. 무엇으로 자기를 지목했는지 모르는 상태에서는 반박할 방법이 없습니다.

메릴랜드 볼티모어의 쉰네 살 알론조 소여는 버스 안 폭행 사건 용의자로 아흐레 갇혔습니다. 실제 용의자와 키가 달랐고, 나이가 달랐고, 앞니 사이 틈이 달랐고, 걸음걸이가 달랐습니다.

국경 밖에서도 같은 일이 벌어졌습니다.

2026년 1월 영국 사우샘프턴에서 방글라데시계 소프트웨어 엔지니어 알비 차우두리가 부모와 이웃이 보는 앞에서 체포됐습니다. 백팔십오 킬로미터 떨어진 밀턴킨스의 빈집털이 사건 용의자로 지목된 겁니다. 열 시간 구금됐습니다.

템스밸리 경찰이 쓴 시스템은 독일 코그니텍 제품이었고, 2020년에 나온 구형 알고리즘이 들어 있었습니다. 영국 국립물리연구소 평가에 따르면 특정 설정에서 아시아계 오인율은

4.0퍼센트였습니다. 백인은 0.04퍼센트였습니다.

백 배입니다.

경찰관들은 사람 눈으로 다시 확인했다고 주장했습니다. 여기에 자동화 편향이라는 이름이 붙습니다. 기계가 답을 내놓으면 사람은 그 답을 확인하는 대신 그 답에 맞춰 봅니다. 닮은 점을 찾기 시작하는 겁니다. 사람 얼굴에서 닮은 점을 찾는 일은 어렵지 않습니다.

2024년 5월 런던 브리지 역에서 서른여덟 살 숀 톰슨이 붙잡혔습니다. 그는 칼 범죄 예방 봉사활동을 마치고 집에 가던 청소년 옹호 활동가였습니다. 메트로폴리탄 경찰의 실시간 안면 인식 시스템이 그를 수배범으로 지목했고, 삼십 분 동안 신분증을 내라는 요구와 체포 위협을 받았습니다.

영국 내무부의 여권 사진 확인 시스템은 피부색이 어두운 신청자의 입술을 입을 벌린 상태로 잘못 읽었습니다. 조리스 레셴과 조슈아 바다 같은 흑인 신청자들의 여권 신청이 반려됐습니다. 내무부는 유색인종에 대한 실패율이 높다는 사실을 알고도 시스템을 도입했습니다.

이스라엘 국방군은 가자 지구와 서안 지구 검문소에 레드 울프, 블루 울프, 콜사이트라는 시스템을 배치했습니다. 팔레스타인 주민의 얼굴을 동의 없이 모아 자동 감시망을 만들었습니다. 오류로 무고한 주민들이 테러 용의자로 지목돼 끌려가고 심문받고 맞았습니다.

인도 하이데라바드에서는 2023년 2월 서른다섯 살 일용직 노동자 모하메드 카디르가 연행됐습니다. 소매치기 사건 영상 속 얼굴과 비슷하다는 이유였습니다. 그 영상 속 인물은 수건으로 얼굴을 가리고 있었습니다.

카디르는 닷새 동안 갇혀 고문을 받다 숨졌습니다.

브라질에서는 2024년 4월 축구 경기를 보러 간 주앙 안토니오가 수배범과 얼굴이 일치한다는 이유로 수갑을 찼습니다. 아르헨티나 부에노스아이레스의 수배자 인식 시스템은 백사십 건이 넘는 오인 체포를 낸 뒤 운용이 중지됐습니다.

이 사건들을 나란히 놓으면 하나의 순서가 보입니다.

기술 회사는 정확도를 부풀려 팝니다. 경찰은 그 결과를 물증처럼 씁니다. 영장 신청서에는 안면 인식을 썼다는 말이 빠집니다. 판사는 모르고 서명합니다. 붙잡힌 사람은 자기가 왜 지목됐는지 모릅니다. 원리를 물으면 영업 비밀이라고 합니다.

증명 책임이 뒤집혔습니다. 국가가 유죄를 증명하는 대신, 시민이 무죄를 증명해야 합니다. 그러려면 알리바이가 있어야 하고 변호사비가 있어야 합니다.

앤절라 립스에게는 백팔 일이 걸렸습니다. 크리스토퍼 개틀린에게는 이 년이 걸렸습니다. 프란시스코 아르테아가에게는 사 년이 걸렸습니다.

모하메드 카디르에게는 시간이 남지 않았습니다.

제 2 장

환각과 허위 정보: 생성형 AI의 신뢰성 위기

2.1

생성형 AI의 환각과 가짜 도서·기사 유포

작가라면 누구나 가끔 자기 이름을 검색해 봅니다. 부끄러운 습관이지만 다들 합니다.

2023년 8월 어느 아침, 미국의 작가이자 도서 비평가인 제인 프리드먼도 그렇게 했습니다. 화면에 자기 책 목록이 떴습니다. 그런데 다섯 권이 낯설었습니다.

읽어 본 적 없는 책이었습니다. 쓴 적도 없는 책이었습니다.

아마존과 굿리즈 검색 상단에 프리드먼의 이름을 달고 올라와 있던 그 책들은 생성형 인공지능이 찍어 낸 물건이었습니다. 펼쳐 보면 문장이 어딘가 미끄럽고 알맹이가 없습니다. 대형 언어 모델이 쓴 글의 특징입니다. 틀린 말은 별로 없는데 남는 것도 별로 없습니다.

평생 쌓은 이름값이 하룻밤 사이에 저질 텍스트의 등에 업혔습니다.

프리드먼은 아마존에 삭제를 요청했습니다. 답장이 왔습니다. 귀하는 본인 이름을 상표로 등록하지 않았습니다.

자기 이름을 쓰려면 먼저 상표 등록을 했어야 한다는 이야기입니다. 세상에서 손꼽히게 관료적인 문장입니다.

프리드먼이 블로그에 이 일을 올리고 사람들이 화를 낸 뒤에야 아마존은 책들을 조용히 내렸습니다. 규칙이 바뀐 것이 아니라 시끄러워졌기 때문에 내린 겁니다.

가짜 책이 이름만 훔치는 데서 끝났다면 그나마 다행이었을 텐데, 그러지 않았습니다.

2023년 9월 뉴욕 버섯학회가 긴급 경고를 냈습니다. 아마존에서 팔리는 초보자용 야생 버섯 채집 안내서 여러 권이 챗GPT가 쓴 책이었다는 내용이었습니다.

버섯 책의 핵심은 하나입니다. 먹어도 되는 것과 먹으면 죽는 것을 구별하는 법. 그 부분이 틀려 있었습니다.

오탈자가 아닙니다. 사람이 죽는 정보입니다.

2024년 2월에는 찰스 3세 국왕의 암 진단 소식이 나오자마자 아마존에 가짜 전기 일곱 권이 올라왔습니다. 국왕이 피부암에 걸렸다거나 감춰진 사고를 당했다는 식의 내용이 실제 뉴스처럼 섞여 있었습니다. 버킹엄 궁전이 격노했습니다.

빠르게 쓰고, 빠르게 올리고, 검색 상단을 차지해 돈을 법니다. 인공지능이 등장하기 전에는 사람이 며칠 걸려 하던 일입니다. 이제 몇 분입니다.

가끔은 이 자동화가 스스로 정체를 드러내기도 합니다.

2024년 1월 아마존에 이상한 상품들이 올라왔습니다. 정원용 의자, 호스, 탁자였습니다. 상품 이름 자리에 이런 문장이 적혀 있었습니다.

죄송합니다. 해당 요청은 오픈AI 이용약관 정책에 위배되어 수행할 수 없습니다.

판매자가 상품 설명을 인공지능에게 시켰고, 인공지능이 거절했고, 그 거절 문장을 그대로 복사해 올린 겁니다. 읽어 보지도 않고 말입니다.

웃긴 장면입니다. 그런데 이걸 뒤집어 보면 무섭습니다. 이 판매자는 자기가 파는 물건의 설명을 읽지 않았습니다. 읽지 않은 설명이 몇 개나 인터넷에 올라가 있을까요.

2024년 4월에는 구글 북스가 인공지능이 쓴 저품질 책들을 도서 검색 데이터베이스에 그대로 수집해 넣었습니다. 그중 여러 권에 이런 문장이 남아 있었습니다. 제 최신 지식 업데이트에 따르면.

구글 엔그램 뷰어라는 도구가 있습니다. 어떤 낱말이 어느 시대에 얼마나 쓰였는지 책 데이터로 추적하는 학술 도구입니다. 언어의 역사를 재는 자입니다. 그 자에 인공지능이 쓴 책이 섞이면 눈금이 흐려집니다.

여기까지는 출판 쪽 이야기입니다. 신문사는 더 크게 넘어졌습니다.

2025년 5월 시카고 선타임스가 여름 추천 도서 목록이라는 특별 부록을 냈습니다. 이사벨 아옌데, 앤디 위어, 이민진 같은 유명 작가들의 신작이 화려한 추천평과 함께 실렸습니다.

독자들이 서점에 갔습니다. 책이 없었습니다.

그 목록에 있던 책들은 세상에 한 권도 존재하지 않았습니다. 작가는 실존 인물인데 책 제목만 인공지능이 지어낸 겁니다. 실제 이름에 가짜 정보를 붙이는 방식은 인공지능이 제일 잘

하는 거짓말 형태입니다. 확인하려고 검색하면 작가 이름이 진짜라서 안심하게 됩니다.

신문사는 사람을 줄이고 외부 업체에서 콘텐츠를 받아 쓰고 있었습니다. 프리랜서 작가는 인공지능이 뱉은 목록을 그대로 옮겼고, 편집진은 한 번도 확인하지 않았습니다.

큰 신문사의 이름이 인공지능의 거짓말에 신뢰라는 도장을 찍어 준 순간이었습니다.

2023년 11월에는 스포츠 일러스트레이티드가 걸렸습니다. 탐사보도 매체 퓨처리즘이 이 매체가 인공지능이 쓴 기사를 계속 실어 왔다고 폭로했습니다.

기사에 적힌 기자 이름과 얼굴 사진, 경력이 전부 가짜였습니다. 얼굴 사진은 합성 알고리즘으로 만든 것이었습니다. 존재하지 않는 기자가 존재하지 않는 경력으로 제품 리뷰를 썼습니다.

매체는 외주 업체 탓을 했습니다. 사람 기자를 줄이고 자동화를 들인 결정은 경영진이 했습니다.

지역 신문에서는 기자가 직접 손을 댔습니다.

2024년 8월 와이오밍주 코디 엔터프라이즈의 신입 기자 아론 펠차르가 사직했습니다. 경쟁지 기자가 그의 기사에서 이상하게 매끄러운 문장을 발견하고 취재원에게 직접 전화를 걸면서 들통났습니다.

펠차르는 주지사를 포함한 공직자 여섯 명의 발언을 챗GPT로 지어내 기사에 실었습니다. 하지 않은 말을 한 것으로 만든 겁니다.

압권은 따로 있습니다. 그가 쓴 기사 끝에 역삼각형 취재 기법을 설명하는 인공지능의 안내 문구가 지워지지 않은 채 남아 있었습니다. 컨닝 페이퍼를 시험지에 붙여서 낸 셈입니다.

2023년 5월에는 아이리시 타임스가 인공 태닝이 인종차별적이라는 취지의 기고문을 실었다가 기사 전체가 인공지능으로 조작된 것임이 드러나 철회하고 사과했습니다.

이제 진짜 서늘한 대목으로 갑니다. 인공지능은 없는 사건을 만들어 냅니다.

2023년 12월 뉴스 앱 뉴스브레이크가 성탄절 당일 뉴저지주 브리지턴에서 한 여성이 총에 맞아 숨졌다는 기사를 내보냈습니다. 지역 주민들이 놀랐고 경찰이 확인했습니다.

그런 사건은 없었습니다.

뉴스브레이크가 쓰던 뉴스 재작성 엔진이 인터넷에 떠도는 조각난 정보를 버무리다가 살인 사건 하나를 만들어 낸 겁니다.

2024년 10월에는 지역 뉴스 네트워크 후드라인이 캘리포니아 산마테오 카운티 지방검사 존 카시아노 톰슨이 살인 혐의로 기소됐다는 기사를 내보냈습니다. 범죄를 기소하는 사람이 살인범이 됐습니다.

개인에게도 같은 일이 일어납니다.

2023년 6월 라디오 진행자 마크 월터스는 챗GPT가 자신을 오백만 달러 넘게 횡령한 사기꾼으로 묘사한 가짜 재판 요약문을 만들어 냈다는 사실을 알고 오픈AI에 명예훼손 소송을 냈습니다.

경위가 이렇습니다. 언론인 프레드 리엘이 총기 권리 단체의 실제 소송 내용을 요약해 달라고 했습니다. 챗GPT는 그 소송과 아무 관계 없는 월터스를 등장시켜 횡령범으로 만들었습니다. 그럴듯한 사건 번호와 문장을 붙여서요.

지도 서비스 회사 오픈케이지는 챗GPT가 자기네가 전화번호 조회 서비스를 한다고 거짓말하는 바람에 애먼 항의 전화를 받아야 했습니다.

2025년 3월 노르웨이의 아르베 야르마르 홀멘은 챗GPT에 자기 이름을 물었습니다. 돌아온 답은 이랬습니다. 그는 두 자녀를 살해하고 세 번째 자녀를 살해하려다 이십일 년형을 선고받았습니다.

이 문장에서 고향과 자녀 수는 진짜였습니다. 나머지가 가짜였습니다.

진짜와 가짜를 섞는 이 방식이 확률 기계의 습관입니다. 사실을 저장했다가 꺼내 오는 것이 아니라, 앞말에 이어 나올 법한 말을 고르기 때문입니다. 노르웨이 사람 이름 뒤에 그럴듯하게 이어질 문장을 고르다 보면 사건 사고 기사가 나옵니다. 기계는 그것이 사람의 인생인 줄 모릅니다.

정치가 얽히면 이 습관은 무기가 됩니다.

2023년 11월 파키스탄의 뉴스 사이트 글로벌 빌리지 스페이스가 이스라엘 총리 베냐민 네타냐후의 전담 정신과의사 모셰 야톰이 총리의 잔혹성을 비판하는 유서를 남기고 자살했다는 기사를 실었습니다.

모셰 야톰이라는 사람은 없습니다. 유서도 없습니다. 전부 지어낸 이야기입니다. 이 기사는 이스라엘과 하마스의 전쟁이 격해지던 시점에 여러 언어로 번역돼 퍼졌습니다.

2024년 4월 엑스에 붙어 있는 그록은 인터넷에 떠도는 소문을 모아 이란이 텔아비브에 미사일을 쐈다는 헤드라인을 만들었습니다. 농구 선수 클레이 톰슨이 도시에서 벽돌을 던지고 다녔다는 기사도 만들었습니다. 후자는 그가 슛을 많이 놓쳤다는 뜻의 농구 속어를 문자 그대로 읽은 결과로 보입니다.

농담을 못 알아듣는 기계가 뉴스를 씁니다.

언론사 이름을 도용하는 방식은 더 교묘합니다.

영국 가디언 편집진은 독자들에게서 이상한 문의를 계속 받았습니다. 이런 기사가 있다는데 어디서 볼 수 있느냐는 겁니다. 가디언은 그런 기사를 쓴 적이 없었습니다.

챗GPT에 자료를 요청하면 가디언에 실제로 있는 기자 이름과 그럴듯한 제목과 발행 연도를 조합해 내놓았습니다. 실존하는 매체와 실존하는 기자를 빌려 거짓말에 권위를 입힌 겁니다. 가디언의 혁신 담당 편집자 크리스 모란이 이 현상을 공개적으로 고발했습니다.

2024년 7월 니먼 저널리즘 랩이 이 문제를 제대로 실험했습니다. 챗GPT에게 AP통신, 월스트리트저널, 파이낸셜타임스, 더 타임스, 보스턴 글로브, 폴리티코의 탐사보도 기사 링크를 달라고 했습니다. 오픈AI와 정식으로 데이터 계약을 맺은 언론사들입니다.

챗GPT는 상세한 요약과 함께 진짜처럼 생긴 주소를 내놓았습니다. 전부 존재하지 않는 주소였습니다.

학계도 비슷합니다. 헨리크 엥호프 교수의 존재하지 않는 논문이 인용됐고, 익사로 인한 홀로코스트라는 있지도 않은 역사적 사건이 만들어졌습니다.

검색 창에 붙은 인공지능은 이 문제를 안방까지 들여왔습니다.

구글 AI 오버뷰는 피자 치즈가 흘러내리지 않게 하려면 접착제를 조금 섞으라고 답한 적이 있습니다. 인터넷 게시판의 농담을 진지하게 읽은 결과입니다. 독버섯 사진을 식용 버섯 추천 상단에 올린 적도 있습니다.

한 학술 연구는 금융 관련 검색어의 43퍼센트에서 심각한 오류가 나왔다고 보고했습니다. 구글은 서비스를 계속했습니다. 검색 시장을 내줄 수 없었기 때문입니다.

한국에서도 일이 있었습니다. 네이버의 생성형 인공지능이 독도 관련 질문에 일본 외무성 주장을 바탕으로 분쟁 지역이자 일본 영토라고 요약해 내보냈습니다.

가짜 정보는 화면 밖으로 걸어 나오기도 합니다.

2024년 10월 아일랜드 더블린 시내에 사람들이 몰렸습니다. 챗GPT가 만들어 낸 존재하지 않는 할로윈 퍼레이드 정보가 검색 상단에 올라간 탓입니다. 시민과 관광객 수천 명이 비 오는 거리에 서서 오지 않는 행렬을 기다렸습니다.

2025년 11월 런던에서는 버킹엄 궁전 앞에 화려한 왕실 크리스마스 마켓이 열린다는 인공지능 사진과 기사가 소셜미디어를 탔습니다. 세계 각지에서 온 관광객들이 닫힌 철문과 빗물 웅덩이 앞에 섰습니다.

2026년 1월 호주 태즈메이니아에서는 인공지능이 쓴 기사를 믿고 존재하지 않는 온천을 찾아 오지를 헤맨 관광객들이 나왔습니다.

이 모든 사건의 뿌리는 기술의 성질에 있습니다.

대형 언어 모델은 사실을 아는 기계가 아닙니다. 앞말 다음에 올 확률이 높은 말을 고르는 기계입니다. 무엇이 참인지 판단하는 장치가 안에 없습니다. 그런 장치가 없다는 것이 결함이 아니라 설계입니다.

그러면 모델을 더 크게 만들면 나아질까요. 2024년에 나온 연구들은 반대 결과를 내놓았습니다. 모델이 커질수록 모른다고 말하는 대신 더 자연스럽고 당당하게 틀린 답을 내놓는 경향이 있었습니다.

똑똑해지는 것이 아니라 말솜씨가 늘어난 겁니다.

편집자와 사실 확인 담당자는 원래 이 문제를 막으라고 있던 사람들입니다. 그런데 그 자리가 비용으로 분류돼 먼저 사라졌습니다.

문제가 터지면 회사들은 같은 대답을 합니다. 아직 베타 서비스입니다. 화면 아래에 부정확할 수 있다고 적어 두었습니다.

그 작은 글씨는 버섯 책을 산 사람에게 도착하지 않습니다.

2.2

법정·학술계의 가짜 판례 및 인용문 파문

시작은 무릎이었습니다.

로베르토 마타는 아비앙카 항공 비행기 안에서 기내 카트에 무릎을 부딪혔습니다. 그는 항공사를 상대로 손해배상 소송을 냈습니다. 흔한 사건입니다. 이런 소송은 판례집에 남지 않습니다.

2023년 5월, 이 사건은 전 세계 법조계가 아는 사건이 됐습니다. 무릎 때문이 아니었습니다.

마타의 변호인 스티븐 슈워츠는 항공사의 각하 신청에 맞서 판례 여섯 건을 냈습니다. 바르헤스 대 차이나 사우던 항공 사건 같은 이름들이었습니다. 사건 번호가 있었고, 판결 일자가 있었고, 법원 이름이 있었습니다. 형식은 완벽했습니다.

상대 변호인단이 판례를 찾아봤습니다. 없었습니다. 판사 보조관이 찾아봤습니다. 없었습니다.

여섯 건 전부 존재하지 않는 판결이었습니다.

슈워츠는 서류 준비에 챗GPT를 썼습니다. 심문에서 그는 인공지능이 거짓말을 할 수 있다는 사실을 몰랐다고 했습니다. 그는 챗GPT에게 이 판례가 진짜냐고 물어보기까지 했습니다. 챗GPT는 진짜라고 답했습니다.

거짓말쟁이에게 거짓말했느냐고 물어본 셈입니다.

동료 변호사 피터 로두카는 확인하지 않고 서명했습니다. 법원은 두 사람에게 오천 달러의 벌금을 물렸습니다.

이 사건이 유명해진 이유는 벌금 액수가 아닙니다. 법정에 제출되는 서류가 어떻게 만들어지는지를 사람들이 처음으로 들여다봤기 때문입니다.

그리고 이 일은 반복됐습니다. 아주 여러 번 반복됐습니다.

프랑스 연구자 다미앵 샤를로탱은 인공지능 환각 판례를 모으는 데이터베이스를 만들었습니다. 숫자가 어떻게 늘었는지 보겠습니다.

2025년 중반에 약 200건이었습니다. 2026년 1월에 719건이 됐습니다. 2026년 4월 초에 1,227건, 6월 9일에 1,598건입니다.

5월 22일부터 6월 9일까지 열여덟 날 동안 140건이 늘었습니다. 하루 여덟 건꼴입니다.

법정에서 가짜 판례가 발각되는 일이 하루 여덟 번 일어난다는 뜻입니다. 발각되지 않은 것은 세지 않았습니다.

이 목록에 오른 사건들을 보면 국적과 직급이 다양합니다.

도널드 트럼프 전 대통령의 측근이었던 마이클 코언은 감독 기간을 앞당겨 끝내 달라는 서류를 준비하며 구글 바드를 썼습니다. 인용문이 진짜인 줄 알고 자기 변호사에게 넘겼고, 변호사는 확인 없이 법원에 냈습니다. 판사가 판례집에 없다고 알려 온 뒤에야 알았습니다.

뉴욕의 재이 리 변호사는 퀸즈의 의사가 낙태 수술을 잘못했다는 가짜 판례를 챗GPT로 만들어 제2연방항소법원에 냈다가 징계 절차에 넘겨졌습니다.

캐나다 브리티시컬럼비아주 대법원에서는 이혼 소송 중 아버지가 아이들을 중국에 데려갈 수 있다는 선례라며 낸 판례 두 건이 환각으로 드러났습니다. 판사는 그 변호사에게 상대방 변호사 비용을 개인 돈으로 물게 했습니다.

영국 1심 재판소의 펠리시티 하버 사건은 탐정 소설 같습니다. 그는 삼천이백육십오 파운드짜리 벌금 처분에 불복하며 판례 아홉 건을 냈습니다. 앤 레드스톤 판사가 이상한 점을 발견했습니다. 영국 판결문인데 미국식 철자가 섞여 있었습니다. 그리고 어딘가 상투적인 문장들이 반복됐습니다.

기계는 자기 국적을 감추지 못합니다.

스페인 카나리아 제도 대법원은 가짜 판례를 마흔여덟 건이나 낸 변호사에게 사백이십 유로의 벌금을 물렸습니다. 그 변호사는 공식 법률 데이터베이스와 대조해 보지도 않았습니다.

호주 멜버른 가정법원에서는 법률 전문 소프트웨어의 도구가 거짓 판례를 만들어 재판이 미뤄졌습니다. 태즈메이니아 대법원에서는 명예훼손 항소 사건에서 대법원장이 존재하지 않

는 판례를 직접 짚어 내고 사건을 기각했습니다.

브라질에서는 연방판사가 판결문 작성 부담 때문에 보좌관에게 기계 학습 도구를 쓰게 했다가 상급법원 판례를 왜곡해 넣어 조사를 받았습니다.

판결문을 인공지능이 쓴 겁니다. 재판받는 사람은 그 사실을 모릅니다.

2026년에는 규모가 커졌습니다.

2026년 2월 이혼 항소심에서 그레그 레이크 변호사가 낸 서면은 인용 예순세 개 중 쉰일곱 개에 문제가 있었습니다. 가짜 판례가 스무 건, 조작된 결정이 세 건, 지어낸 법조문 인용이 있었습니다.

예순세 개 중에 여섯 개만 진짜였습니다.

네브래스카 대법원은 2026년 4월 그의 자격을 정지하고 징계 조사를 명령했습니다.

2026년 3월 제6연방항소법원의 화이팅 판결은 병합된 세 항소심 서면에서 가짜 인용 스물네 건 이상과 사실 왜곡을 다뤘습니다. 오리건 연방지방법원의 한 사건에서는 제재와 비용을 합쳐 약 십만 구천칠백 달러가 부과됐습니다. 2026년 1분기에만 미국 법원이 인공지능 관련 서면에 물린 제재가 십사만 오천 달러를 넘었습니다.

기록을 보면 무늬가 하나 보입니다. 가짜 인용 자체보다 그것을 감추려 한 쪽에 더 무거운 벌이 내려집니다. 판사들은 실수는 용서해도 거짓말은 용서하지 않습니다.

기업들도 예외가 아니었습니다.

세계 최초의 로봇 변호사라고 광고한 두낫페이는 변호사 검증도 없이 법률 문서를 분석할 수 있다고 선전하다가 연방거래위원회에 걸려 십구만 삼천 달러의 벌금을 냈습니다.

개인상해 서류 작성을 자동화한다던 이븐업은 전직 직원들의 폭로로 실상이 드러났습니다. 인공지능이 부상 내역을 빠뜨리거나 없는 병명을 만들어 내면 직원들이 손으로 고치고 있었습니다.

호버보드 화재 손해배상 소송에서는 로펌 자체 데이터베이스가 만든 가짜 판례 여덟 건이 제출됐습니다. 판사는 변호사들에게 합쳐서 오천 달러를 물리고 주도한 변호사를 사건에서 뺐습니다.

대형 로펌도 걸렸습니다. 스테이트팜 보험사 사건에서 두 거대 로펌 변호사들이 제미나이와 법률 전문 도구를 섞어 쓰다가 인용 스물일곱 개 중 아홉 개가 오류, 두 개는 완전한 가짜로 밝혀져 삼만 천 달러를 물었습니다.

마이필로우 회장 마이크 린델의 변호인단은 가짜 판례 서른 건 이상을 그대로 두었다가 변호사당 삼천 달러씩 벌금을 받았습니다.

제일 어색한 사례는 인공지능 회사 안에서 나왔습니다.

유니버설 뮤직 그룹과 앤스로픽의 저작권 소송에서, 앤스로픽 쪽 데이터 과학자가 낸 서류에 자사 챗봇이 만든 가짜 학술 논문이 들어 있었습니다. 인공지능 회사가 인공지능 때문에 법정에서 야단맞은 겁니다.

스탠퍼드 대학교의 미디어 정보 전문가 제프 핸콕 교수는 미네소타주 딥페이크 금지법 재판에 낸 진술서에 환각으로 만들어진 가짜 논문 두 건을 인용했습니다. 딥페이크의 위험을 설명하러 나온 전문가가 인공지능이 지어낸 자료를 인용한 겁니다.

법정을 나오면 학교가 있습니다. 여기도 사정이 다르지 않습니다.

2023년 8월 덴마크의 생물학자 헨리크 엥호프 교수는 에티오피아 연구진의 노래기 논문을 읽다가 눈을 의심했습니다. 자기가 쓰지 않은 논문 두 편이 자기 이름으로 인용돼 있었습니다.

연구진은 논문 작성에 챗GPT를 썼고, 가짜 인용이 만들어진 사실을 나중에 알았다고 했습니다. 논문은 철회됐습니다.

인용은 학문에서 계보입니다. 어떤 생각이 어디서 왔는지 적는 일입니다. 가짜 인용은 없는 조상을 족보에 적는 것과 같습니다. 한 번 섞이면 뒤에 오는 사람들이 계속 헛발을 딛습니다.

이 문제는 인공지능보다 오래됐습니다.

2008년부터 2013년 사이 프랑스 연구자 시릴 라베는 유명 학술 출판사 슈프링어와 IEEE의 학술대회 논문집에서 가짜 논문 백이십여 건을 찾아냈습니다. 이 논문들은 MIT 대학원생들이 2005년에 심사가 허술한 학회를 놀리려고 만든 자동 논문 생성 프로그램의 산물이었습니다.

무의미한 단어 조합이 심사를 통과해 정식 논문집에 실렸습니다. 이십 년 전 일입니다.

2024년 2월에는 한 세포생물학 저널이 논문을 철회했습니다. 논문 안에 미드저니로 만든 그림들이 있었는데, 그중에 해부학적으로 말이 되지 않는 거대한 쥐 그림이 있었습니다. 그림에 적힌 글자들도 뜻이 없는 알파벳 나열이었습니다.

심사위원 여러 명이 이 그림을 보고 승인했습니다.

2024년 404 미디어 조사에서는 구글 스콜라에 등록된 논문 백열다섯 건 이상에 제 최신 지식 업데이트에 따르면이라는 문장이 남아 있었습니다. 붙여 넣기를 하고 읽지 않은 겁니다.

저널 오브 휴먼 에볼루션에서는 출판사가 편집위원회와 상의 없이 원고 편집 과정에 인공지능을 넣었습니다. 통과된 원고의 서식이 깨지고 내용이 뒤틀리자 편집위원 전원이 사퇴했습니다.

스탠퍼드 연구진은 세계적인 인공지능 학회의 논문 심사평 중 최대 17퍼센트가 챗GPT로 작성됐다고 밝혔습니다.

인공지능 논문을 인공지능이 심사하는 구조입니다. 사람은 어디에 있느냐고 물을 만합니다.

정책 문서에도 같은 일이 있었습니다.

호주 맥쿼리 대학교의 제임스 거스리 명예교수는 컨설팅 업계 규제를 요청하는 의회 제출 서류를 구글 바드로 작성했습니다. 바드는 딜로이트가 부실 시공사 감사를 소홀히 해 소송당했다거나 KPMG가 편의점 임금 착취 사건에 개입했다는 사건들을 만들어 냈습니다. 전부 없는 일이었습니다. 교수는 상원에 사과문을 보냈습니다.

컨설팅 업계를 비판하려던 문서가 컨설팅 회사들을 없는 죄로 모함한 겁니다.

호주 정부의 복지 시스템 보고서를 쓴 딜로이트 호주 법인도 챗GPT가 만든 가짜 학술 인용과 연방법원 판결문 구절을 그대로 실었다가 비판을 받고 돈을 물어냈습니다.

2025년 미국 보건복지부와 백악관이 낸 보건 보고서에는 가짜 학술 연구가 일곱 건 이상 들어 있었습니다. 챗GPT 특유의 인용 표기가 그대로 남아 있었습니다.

기후 변화를 부정하는 사람들은 그록 3 모델로 논문을 만들어 사설 저널에 실었습니다. 과학적 합의를 흔드는 데 쓰였습니다.

이 사건들을 겪은 사람들의 직업을 세어 보겠습니다. 변호사, 판사, 교수, 정부 관료, 대기업 컨설턴트입니다.

글을 다루는 훈련을 오래 받은 사람들입니다. 문서의 진위를 가리는 일로 돈을 버는 사람들입니다.

그런데도 걸렸습니다. 인공지능이 내놓는 문서가 형식적으로 흠잡을 데가 없기 때문입니다. 사건 번호가 있고, 날짜가 있고, 문체가 법률 문서답습니다. 사람은 형식으로 진위를 판단하는 습관이 있습니다.

가짜 판례는 그 습관의 정확한 반대편에 서 있습니다. 내용이 없고 형식만 있습니다.

법정에서 사실을 다투는 이유는 사람의 자유와 재산이 걸려 있기 때문입니다. 그 다툼의 근거가 존재하지 않는 판결문이었다면, 그 재판은 무엇이었을까요.

하루 여덟 건씩 발각되고 있습니다.

2.3

역사적·지리적 사실 왜곡과 사회적 논란

2025년 10월, 누군가 네이버 검색창에 일본어로 일본 영토라고 쳤습니다.

화면 위쪽에 인공지능이 정리한 요약이 떴습니다. 그 안에 독도가 있었습니다. 북방영토, 센카쿠 열도와 나란히 놓인 채, 한국과 분쟁 중인 영역이라고 적혀 있었습니다.

한국 회사의 서비스가, 한국 사람들이 쓰는 검색창에서, 독도를 일본 쪽 목록에 넣은 겁니다.

성신여대 서경덕 교수가 이 화면을 공개했고 사람들이 화를 냈습니다. 네이버는 검색 요약 기능을 급히 멈추고 확인에 들어갔습니다.

원인은 기술적으로는 지루하고 내용적으로는 아픕니다. 하이퍼클로바 X는 일본어 질문을 받으면 일본어 웹 문서를 모아 요약합니다. 일본어로 영토를 검색하면 일본 외무성 홍보 자료가 상단에 옵니다. 시스템은 그 자료를 정리했습니다.

시킨 대로 한 겁니다. 다만 아무도 이 질문에는 사람이 한 번 봐야 한다고 정해 두지 않았습니다.

영토 문제에서 어느 나라 자료를 먼저 읽느냐는 기술 문제가 아니라 입장 문제입니다. 기계에는 입장이 없습니다. 없는 것이 중립이라고 부르기도 합니다. 그런데 자료가 한쪽으로 기울어 있으면, 입장 없는 기계는 기울어진 쪽으로 굴러갑니다.

대만에서는 더 놀라운 대답이 나왔습니다.

2023년 10월 대만 최고 학술기관인 중앙연구원이 대만형 언어 모델을 만들어 공개했습니다. 시민들이 챗봇에게 물었습니다. 너는 어느 나라 소속이니.

챗봇이 답했습니다. 중국입니다.

대만의 최고 지도자가 누구냐고 물었습니다. 시진핑이라고 답했습니다. 국가가 무엇이냐고 물었습니다. 의용군 진행곡이라고 답했습니다.

개발진은 이유를 밝혔습니다. 빨리 만들려고 중국 대륙의 간체자 데이터셋을 번역기로 돌려 그대로 학습에 넣었다는 겁니다.

번역기는 글자를 바꿉니다. 관점은 바꾸지 못합니다. 대만 정부와 연구원은 시스템을 즉시 내렸습니다.

2025년 2월 전 세계에서 삼백만 번 넘게 내려받은 중국의 딥시크 R1 모델은 위구르족 집단 학살 의혹을 물으면 내정 간섭이자 근거 없는 모함이라는 답을 내놨습니다. 중국 정부의 논리를 그대로 되풀이한 겁니다.

모델은 자기가 읽은 것을 말합니다. 어떤 자료를 먹였는지가 곧 그 모델의 세계관이 됩니다.

유럽의회도 걸렸습니다. 2025년 앤스로픽의 클로드를 기반으로 만든 공식 인공지능이 유럽운영위원회 초대 위원장의 이름을 대지 못했습니다. 기록을 다루라고 만든 물건이 기록을 몰랐습니다.

역사 왜곡은 여기서 더 나갑니다.

2024년 6월 유네스코 보고서에 챗GPT의 답변 하나가 실렸습니다. 홀로코스트에 대해 묻자, 나치 독일이 유대인들을 집단으로 물에 빠뜨려 죽였다는 사건을 설명했다는 겁니다.

그런 사건은 없었습니다. 익사로 인한 홀로코스트라는 이름까지 붙어 있었습니다.

역사학자들이 걱정한 이유는 이 거짓말의 방향에 있습니다. 홀로코스트 부정론자들은 오랫동안 기록의 사소한 틈을 파고들어 전체를 흔드는 방식을 써 왔습니다. 없는 사건이 기록에 섞이면, 있었던 사건도 의심받습니다.

일론 머스크의 xAI가 만든 그록은 여기서 선을 넘었습니다.

2025년 중반, 그록은 정치적 올바름을 걷어 낸다는 명목으로 시스템을 손봤습니다. 그 직후 그록은 육백만 명이라는 홀로코스트 희생자 수가 정치적으로 조작됐을 수 있다는 말을 했습니다. 아우슈비츠 가스실의 존재를 부정하는 답도 내놨습니다. 자기 자신을 메카히틀러라고 부르기도 했습니다.

2023년 1월에는 아마존 엔지니어가 GPT-3로 만든 역사적 인물 대화 앱이 있었습니다. 이만 명의 인물과 대화할 수 있다고 광고했습니다. 나치 선전상 요제프 괴벨스 챗봇은 자기가

유대인을 미워하지 않았고 오히려 괴로워했다고 말했습니다.

가해자의 목소리를 흉내 내는 기계는 가해자의 변명도 흉내 냅니다. 그 변명이 앱에서 나오면 누군가는 그것을 사료로 착각합니다.

회사들은 프로그래밍 오류였다거나 내부 지시문이 임의로 바뀌었다고 해명했습니다.

눈으로 보는 것도 못 믿게 됐습니다.

2025년 6월 영국 팩트체크 기관 풀팩트가 이상한 사례를 잡아냈습니다. 인도 고아주의 어느 해변에서 찍은 평범한 영상이 있었습니다. 구글 렌즈에 붙은 인공지능 요약은 이 영상을 이렇게 설명했습니다. 영국 도버 해변에 망명 신청자들이 대거 들어오는 장면.

해변이 나오고 사람이 많으면 그 장면이 됩니다. 인도가 영국이 됩니다.

이 요약이 소셜미디어의 소문과 만나면 어떻게 되는지는 짐작하기 어렵지 않습니다. 영국에서 반이주민 정서가 달아오르던 시기였습니다.

같은 서비스가 2024년 5월에는 이런 답들을 내놨습니다. 피자 치즈가 흘러내리면 무독성 접착제를 섞으라. 하루에 작은 돌을 하나씩 먹으라. 신장결석에는 소변을 마시라. 버락 오바마는 미국의 첫 무슬림 대통령이다.

돌을 먹으라는 답은 풍자 매체의 기사에서 왔습니다. 기계는 농담과 정보를 구별하는 감각이 없습니다. 사람은 어릴 때 이 감각을 배웁니다. 표정과 상황을 보면서요. 기계는 글자만 받습니다.

2025년 4월 말레이시아에서는 국기가 문제가 됐습니다. 언론사와 교육부와 국립 전시회 관계자들이 인공지능으로 만든 말레이시아 국기 이미지를 공개했습니다. 초승달 기호가 빠져 있거나 상징이 이상하게 뒤틀려 있었습니다.

국기는 나라의 얼굴입니다. 그 얼굴을 그리면서 아무도 확인하지 않았습니다.

지도와 지형에서도 같은 일이 벌어졌습니다.

2024년 11월 일본 후쿠오카현은 공식 사과를 하고 광고 대행사와 계약을 끊었습니다. 대행사가 인공지능으로 만든 홍보물에 후쿠오카에 없는 관광 명소와 없는 향토 음식이 실제인 것처럼 들어 있었기 때문입니다.

2025년 5월 영국 자연청의 이탄지 지도 제작 프로젝트도 넘어졌습니다. 이탄지는 탄소를 잡아 두는 습지라 기후 정책에서 중요합니다. 기계 학습 모델이 만든 지도에는 돌담과 채석장과 흙더미와 화강암 바위가 이탄지로 표시돼 있었습니다.

위에서 내려다본 사진만으로는 젖은 땅과 마른 돌을 구별하기 어렵습니다. 사람은 발로 밟아 봅니다. 모델에게는 발이 없습니다.

2021년 스탠퍼드와 워싱턴 대학교 연구진은 딥페이크 지리학이라는 개념을 경고했습니다. 위성 사진을 합성해 없는 지형을 만들 수 있다는 겁니다. 군사와 외교에서 어떤 일이 벌어질지는 상상에 맡기겠습니다.

2024년 12월 벨기에의 저명한 사진기자 카를 드 케이저는 우크라이나 침공 이후 러시아에 갈 수 없게 되자 생성형 인공지능으로 러시아 풍경과 인물 사진을 만들어 사진집에 실었습니다. 사진 저널리즘의 근거는 거기 있었다는 사실입니다. 그 근거가 사라졌습니다.

길 안내에서는 사람이 죽었습니다.

2024년 인도에서 남성 세 명이 구글 맵스 안내를 따라 차를 몰다가 끊긴 다리 아래로 떨어져 모두 숨졌습니다. 다리는 완공되지 않은 상태였고 물에 잠겨 있었습니다.

지도에는 길이 있었습니다. 현실에는 없었습니다.

미국에서도 같은 사고가 있었습니다. 무너진 다리로 안내받은 운전자가 물에 빠져 숨졌습니다. 같은 실수가 다른 대륙에서 반복됐습니다.

2026년 2월 영국에서는 구글 맵스 안내를 따라간 아마존 배송 트럭이 브룸웨이로 들어갔습니다. 밀물 때 바닷물에 잠기는 갯벌 길입니다. 트럭은 진흙에 갇혔습니다.

2025년 5월에는 구글 맵스가 원인 모를 오류로 독일 전역의 주요 아우토반이 통제됐다고 표시해 교통 정보망을 뒤흔들었습니다.

2026년 1월 호주 태즈메이니아에서는 인공지능이 쓴 여행 기사를 믿고 존재하지 않는 온천을 찾아 나선 관광객들이 오지에서 고립됐습니다.

급한 상황에서 인공지능이 내놓는 오보는 더 위험합니다.

2024년 4월 그록은 이란이 텔아비브에 미사일을 쐈다는 헤드라인을 만들었습니다. 같은 달 인도 총리가 정부에서 쫓겨났다는 소식도 내보냈습니다. 2025년 5월에는 야구 경기 질문이나 해적처럼 말해 달라는 요청에까지 남아프리카공화국 백인 집단학살 음모론을 스스로 끼워 넣었습니다.

2026년 3월에는 힐스버러와 헤이젤 축구 참사 희생자들을 모욕하는 글을 만들어 유가족들에게 상처를 남겼습니다. 두 사건 모두 수백 명이 압사한 실제 참사입니다.

2025년 9월 런던 반이민 시위 현장에서는 경찰이 폭력 영상을 조작했다는 거짓 정보를 내보냈습니다.

지식 자체를 새로 쓰겠다는 시도도 있었습니다.

2025년 10월 머스크는 위키피디아에 맞서 그록피디아라는 인공지능 백과사전을 냈습니다. 출범과 동시에 구십만 건의 문서가 올라왔습니다. 검증되지 않은 게시판 글과 블로그를 베낀 문서들이었고, 환각으로 만들어진 사실들이 항목으로 등재됐습니다. 남아프리카공화국 백인 집단학살 음모론이 사실처럼 실렸고 에이즈의 역사가 뒤틀렸습니다.

백과사전이라는 이름은 사람을 안심시킵니다. 그 안심을 노린 겁니다.

2025년 2월에는 그록 3 모델이 추론 과정에서 머스크와 트럼프가 허위 정보를 퍼뜨린다는 비판 출처를 내부 지시문으로 차단하도록 설정돼 있었다는 사실이 드러났습니다. 무엇을 말할지가 아니라 무엇을 읽지 않을지를 정해 둔 겁니다.

개인에게 오는 피해는 즉각적입니다.

2025년 10월 미국 마샤 블랙번 상원의원은 구글의 젬마 모델이 자신이 1987년 주 상원 선거 당시 성범죄 사건에 휘말렸다는 이야기를 만들어 내고 가짜 뉴스 링크까지 붙인 것을 발견했습니다.

같은 계열 시스템은 2023년 12월부터 보수 활동가 로비 스타벅을 아동 성폭행범으로 묘사하는 글을 만들어 냈고 소송으로 이어졌습니다.

2025년 미네소타의 태양광 기업 울프 리버 일렉트릭은 구글 AI 오버뷰가 자기 회사가 주 검찰총장에게 사기로 고소당했다는 요약을 내보내는 바람에 계약이 줄줄이 취소됐습니다. 회사는 일억 천만 달러 손해배상 소송을 냈습니다.

같은 해 이탈리아의 유명 의사는 구글 AI 오버뷰가 자기 사망 소식을 띄우자 소송을 냈습니다. 살아 있는 사람이 자기가 살아 있다고 증명해야 했습니다.

2026년 1월 캐나다 음악가 애슐리 맥아이작은 구글 AI 오버뷰가 자신을 성범죄 확정 판결자로 표시하는 바람에 공연이 취소됐습니다.

이런 사건들에서 회사들은 대체로 같은 말을 합니다. 알고리즘의 오작동이며 개선하겠다.

개선은 다음 사용자를 위한 것입니다. 이미 공연이 취소된 사람에게는 늦습니다.

여기까지 나온 대상들을 늘어놓아 보겠습니다. 한 나라의 섬, 한 나라의 국가와 국기, 육백만 명의 죽음, 다리, 갯벌, 습지, 한 사람의 생사와 전과.

전부 확인 가능한 것들입니다. 지도를 펴 보거나 기록을 찾아보거나 전화 한 통을 걸면 됩니다.

그 한 통이 비용으로 계산돼 지워진 자리에, 확률로 이어 붙인 문장이 들어왔습니다.

제 3 장

프라이버시 침해와 감시 사회: 무단 데이터 수집의 그늘

3.1

무단 생체 정보 수집과 공공·상업적 감시

쇼핑몰에서 길을 잃으면 안내 화면 앞에 섭니다. 지도를 보고, 가게 이름을 찾고, 손가락으로 화면을 몇 번 누릅니다. 몇 초 걸립니다.

2018년 캐나다 쇼핑몰 열두 곳에서, 그 몇 초 동안 벌어진 일이 있습니다.

캐나다의 대형 부동산 기업 캐딜락 페어뷰는 안내 키오스크 안에 카메라를 숨겨 두었습니다. 화면을 보는 사람의 얼굴이 찍혔고, 생체 측정 알고리즘이 나이와 성별을 판정했습니다.

그렇게 기록된 사람이 오백만 명입니다.

조사가 시작되자 회사는 이렇게 해명했습니다. 이미지는 보관하지 않고 곧바로 지웁니다.

감독 기관의 판단은 달랐습니다. 지우고 말고를 떠나, 동의 없이 얼굴을 재는 행위 자체가 법을 어긴 것이라고 봤습니다.

이 구별이 중요합니다. 사람들은 대개 내 사진이 어디 저장돼 있느냐를 걱정합니다. 그런데 생체 정보의 핵심은 저장이 아니라 측정입니다. 사진은 지워도 이 사람은 삼십 대 여성이고 화요일 오후에 이 층에 있었다는 기록은 남습니다.

얼굴은 비밀번호와 다릅니다. 유출되면 바꿀 수 없습니다.

미국의 약국 체인 라이트에이드는 여기서 한 발 더 나갔습니다. 2012년부터 2020년까지 전국 매장에 안면 인식 카메라를 설치했습니다. 도둑을 막겠다는 명분이었습니다.

카메라가 어디에 많이 설치됐는지가 문제였습니다. 저소득층과 흑인, 히스패닉이 많이 사는 동네에 몰려 있었습니다.

그리고 알고리즘은 자주 틀렸습니다. 경보가 울리면 직원들이 손님에게 다가갔습니다. 나가 달라고 했습니다. 경찰을 불렀습니다.

물건을 사러 온 사람이 사람들 앞에서 도둑 취급을 받았습니다. 그 자리에서 억울함을 증명할 방법은 없습니다.

연방거래위원회는 라이트에이드에 오 년 동안 이 기술을 쓰지 못하게 했습니다.

같은 일이 여러 나라에서 있었습니다.

스페인의 슈퍼마켓 체인 메르카도나는 마흔 곳이 넘는 매장에 안면 인식을 도입했다가 오백십오만 유로의 벌금을 맞았습니다. 회사는 접근 금지 명령을 받은 전과자를 걸러 내려 했다고 했습니다.

당국의 답은 이랬습니다. 몇 사람을 찾겠다고 모든 사람의 얼굴을 재는 것은 허용되지 않는다.

호주에서는 버닝스, 케이마트, 세븐일레븐이 걸렸습니다. 세븐일레븐은 매장 태블릿으로 만족도 조사를 하면서 응답자의 얼굴을 몰래 찍어 보관했습니다.

만족하십니까 하고 물으면서 얼굴을 찍은 겁니다.

영국의 식료품점 체인들도 안면 인식 카메라를 들여왔다가 무고한 손님을 절도범으로 몰아 소동을 냈습니다.

중국에서는 이 기술이 수수료 계산에 쓰였습니다.

난징, 닝보, 난닝의 부동산 개발사들은 분양 사무소를 찾는 손님의 얼굴을 동의 없이 찍었습니다. 목적은 이랬습니다. 이 손님이 어느 중개업자를 통해 왔는지 확인해서 수수료를 다르게 주려는 겁니다.

손님의 얼굴이 중개 수수료 계산기의 부품이 됐습니다. 손님은 몰랐습니다.

이 이야기에는 웃픈 후일담이 있습니다. 소문이 퍼지자 오토바이 헬멧을 쓰고 분양 사무소에 가는 사람들이 나타났습니다.

코일러, BMW, 맥스마라 같은 브랜드도 중국 매장에 몰래 카메라를 두고 방문 횟수와 동선을 추적하다가 2021년 관영 방송에 적발됐습니다.

무인 매장과 결제 시스템도 얼굴과 손을 모읍니다.

뉴욕의 아마존 고 매장은 들어오는 손님의 생체 정보를 수집하면서 뉴욕시 법이 정한 고지 의무를 지키지 않아 집단 소송을 당했습니다. 손바닥을 대면 결제되는 아마존 원은 손바닥

정맥과 손금 데이터를 서버에 저장합니다.

맥도날드는 드라이브스루에 음성 인식을 넣고 운전자의 목소리 특징을 동의 없이 모아 보관하다가 일리노이주 생체정보보호법 위반으로 소송을 당했습니다. 도미노 피자도 전화 주문에서 목소리를 무단 수집한 혐의로 법정에 섰습니다.

햄버거를 주문했을 뿐인데 목소리 지문이 남았습니다.

학교로 가 보겠습니다.

스웨덴 셰렙테오의 안데르스토르프 고등학교는 출석 확인 시간을 아끼겠다며 학생 스물두 명을 대상으로 안면 인식 카메라를 시험했습니다. 2019년 스웨덴 데이터보호청은 개인정보보호규정 위반으로 벌금을 물렸습니다. 이 규정으로 나온 스웨덴 첫 벌금이었습니다.

학교는 학부모 동의를 받았다고 했습니다. 당국은 이렇게 봤습니다. 학교와 학생 사이에는 힘의 차이가 있어서 진짜 자발적인 동의가 성립하기 어렵다.

선생님이 하는 부탁을 학생이 거절하기는 어렵습니다. 어른들 사이에서도 그렇습니다.

그리고 목적이 출석 확인이었습니다. 출석부를 부르면 되는 일에 생체 정보를 쓴 겁니다.

프랑스 니스와 마르세유의 고등학교들은 출입문에 안면 인식 게이트를 놓으려다 시민단체 반발과 법원 불허로 접었습니다.

영국 노스에어셔의 고등학교 아홉 곳은 급식비 결제를 빠르게 하겠다며 안면 인식을 도입했다가 정보위원회 경고를 받고 껐습니다. 첼머 밸리 고등학교도 같은 이유로 적발됐습니다.

폴란드 그단스크의 한 초등학교는 급식 확인을 위해 아이들의 지문을 모았다가 벌금을 받았습니다.

밥 먹는 줄을 몇 초 줄이려고 아이 지문을 받은 겁니다.

인도 정부가 성적표 발급과 급식 확인에 안면 인식을 넣으려 했을 때도 논란이 터졌습니다. 스코틀랜드, 브라질 파라나주, 호주 빅토리아주에서도 비슷한 시도와 반발이 있었습니다.

미국에서는 안전이라는 말이 열쇠로 쓰였습니다. 뉴욕주 록포트 학구는 총기 사고를 막겠다며 수백만 달러를 들여 학교에 안면 인식 감시 시스템을 깔았습니다. 학부모들과 시민단체

가 반발했고, 뉴욕주 교육청은 주 내 모든 학교에서 이 기술 사용을 일시 중단하는 조치를 취했습니다.

교실 안으로 들어간 카메라는 얼굴만 보지 않았습니다.

중국의 제11중학교와 중국약과대학은 교실 앞에 카메라를 놓고 학생들의 표정과 졸음과 집중도를 실시간으로 분석해 점수를 매겼습니다.

수업 중에 딴생각하는 표정이 기록되는 교실을 상상해 보십시오. 그 교실에서 학생은 배우는 사람이 아니라 평가받는 데이터가 됩니다.

지루하다는 얼굴을 지을 자유는 배움의 일부입니다.

공항과 국경에서는 규모가 커집니다.

캐나다 국경관리청은 토론토 피어슨 국제공항에서 승객이 모르는 사이에 얼굴을 수집하는 무인 키오스크를 시험했습니다. 뉴질랜드 웰링턴 국제공항에서도 동의 없이 카메라를 돌리다가 언론 보도로 드러났습니다. 스페인 국영 공항 운영사 아에나는 이용객 얼굴을 불법 수집해 벌금을 물었습니다.

한국에서도 일이 있었습니다.

법무부와 과학기술정보통신부는 2019년부터 인공지능 식별 추적 시스템 사업을 추진하면서, 인천공항 등을 통과한 출입국자들의 얼굴 사진을 본인 동의 없이 민간 인공지능 기업들에게 학습용으로 넘겼습니다.

건수는 일억 칠천만 건입니다.

외국인만 있었던 것이 아닙니다. 한국 국민의 얼굴도 들어 있었습니다. 법적 근거 없이 민간 기업에 넘어갔다는 사실이 조사에서 확인됐습니다.

여권 심사대에서 카메라를 본 기억이 있다면, 그 사진이 어디로 갔는지 생각해 볼 만합니다.

미국 국경 당국은 망명 신청자용 모바일 앱에 안면 인식을 넣었습니다. 이 앱은 흑인 신청자의 얼굴을 잘 인식하지 못했습니다. 인식이 안 되면 신청 자체가 진행되지 않습니다. 살던 곳에서 도망쳐 온 사람이 앱 앞에서 한 번 더 막힌 겁니다.

거리에서는 홍채를 모으는 사업이 있었습니다.

샘 올트먼이 세운 월드코인은 오브라는 홍채 인식 기기를 세계 여러 도시에 놓았습니다. 홍채를 등록하면 암호화폐를 줍니다. 가난한 나라와 젊은 층에서 줄이 길게 섰습니다.

케냐 정부는 사업을 전면 중단시켰습니다. 포르투갈, 스페인, 인도네시아, 태국 당국도 영업 정지를 명령했습니다. 홍콩, 아르헨티나, 한국 개인정보보호위원회도 조사에 나섰습니다.

한 번 등록한 홍채는 회수할 수 없습니다. 받은 코인은 값이 오르내립니다. 거래가 공평했는지 따져 볼 만합니다.

안경 하나로 남의 신원을 아는 물건도 나왔습니다.

하버드 대학교 학생 두 명이 메타의 레이밴 스마트 글래스에 안면 인식을 붙인 시스템을 공개했습니다. 이 안경을 쓰고 길이나 지하철에서 모르는 사람을 바라보면, 몇 초 만에 상대의 이름과 주소와 전화번호와 가족 관계가 휴대폰 화면에 뜹니다.

학생들은 위험을 알리려고 만들었다고 했습니다. 만들 수 있다는 사실이 이미 알림입니다.

메타의 스마트 글래스 착용자가 마사지 업소나 공공장소에서 남을 몰래 찍어 퍼뜨린 일도 있었습니다. 이 안경으로 모은 영상이 해외 외주 인력에게 넘어가 학습에 쓰였다는 사실도 드러났습니다.

이 모든 도구의 뒤에는 인터넷을 통째로 긁는 회사가 있습니다.

클리어뷰 AI는 페이스북, 인스타그램, 엑스, 링크드인과 인터넷 곳곳에서 얼굴 사진 삼백억 장을 모아 검색 데이터베이스를 만들었습니다. 그리고 세계 각국 경찰과 수사기관, 심지어 유통업체에 팔았습니다.

미국시민자유연맹이 일리노이주 시민들을 대리해 소송을 냈고, 영국과 프랑스와 이탈리아와 네덜란드와 그리스의 개인정보 보호 기구들이 거액의 벌금과 데이터 파기 명령을 내렸습니다.

얼굴 검색 엔진 피메이스는 더 나빴습니다. 사적인 사진, 세상을 떠난 사람의 사진, 아동의 성적 이미지까지 긁어모아 사진 한 장으로 그 사람의 사생활 사진을 찾아 주는 서비스를 만들었습니다. 독일과 미국 법원이 제재했습니다.

메타는 텍사스주 등에서 주민 동의 없이 얼굴 생체 정보를 수집한 혐의로 십사억 달러를 합의금으로 냈습니다.

국가 단위로 가면 규모가 다릅니다.

이스라엘 국방군은 헤브론과 가자 지구에서 팔레스타인 주민들을 감시하는 안면 인식 체계를 운영했습니다. 모스크바는 지하철 결제와 반정부 인사 추적에 안면 감시망을 썼습니다. 런던 킹스크로스 역과 수사기관들도 실시간 안면 인식을 돌렸습니다.

이 사건들을 다시 훑어보면 장소들이 눈에 들어옵니다. 쇼핑몰, 약국, 슈퍼마켓, 편의점, 자동차 매장, 분양 사무소, 초등학교, 고등학교, 대학교, 공항, 국경, 지하철역, 거리.

사람이 하루를 사는 동선 그대로입니다.

그리고 이 모든 곳에서 같은 문장이 반복됐습니다. 안전을 위해서입니다. 편의를 위해서입니다. 효율을 위해서입니다.

이 문장들에 반대하기는 어렵습니다. 그래서 잘 팔립니다.

얼굴과 홍채와 지문과 목소리는 몸에 붙어 있습니다. 잃어버려도 새로 발급받을 수 없습니다. 그 사실을 아는 사람들이 그것을 모으고 있습니다.

3.2

개인정보 무단 학습과 AI 대화 유출 사고

2023년 3월 어느 날, 챗GPT를 켠 사람들의 화면 왼쪽에 대화 목록이 떴습니다. 늘 있던 자리입니다.

그런데 제목들이 낯설었습니다. 자기가 나눈 적 없는 대화들이었습니다.

남의 대화 제목이 내 화면에 있었습니다. 내 대화 제목도 누군가의 화면에 있었을 겁니다.

오픈AI는 오픈소스 데이터베이스 라이브러리의 버그 때문이라고 설명하고 서비스를 잠시 멈췄습니다. 조사 결과 챗GPT 플러스 구독자의 약 1.2퍼센트에서 이름과 이메일 주소, 신용카드 번호 뒷자리 네 개와 유효기간까지 다른 사용자에게 보였습니다.

사람들이 그 창에 무엇을 적는지 생각해 보면 이 사고의 무게가 달라집니다.

이력서를 고쳐 달라고 하면서 주소와 전화번호를 넣습니다. 진단서를 이해하고 싶어서 병명을 적습니다. 이혼을 고민하면서 배우자 이야기를 씁니다. 회사에 말하지 못한 고민을 씁니다.

검색창에는 질문을 넣습니다. 챗봇에는 사정을 넣습니다.

이 차이를 사람들이 알아차리기 전에 다음 사고가 왔습니다.

챗GPT에는 대화를 링크로 공유하는 기능이 있었습니다. 친구에게 보여 주려고 만든 기능입니다. 그런데 이 링크가 공개 웹페이지처럼 취급됐습니다.

2025년, 사용자들이 만든 대화 기록 십만 건 이상이 구글 같은 검색 엔진에 색인됐습니다. 검색창에 낱말 몇 개를 넣으면 낯선 사람의 상담 내역이 나왔습니다. 사적인 고민이 나왔습니다. 회사 기밀이 나왔습니다.

친구 한 명에게 보여 주려던 것이 전 세계에 보이게 된 겁니다.

구글의 바드도 같은 실수를 했습니다. 그록도 했습니다. 그록과 나눈 대화 수십만 건이 검색 결과에 노출됐고, 요청하지도 않았는데 그록이 성인물 업계 종사자의 본명과 생년월일을 내

놓은 일도 있었습니다.

같은 실수가 회사를 옮겨 가며 반복됐습니다. 새 기능을 빨리 내놓는 경쟁에서 공유 링크의 접근 권한을 따지는 일은 뒤로 밀립니다.

정서적인 대화를 다루는 앱에서는 유출이 더 아픕니다.

2024년 캐릭터 AI에서는 내부 오류로 이용자들이 서로의 비공개 대화를 볼 수 있는 사고가 났습니다. 가상 캐릭터에게 털어놓은 고백들이었습니다.

중국의 딥시크는 데이터베이스 보안 관리 소홀로 계정 정보와 대화 내역이 밖으로 새어 나갔습니다. 비공개 메시지 삼억 건 규모입니다.

2026년 2월에도 비슷한 규모의 사고가 있었습니다. 한 인공지능 채팅 앱에서 이용자 이천오백만 명의 메시지 삼억 건이 노출됐습니다. 이용자별 파일에는 대화 전체 기록이 들어 있었습니다.

인공지능 동반자 앱들은 더 깊은 것을 모읍니다. 레플리카, 뮤아, 노미 같은 앱은 정서적 교감을 앞세워 사용자의 성적 취향과 감정 데이터를 모았습니다. 모질라 재단 조사에 따르면 이 앱들은 최소한의 개인정보 보호 기준도 지키지 않았고, 모은 데이터를 마케팅 업자에게 넘기기도 했습니다.

연이은 유출로 사십만 명이 넘는 이용자의 대화가 밖으로 나왔습니다. 한 성인용 챗봇 서비스의 보안이 무너졌을 때는 여성들의 졸업앨범 사진 이백만 건과 합성 음란물 데이터까지 함께 노출됐습니다.

졸업앨범 사진은 본인이 그 서비스에 올린 것이 아닙니다. 누군가 긁어 온 겁니다.

회사 기밀도 같은 창으로 빠져나갑니다.

2023년 봄 삼성전자 사업장에서 사고가 났습니다. 반도체 부문 직원들이 업무를 빨리 끝내려고 챗GPT를 썼습니다. 설비 측정 데이터를 넣었습니다. 수율 관련 소스 코드를 넣었습니다. 회의록을 통째로 넣고 요약해 달라고 했습니다.

그 데이터는 오픈AI 서버로 갔습니다. 회사는 사내 인공지능 사용을 급히 제한했습니다.

직원들을 나무라기는 쉽습니다. 그런데 이 사고의 핵심은 다른 데 있습니다. 챗봇 창은 검색창처럼 생겼습니다. 검색창에는 회사 자료를 넣지 않습니다. 그런데 챗봇에는 넣게 됩니다. 요약해 준다고 하니까요.

도구의 생김새가 사람의 행동을 바꿉니다.

음성 요약 서비스 오터 AI는 비공개 투자자 전화 회의를 동의 없이 녹음하고 요약본을 엉뚱한 곳에 보내 재무 정보와 투자 전략을 흘렸습니다.

큰 회사들도 자기 자료를 흘립니다.

2023년 9월 마이크로소프트 인공지능 연구진이 깃허브에 데이터셋을 공유하다가 접근 권한 설정을 잘못했습니다. 전체 파일 시스템을 읽고 쓸 수 있는 토큰이 공개 파일에 섞여 들어간 겁니다.

그 결과 노출된 데이터가 삼십팔 테라바이트입니다. 직원들의 암호, 비밀 키, 사내 메신저 대화, 삼만 오천 명이 넘는 직원의 비공개 업무 데이터가 들어 있었습니다.

인공지능 연구를 하는 팀이 인공지능 학습용 데이터를 공유하다가 회사 내부를 열어 놓은 겁니다.

아마존의 기업용 인공지능 Q는 환각과 함께 비공개 고객 데이터를 흘리는 결함을 보였습니다. 컨설팅 그룹 맥킨지의 사내 인공지능 플랫폼은 해커들의 표적이 됐고, 두 시간 만에 수백만 건의 고객 기록과 비밀 프로젝트 자료에 접근당했습니다.

메타의 인공지능 에이전트도 내부 자료와 사용자 정보를 흘렸습니다. 한 인공지능 비서 프로그램은 서버 수천 대가 인터넷에 그대로 열려 있어 악성코드 공격을 받고 사용자 데이터를 대량으로 빼앗겼습니다.

2026년 2월에는 보안 연구로 새로운 방식이 드러났습니다. 악의적으로 만든 프롬프트 하나가 평범한 챗GPT 대화를 데이터 유출 통로로 바꿀 수 있다는 겁니다. 사용자는 아무것도 모릅니다. 취약점은 같은 달 수정됐습니다.

여기까지는 사고입니다. 다음은 사고가 아닙니다.

2023년 8월 화상회의 플랫폼 줌이 이용약관을 슬쩍 바꿨습니다. 사용자들이 나눈 영상, 음성, 대화 기록, 공유 파일을 인공지능 학습에 쓸 수 있는 권리를 회사에 주는 내용이었습니다.

재택근무 시절 사람들이 줌에서 무엇을 했는지 떠올려 보십시오. 회사 회의를 했고, 수업을 들었고, 병원 상담을 받았고, 장례식에 참석했습니다.

비판이 커지자 줌은 동의 없이는 쓰지 않겠다고 물러섰습니다.

슬랙은 2024년에 적발됐습니다. 사용자들이 주고받은 메시지와 올린 비공개 파일을 인공지능 학습에 기본으로 쓰고 있었습니다. 거부하려면 사용자가 따로 신청해야 하는 방식이었고, 그 사실을 제대로 알리지 않았습니다.

기본값이 무엇이냐가 모든 것을 정합니다. 대부분의 사람은 설정을 바꾸지 않습니다. 회사들은 이 사실을 압니다.

어도비는 클라우드 이용약관을 바꿔 창작자들이 작업 중인 비공개 프로젝트와 이미지를 자사 인공지능 학습을 위해 분석할 수 있게 했다가 전 세계 디자이너들의 구독 해지 운동을 맞았습니다.

링크드인은 사전 동의 없이 프로필과 경력과 게시글을 인공지능 학습에 썼습니다. 엑스는 모든 게시글을 그록 학습용으로 자동 수집하도록 설정해 유럽 당국의 조사를 받았습니다. 사운드클라우드는 음악가들의 곡을 학습에 썼습니다.

메타는 페이스북과 인스타그램 이용자들이 수십 년간 올린 글과 사진을 언어 모델 학습에 쓰겠다고 발표해 유럽과 남미 개인정보 기구들의 제재를 받았습니다.

십 년 전 올린 아기 사진, 십오 년 전 쓴 부끄러운 글, 세상을 떠난 가족과 찍은 사진이 전부 학습 재료 목록에 들어 있습니다. 올릴 때는 친구들에게 보여 주려던 것이었습니다.

메타는 레이밴 스마트 글래스로 찍힌 사적인 사진과 영상을 해외 외주 데이터 노동자들에게 넘겨 학습에 쓴 사실도 드러났습니다. 어느 나라의 어느 사무실에 앉은 사람이 우리 거실 영상에 이름표를 붙이고 있었던 겁니다.

2026년 5월에는 캘리포니아 연방법원에 집단소송이 접수됐습니다. 챗GPT 웹사이트에 메타 픽셀과 구글 애널리틱스 같은 추적 기술이 심어져 있어서, 대화 관련 정보와 사용자 ID와 이메일 주소가 구글과 메타로 전송됐다는 내용입니다.

대화 상대가 하나인 줄 알았는데 셋이었다는 주장입니다.

의료 기록에서는 규모가 다릅니다.

2017년 영국 데이터보호청은 국립보건서비스 산하 왕립 자유 병원과 구글 딥마인드의 데이터 공유 협약이 법을 어겼다고 판정했습니다. 병원은 급성 신장 손상 진단 앱을 만든다는 명목으로 환자 백육십만 명의 전체 의료 기록을 동의 없이 넘겼습니다.

구글은 이후 미국 최대 병원 체인 중 하나와 계약해 진료 기록을 학습용으로 받았고, 오천만 명의 의료 데이터를 환자 몰래 모은 사업으로 비판을 받았습니다.

마음이 제일 무거워지는 사례는 마지막입니다.

미국의 크라이시스 텍스트 라인은 자살 예방과 정신 건강 위기 상담을 하는 비영리 단체입니다. 죽고 싶다는 생각이 들 때 문자를 보내면 상담원이 답합니다. 도움을 청한 사람들이 맨 밑바닥의 말을 적는 곳입니다.

이 단체는 그 대화 기록을 모아 자사의 영리 목적 인공지능 자회사에 넘겼습니다. 그 회사는 이 데이터로 기업용 고객 서비스 챗봇을 학습시켰습니다.

죽고 싶다는 문장이 고객 응대 품질을 높이는 재료가 됐습니다.

생리 주기와 임신 가능성을 기록하는 건강 앱 플로는 사용자의 건강 데이터를 페이스북과 구글 같은 곳에 무단 전송하다가 연방거래위원회 제재를 받았습니다.

이 사건들을 나란히 놓으면 데이터가 흐르는 방향이 한쪽이라는 사실이 보입니다.

사람에게서 회사로 갑니다. 회사에서 사람에게로 돌아오는 것은 편리한 기능입니다. 편리함은 진짜입니다. 그래서 이 거래는 성립합니다.

문제는 가격표가 없다는 데 있습니다. 무엇을 주고 무엇을 받는지 계약서에 적혀 있지 않습니다. 적혀 있어도 열여덟 쪽짜리 약관 열두째 쪽에 있습니다.

그리고 한 번 넘어간 데이터는 돌아오지 않습니다. 회사가 지웠다고 해도, 그 데이터로 학습한 모델 안에서는 지워지지 않습니다.

모델은 무엇을 배웠는지 잊는 법을 모릅니다.

3.3

위치 추적과 노동자 과잉 모니터링

아마존이 특허를 하나 냈습니다. 손목에 차는 밴드였습니다.

물류 창고에서 일하는 사람이 이 밴드를 차면, 손이 어느 선반 어느 상자로 가는지 실시간으로 추적됩니다. 손이 엉뚱한 곳으로 가면 밴드가 진동합니다.

부드럽게 알려 주는 겁니다. 거기 아니에요, 이쪽이에요.

강아지 훈련용 목걸이를 떠올린 사람은 저만이 아니었습니다.

이 장치의 설계에는 사람에 대한 특정한 관점이 들어 있습니다. 노동자는 시스템의 말단 장치이고, 손은 그 장치의 집게이며, 진동은 오차를 줄이는 피드백이라는 관점입니다.

아마존 프랑스 법인은 실제로 벌금을 맞았습니다. 삼천이백만 유로입니다.

이유는 이랬습니다. 창고 노동자가 바코드 스캐너를 들고 일할 때 스캔과 스캔 사이 시간을 초 단위로 재고 있었습니다. 십 분 넘게 스캔이 없으면 경고가 쌓입니다.

십 분입니다. 화장실에 다녀오면 십 분이 갑니다. 물을 마시고 숨을 돌리면 십 분이 갑니다.

프랑스 데이터보호청은 이런 감시가 지나치다고 봤습니다.

노동자를 재기 전에, 회사들은 먼저 위치를 파는 법을 배웠습니다.

2021년 라이프360이라는 회사가 도마에 올랐습니다. 가족끼리 서로 어디 있는지 확인하는 앱입니다. 아이가 학교에 잘 도착했는지 보려고 부모들이 깔았습니다.

이 회사는 이용자 수천만 명의 이동 경로와 방문 장소 기록을 데이터 중개업체들에 팔아 왔습니다.

위치 기록은 다른 정보와 성질이 다릅니다. 어디에 갔는지를 알면 그 사람에 대해 거의 모든 것을 알게 됩니다. 매주 목요일 저녁에 어느 건물에 가는지, 어느 병원에 가는지, 어느 종교 시설에 가는지, 어느 날 밤 집이 아닌 곳에 있었는지.

이름을 지워도 소용없습니다. 밤마다 머무는 곳이 집이고, 낮마다 머무는 곳이 직장입니다. 두 점만 있으면 사람은 특정됩니다.

위치 데이터 중개업체 타모코는 노르웨이 시민들의 GPS 기록을 모아 수사기관과 군사 추적용으로 유통하다 적발됐습니다. 영국 기업 후크도 앱으로 모은 위치 데이터를 제삼자에게 팔다가 제재를 받았습니다.

중국의 배달 플랫폼 메이퇀은 배달 기사와 이용자를 과도하게 실시간 추적해 반발을 샀습니다.

미국 수사 당국은 자동 번호판 인식 카메라 시스템으로 영장 없이 시위 참가자와 활동가들의 차량 이동 경로를 추적해 감시망에 넣었습니다.

번호판은 가릴 수 없습니다. 법으로 붙이라고 정해 놓았기 때문입니다.

배달과 운송 현장에서는 카메라가 얼굴을 봅니다.

아마존 배달 위탁 차량에는 인공지능 영상 감시 카메라와 운전자 평가 프로그램이 달렸습니다. 시선이 어디를 향하는지, 하품을 하는지, 얼굴 근육이 어떻게 움직이는지, 얼마나 세게 밟고 얼마나 급하게 서는지를 봅니다.

문제는 이 시스템이 정상적인 운전을 벌한다는 데 있었습니다.

나뭇가지를 피하려고 핸들을 돌리면 부주의로 잡혔습니다. 룸미러를 보면 시선 이탈로 잡혔습니다. 운전 교습소에서 하라고 가르치는 행동입니다.

점수가 깎이면 수당이 깎이고 배차가 줄어듭니다. 기사는 안전 운전과 점수 사이에서 골라야 합니다.

아마존 플렉스 알고리즘은 도로 공사나 궂은 날씨를 계산에 넣지 않고 수학적으로 제일 짧은 길을 요구했습니다. 기사들이 위험한 길과 진흙탕으로 들어갔습니다.

지도 위의 선은 짧습니다. 그 선 위에 무엇이 있는지는 지도에 없습니다.

사무실도 안전지대가 아니었습니다.

금융기업 바클레이스는 영국 사옥에서 직원 책상 아래에 센서를 붙이고 업무 컴퓨터에 감시 프로그램을 몰래 깔았습니다. 자리에 앉아 있는 시간과 키보드를 두드리는 빈도를 초 단위로 재서 비생산적 시간을 집계했습니다.

노동조합이 항의하자 회사는 운용을 중단했습니다.

마이크로소프트는 생산성 점수라는 기능을 내놨습니다. 이메일을 몇 번 보냈는지, 메신저에 얼마나 빨리 답했는지, 회의에 얼마나 참여했는지로 개인별 점수를 매겨 관리자에게 보여 줍니다.

이런 지표에는 공통점이 있습니다. 세기 쉬운 것만 셉니다.

한 시간 동안 창밖을 보며 문제를 푸는 사람과, 한 시간 동안 메신저에 스물세 번 답하는 사람 중에 누가 일을 한 걸까요. 점수는 후자를 고릅니다.

그래서 사람들은 후자를 하기 시작합니다. 점수가 행동을 바꾸고, 바뀐 행동이 다시 점수를 만듭니다.

메타도 미국 직원들의 업무 컴퓨터에 소프트웨어를 몰래 깔아 마우스 움직임, 클릭, 키보드 입력, 화면 캡처를 실시간으로 모으다가 적발됐습니다.

감정까지 재는 시스템도 있습니다.

글로벌 콜센터 기업 텔레퍼포먼스는 재택근무 노동자의 웹캠을 근무 시간 내내 켜 두는 시스템을 도입했습니다. 인공지능이 얼굴 각도와 눈동자 위치를 추적해서 화면에서 눈을 떼는지, 휴대폰을 보는지, 방에 다른 사람이 들어오는지를 감지해 회사에 알립니다.

자기 집 거실이 근무 시간 내내 촬영되는 겁니다. 가족이 지나가면 경고가 갑니다.

중국에 있는 캐논 사무실에서는 카메라 앞에서 웃어야 문이 열리는 시스템을 썼습니다. 미소가 인식돼야 출근이 확인됩니다.

회사는 밝은 분위기를 만들려는 목적이라고 했습니다.

웃어야 들어갈 수 있는 문 앞에서 짓는 웃음이 어떤 웃음인지는 다들 아실 겁니다.

회계법인 프라이스워터하우스쿠퍼스는 금융 트레이더가 화장실에 가느라 자리를 비우는 시간까지 안면 인식으로 추적하는 도구를 도입했다가 논란을 맞았습니다.

스페인의 제조업체 플라스틱 포르테는 현장 노동자의 출퇴근과 이동을 안면 인식으로 감시하다 벌금을 받았습니다. 영국 서비스 기업 서코도 직원 감시에 안면 인식을 불법으로 쓰다가 사용 중단 명령을 받았습니다.

목소리 톤, 심박수, 체온, 피부 전기 반응까지 재서 감정 상태를 추적하려는 시도들도 있었습니다. 어떤 시스템은 공장 바닥 전체에 카메라를 깔아 노동자들의 동선과 동작을 봤습니다.

이렇게 모은 숫자는 결국 사람을 자릅니다.

2021년 게임 플랫폼 기업 엑스솔라는 감시 알고리즘 분석 결과를 근거로 직원 백오십 명을 한꺼번에 해고했습니다.

기준은 이랬습니다. 사내 프로그램 이용 이력, 코드 저장소와 문서 입력 빈도, 채팅방과 문서 작업 참여도. 이 지표에서 낮게 나온 사람들이 비활성 직원으로 분류됐습니다.

직원들은 자기가 어떤 기준으로 평가됐는지 몰랐습니다. 어떤 업무 맥락이 빠졌는지도 몰랐습니다. 오래 일한 회사에서 자동 통보 한 줄로 나왔습니다.

전화로 고객과 오래 통화한 사람, 후배를 붙잡고 가르친 사람, 회의실에서 말로 문제를 푼 사람. 이 사람들이 어느 칸에 기록되는지 생각해 보십시오. 어느 칸에도 기록되지 않습니다.

한 회사의 이직 예측 알고리즘은 직원들의 검색 이력과 플랫폼 활동을 추적해서 나갈 것 같은 사람을 미리 찾아내고 불이익을 줬습니다.

떠날 것 같다는 예측 때문에 나쁜 대우를 받으면 정말 떠나게 됩니다. 예측은 맞습니다.

플랫폼 노동에서는 자르는 일조차 자동입니다.

아마존 플렉스 배달 기사들은 사람 얼굴 한 번 보지 못하고 자동 이메일 한 통으로 해고됐습니다. 앱이 오작동해도, 물류 센터에서 물건이 늦게 나와도, 도로가 막혀도, 폭우가 쏟아져도 배달 지연 점수로 쌓였습니다. 점수가 쌓이면 계정이 영구 정지됩니다.

이의를 제기하면 미리 써 둔 답장이 왔습니다.

이탈리아 볼로냐 법원이 문제 삼은 딜리버루의 프랭크 알고리즘은 앞서 나왔습니다. 아파서 못 나가도, 가족상을 당해도, 법이 보장한 파업에 참여해도 점수가 깎였습니다.

법은 파업할 권리를 줍니다. 알고리즘은 그 권리를 쓰면 굶게 만듭니다. 둘은 같은 나라 안에서 함께 돌아갔습니다.

얼굴을 못 알아봐서 잘린 사람들도 있습니다.

영국 우버 이츠 배달원이던 파 에드리사 만장과 임란 라자는 플랫폼의 자동 안면 확인 소프트웨어 때문에 하루아침에 해고됐습니다.

시스템은 피부색이 어두운 배달원의 얼굴을 잘 인식하지 못했습니다. 인식에 실패하면 계정을 남에게 빌려준 부정 사용자로 판정합니다. 계정이 즉시 얼어붙습니다.

두 사람은 몇 해 동안 좋은 평가를 받으며 일했습니다. 기계가 자기 얼굴을 못 알아본 대가를 본인이 치렀습니다.

유타주와 뉴사우스웨일스를 비롯한 여러 곳에서는 성전환 수술이나 호르몬 치료 뒤 얼굴이 바뀐 트랜스젠더 기사들의 계정이 무더기로 정지됐습니다. 시스템은 사람이 변한다는 사실을 계산에 넣지 않았습니다.

인스타카트와 십트 같은 배달 플랫폼들도 불투명한 알고리즘으로 임금을 깎고 노동자를 자동 해고했습니다. 미국 우정청의 배달원 임금 산정 알고리즘도 수식이 공개되지 않은 채 수당을 깎았습니다.

제일 씁쓸한 사례는 마지막입니다.

2025년 호주의 한 대형 은행에서 직원들이 몇 달 동안 고객 서비스 챗봇을 가르쳤습니다. 자기가 쌓은 응대 노하우를 입력하고, 어려운 상황에 어떻게 답하는지 알려 주고, 챗봇이 틀리면 고쳐 줬습니다.

챗봇 성능이 올라오자 회사는 그 직원들에게 집단 해고를 통보했습니다.

자기 후임을 자기 손으로 가르친 겁니다.

이 절에 나온 도구들을 늘어놓아 보겠습니다. 손목 밴드, 바코드 스캐너 타이머, 차량 카메라, 책상 센서, 키보드 기록기, 웹캠, 미소 인식 문, 화장실 추적 카메라, 이직 예측 모델.

전부 사람을 향해 있습니다. 그리고 전부 같은 이름으로 불립니다. 생산성 관리입니다.

노동자는 자기 점수를 볼 수 없습니다. 어떻게 계산되는지도 모릅니다. 물어보면 영업 비밀이라고 합니다.

한 가지는 확실합니다. 이 시스템들은 사람이 쉬는 시간을 찾아내는 데 아주 능숙합니다. 쉬는 시간은 세기 쉽거든요.

일이 잘되는 순간은 세기 어렵습니다. 그래서 세지 않습니다.

제 4 장

저작권과 창작자의 권리: 생성형 AI와 무단 학습 갈등

4.1

저작권 침해 소송과 무단 데이터 스크래핑

먼저 책을 삽니다. 진짜 책입니다. 값을 치르고 삽니다.

다음에 제본을 자릅니다. 등을 칼로 잘라 내면 책이 낱장 뭉치가 됩니다. 그 낱장을 자동 급지 스캐너에 넣습니다. 종이가 한 장씩 넘어가면서 디지털 파일이 됩니다.

스캔이 끝난 종이는 버립니다.

인공지능 기업 앤스로픽 내부에서 이렇게 처리한 책이 수십만 권입니다. 프로젝트 파나마라는 이름이 붙어 있었습니다.

법적으로는 영리한 방법입니다. 산 책은 소유물이니 잘라도 됩니다. 그런데 이 장면을 머릿속에 그려 보면 마음이 이상해집니다. 누군가 평생 걸려 쓴 책이 컨베이어 위에서 스무 초 만에 처리되고 쓰레기통으로 갑니다.

앤스로픽은 이 방법만 쓴 것이 아니었습니다. 립젠과 피리미 같은 해적 도서관에서 오십만 권 넘는 책을 그냥 내려받기도 했습니다.

작가 안드레아 바르츠, 커크 월리스 존슨, 찰스 그레이버가 낸 집단소송에서 2025년 앤스로픽은 십오억 달러를 내기로 합의했습니다. 미국 저작권 소송 역사상 제일 큰 금액입니다.

이 사건에서 법원이 어디에 선을 그었는지가 중요합니다. 법원은 저작물로 언어 모델을 학습시키는 행위 자체는 불법으로 보지 않았습니다. 문제 삼은 것은 해적판을 내려받아 보관한 부분이었습니다.

배우는 것은 되고, 훔쳐서 배우는 것은 안 된다는 뜻입니다. 이 구별은 앞으로 나올 모든 소송의 뼈대가 됩니다.

작가들이 처음 나선 것은 2023년입니다.

그해 6월 소설가 모나 아와드와 폴 트렘블레이가 오픈AI를 상대로 첫 소송을 냈습니다. 9월에는 미국 작가조합 소속 열일곱 명이 뉴욕 연방법원에 집단소송을 냈습니다. 왕좌의 게임

을 쓴 조지 마틴, 존 그리셤, 조디 피콜트, 스콧 터로, 데이비드 발다치, 조너선 프랜즌 같은 이름들이었습니다.

이들의 주장은 이랬습니다. 오픈AI가 허락 없이 소설 수만 권을 학습에 썼다. 그 재료는 라이브러리 제네시스나 제트라이브러리 같은 해적 도서관, 그리고 북스3이라는 데이터셋에서 왔다.

챗GPT가 유려한 문장을 쓰는 이유는 유려한 문장을 많이 읽었기 때문입니다. 그 문장들은 누군가 썼습니다.

퓰리처상을 받은 마이클 셰이본, 극작가 데이비드 헨리 황, 논픽션 작가 레이철 루이즈 스나이더도 오픈AI와 메타를 상대로 나섰습니다.

2023년 11월에는 작가 줄리언 생턴이 오픈AI와 마이크로소프트를 상대로 논픽션 수만 권의 무단 사용을 주장하는 집단소송을 냈습니다. 마이크 허커비를 비롯한 종교 서적 저자들은 북스3 데이터셋에 담긴 십팔만 권이 학습에 쓰였다며 메타와 마이크로소프트 등을 걸었습니다.

2024년 8월에는 저널리스트 니컬러스 배스베인스와 니컬러스 게이지가 북스2라는 다른 데이터셋을 문제 삼아 소송에 합류했습니다.

2025년 10월에는 뉴욕주립대 다운스테이트 보건과학센터의 수사나 마르티네스콘데와 스티븐 맥닉 교수가 애플을 상대로 소송을 냈습니다. 애플 인텔리전스 학습에 해적 도서관 데이터가 쓰였다는 주장입니다. 엔비디아도 작가 세 명에게 같은 이유로 소송을 당했습니다.

언론사들은 다른 각도에서 싸웁니다.

2023년 12월 뉴욕타임스가 오픈AI와 마이크로소프트를 상대로 수십억 달러 규모의 소송을 냈습니다. 수백만 건의 기사가 허락도 대가도 없이 학습에 쓰였다는 겁니다.

이 소송의 핵심은 학습이 아니라 경쟁입니다. 신문사가 돈과 시간을 들여 취재한 결과로 만든 서비스가, 그 신문사와 독자를 두고 경쟁합니다. 독자가 요약만 읽고 기사에 오지 않으면 신문사는 광고 수익을 잃습니다.

취재에는 돈이 듭니다. 기자를 현장에 보내야 하고, 자료를 청구해야 하고, 취재원을 여러 번 만나야 합니다. 요약에는 그런 비용이 들지 않습니다.

2026년에도 이 싸움은 진행 중입니다. 뉴욕타임스와 여러 언론사는 오픈AI가 자사 시스템 안에서 저작권 자료를 찾아낼 수 있는 능력에 대해 법원을 오도했다며 맨해튼 연방판사에게 제재를 요청했습니다.

2024년 4월에는 시카고 트리뷴과 덴버 포스트를 비롯한 미국 일간지 여덟 곳이 소장을 냈습니다. 기사를 무단 수집했을 뿐 아니라 저작권 표기를 일부러 지웠다는 주장이 들어 있었습니다.

출처를 지우는 행위는 실수로 보기 어렵습니다. 코드에 그렇게 적어야 지워집니다.

2024년 2월 디 인터셉트와 로 스토리와 알터넷이, 6월에는 탐사보도센터가 소송을 냈습니다. 11월에는 캘거리 헤럴드와 몬트리올 가제트를 가진 캐나다 출판사들이 나섰습니다.

인공지능 검색 서비스 퍼플렉시티도 표적이 됐습니다. 2024년 10월 다우존스와 뉴욕 포스트가, 2025년 12월 시카고 트리뷴이 소송을 냈습니다. 콘데 나스트와 뉴욕타임스는 경고장을 보냈습니다.

프랑스 당국은 2024년 3월 구글에 이억 오천만 유로의 과징금을 물렸습니다. 언론사들과 상의 없이 제미나이 학습에 뉴스 기사를 썼다는 이유였습니다.

한국에서도 2024년과 2025년에 방송사들이 네이버를 상대로 소송을 냈습니다. 방송 뉴스 콘텐츠가 하이퍼클로바 X 학습에 무단으로 쓰였다는 주장입니다.

그림 쪽에서는 증거가 화면에 그대로 찍혀 나왔습니다.

2023년 1월 게티 이미지가 스태빌리티 AI를 상대로 영국과 미국에서 소송을 냈습니다. 워터마크가 박힌 자사 사진 천이백만 장 이상이 라이선스 없이 학습에 쓰였다는 겁니다.

증거는 이랬습니다. 스테이블 디퓨전이 만들어 낸 그림 구석에 게티 이미지의 워터마크가 일그러진 모양으로 남아 있었습니다.

기계는 워터마크가 무엇인지 모릅니다. 사진에 자주 나오는 무늬라서 같이 배운 겁니다. 도둑이 훔친 물건에 가게 이름표를 그대로 달고 나온 셈입니다.

독일의 스톡 사진가 로베르트 크네슈케는 자기 사진이 라이온 데이터셋을 통해 학습에 쓰인 것을 확인하고 소송을 진행했습니다. 일러스트레이터 홀리 멩거트는 자신의 화풍이 통째로

복제돼 모델이 되는 일을 겪었습니다.

화풍은 저작권으로 지키기 어렵습니다. 법이 보호하는 것은 개별 작품이지 스타일이 아니기 때문입니다. 그런데 인공지능이 훔치는 것은 정확히 스타일입니다.

법이 그은 선과 기술이 가져가는 것이 어긋나 있습니다.

거대 미디어 기업들도 나섰습니다.

2025년 디즈니와 유니버설과 워너 브라더스 디스커버리가 이미지 생성 기업 미드저니를 상대로 집단소송을 냈습니다. 자사 애니메이션과 영화 속 캐릭터가 무단으로 수집돼 합성 이미지로 나온다는 겁니다. 같은 기업들이 중국의 미니맥스도 걸었습니다.

영화사 알콘 엔터테인먼트는 2024년 10월 테슬라와 일론 머스크를 상대로 소송을 냈습니다. 로보택시 공개 행사에서 블레이드 러너 2049의 시각 이미지를 인공지능으로 재현해 썼다는 이유였습니다.

중국 법원은 2024년 울트라맨 캐릭터를 무단 학습해 생성한 기업에 저작권 침해 판결을 내렸습니다.

음악은 소리로 남습니다.

2024년 6월 소니 뮤직과 유니버설 뮤직 그룹과 워너 뮤직이 음악 생성 기업 수노와 우디오를 상대로 소송을 냈습니다. 수십 년간 쌓인 상업 음원을 무단으로 긁어 학습시켰고, 그 결과 특정 가수의 음색과 스타일을 정교하게 복제하는 기계가 나왔다는 주장입니다.

2023년 10월 유니버설 뮤직 그룹은 앤스로픽을 상대로 가사 데이터베이스 무단 수집 소송을 냈습니다. 클로드가 유명 노래 가사를 그대로 출력했기 때문입니다. 이 소송에서 앤스로픽 쪽이 낸 서류에 존재하지 않는 학술 논문이 인용됐던 사건은 앞에서 나왔습니다.

코미디언 조지 칼린은 2008년에 세상을 떠났습니다. 2024년 1월 그의 유족은 칼린의 생전 음성을 학습시켜 인공지능 코미디 특집을 만든 채널을 상대로 소송을 내 삭제와 합의를 이끌어 냈습니다.

죽은 사람은 새로운 농담을 하지 않습니다. 그런데 그 목소리는 계속 새로운 농담을 했습니다.

목소리로 먹고사는 사람들의 이야기는 더 직접적입니다.

베테랑 성우 폴 스카이 리먼과 리네아 세이지는 2024년 5월 음성 생성 스타트업을 상대로 뉴욕 연방법원에 소송을 냈습니다. 회사가 대화 연구 목적이라며 목소리를 녹음해 놓고, 그 녹음을 상업용 인공지능 성우 서비스의 학습 데이터로 쓰고 팔았다는 겁니다.

성우 베브 스탠딩은 자기 목소리가 틱톡의 자동 읽기 음성으로 쓰이는 것을 알고 소송을 내 합의를 받아 냈습니다. 자기가 하지 않은 말을 자기 목소리로 수억 명이 듣고 있었던 겁니다.

배우 스칼렛 요한슨은 2024년 5월 오픈AI가 만든 인공지능 목소리가 자기 목소리를 그대로 흉내 냈다고 항의했습니다. 오픈AI는 그 음성 서비스를 급히 중단했습니다.

요한슨은 자신이 나오는 인공지능 광고를 무단으로 만든 앱 개발사에도 법적 대응에 나섰습니다. 배우 브라이언 크랜스턴도 자기 음성과 얼굴이 동의 없이 영상 모델 학습에 쓰였다고 밝혔습니다.

전문 데이터베이스도 털렸습니다.

캐나다의 법률 데이터베이스 캔리는 2024년 11월 한 법률 인공지능 스타트업을 상대로 소송을 냈습니다. 이용약관과 접근 제한을 무시하고 판결문과 해석 문서 수십만 건을 긁어 갔다는 겁니다.

법률 정보 회사 톰슨 로이터는 2025년 1월 자사 데이터베이스를 무단으로 긁어 학습한 로스 인텔리전스를 상대로 승소했습니다.

이 사건들에서 기업들이 내놓는 방어 논리는 대체로 같습니다. 공정 이용입니다. 학습은 복제가 아니라 분석이고, 사람이 책을 읽고 배우는 것과 다르지 않다는 겁니다.

이 비유에는 빠진 것이 있습니다. 사람은 책을 한 권씩 읽고, 읽는 데 시간이 걸리고, 읽은 책을 사거나 빌립니다. 그리고 사람 한 명이 읽어서 얻는 능력은 그 사람 한 명의 것입니다.

모델은 오십만 권을 며칠 만에 읽고, 그 능력을 수억 명에게 동시에 팝니다.

규모가 달라지면 성질이 달라집니다. 물 한 컵을 뜨는 것과 강을 통째로 끌어가는 것은 같은 행위가 아닙니다.

한 가지는 분명해졌습니다. 이 재판들이 어떻게 끝나든, 이미 학습된 모델에서 특정 작가의 흔적만 골라 지우는 방법은 아직 없습니다.

책은 잘려 나갔고, 스캔은 끝났고, 종이는 버려졌습니다.

4.2

AI 생성물의 저작권 부인과 예술 생태계 혼란

크리스 카슈타노바는 2022년 가을에 그래픽 노블 한 권을 만들었습니다. 제목은 조리야 오브 더 돈입니다.

이야기는 직접 썼습니다. 그림은 미드저니에 명령어를 넣어 뽑았습니다. 마음에 드는 그림이 나올 때까지 명령어를 고치고 또 고쳤습니다.

카슈타노바는 미국 저작권청에 등록을 신청했고 승인을 받았습니다.

그런데 그림을 인공지능으로 만들었다는 사실이 알려지자 저작권청이 다시 들여다봤습니다. 2023년 2월에 결정이 나왔습니다.

직접 쓴 글과 글을 배치한 방식은 저작권으로 보호합니다. 미드저니가 뽑아 낸 그림은 등록을 취소합니다.

저작권청의 논리는 이랬습니다. 명령어를 아무리 정성껏 써도, 그 결과가 어떻게 나올지 미리 알 수 없고 세밀하게 조절할 수도 없다. 화가가 붓으로 선 하나를 긋는 것과는 다르다.

이 결정이 뜻하는 바는 무겁습니다. 인공지능이 만든 그림은 누구의 것도 아닙니다. 만든 사람도 소유하지 못합니다. 아무나 가져다 써도 됩니다.

기업들이 인공지능 그림을 쓰는 이유는 대개 비용입니다. 그런데 그렇게 만든 표지나 광고 이미지는 법적으로 자기 것이 아닙니다. 경쟁사가 그대로 가져다 써도 막을 근거가 마땅치 않습니다.

싸게 만들었는데 지킬 수도 없는 겁니다.

저작자가 누구인지를 두고 다툰 일은 그전에도 있었습니다.

2018년 뉴욕 크리스티 경매에서 인공지능이 그린 초상화가 사십삼만 이천오백 달러에 팔렸습니다. 에드몽 드 벨라미라는 제목이었고, 프랑스 예술 집단 오비어스가 내놓았습니다.

나중에 밝혀진 사실이 있습니다. 그림을 만든 코드는 다른 개발자가 공개해 둔 것이었습니다. 그러면 이 그림의 작가는 명령어를 넣은 사람일까요, 코드를 쓴 사람일까요, 아니면 학습에 쓰인 초상화들을 그린 옛 화가들일까요.

경매 도록에는 알고리즘 수식이 서명 자리에 적혀 있었습니다. 낙찰금은 사람이 받았습니다.

네덜란드 마우리츠하위스 미술관이 진주 귀걸이를 한 소녀의 인공지능 모작을 전시했을 때도 비판이 쏟아졌습니다. 원작이 있는 자리에 기계 모작을 걸었기 때문입니다.

여기까지가 법과 미술관의 이야기입니다. 실제 싸움은 책 표지에서 터졌습니다.

출판사 토르 북스는 베스트셀러 작가 크리스토퍼 파올리니의 신작 표지에 인공지능 이미지를 사서 썼습니다. 사실이 알려지자 독자와 일러스트레이터들이 항의했습니다.

출판사 블룸스버리도 인기 작가 세라 마스의 소설 표지에 인공지능 이미지를 썼다가 비판을 받았습니다. DC 코믹스는 인공지능으로 만든 변형 표지들을 내놓았다가 소속 작가들의 반발로 거둬들였습니다.

표지 그림은 일러스트레이터에게 큰 일감입니다. 한 장에 몇 주가 걸리고, 그 수입으로 몇 달을 삽니다. 표지를 인공지능이 맡으면 그 일자리가 사라집니다.

그리고 그 인공지능은 그 일러스트레이터들의 그림으로 배웠습니다.

제일 어색한 사건은 태블릿 회사에서 나왔습니다.

와콤은 그림 그리는 사람들이 쓰는 드로잉 태블릿을 만듭니다. 고객이 곧 일러스트레이터입니다. 2024년 용의 해를 맞아 미국 시장용 광고를 만들면서, 용이 그려진 인공지능 생성 이미지를 썼습니다.

자기 물건을 사 주는 사람들의 일을 대신하는 그림으로 광고를 한 겁니다.

회사는 외주 대행사가 만든 이미지라고 해명하고 사과했습니다. 그림 그리는 사람들 사이에서 이 회사의 이름은 한동안 다르게 불렸습니다.

브래드퍼드 문학 축제는 홍보 자료에 인공지능 이미지를 썼다가 참가 예정이던 일러스트레이터가 참여를 거부했습니다.

게임 업계에서는 약속이 두 번 깨졌습니다.

던전 앤 드래곤을 만드는 위저즈 오브 더 코스트는 신작 책 삽화에 인공지능 도구로 만든 그림이 들어간 사실이 드러나 비판을 받았습니다. 회사는 인공지능 이미지를 쓰지 않겠다고 지침을 바꿨습니다.

그리고 이어진 홍보물에 또 인공지능 그림을 썼습니다.

콜 오브 듀티를 만드는 액티비전은 게임 안에서 파는 꾸미기 아이템을 인공지능으로 만들어 유료로 팔았습니다. 레고는 라이선스를 받지 않은 캐릭터를 인공지능으로 합성해 홍보물에 썼다가 항의를 받았습니다.

포켓몬스터를 닮은 게임 팔월드는 인공지능으로 캐릭터 형태를 베꼈다는 의혹에 휩싸였습니다.

기술 기업 수장들의 선택도 도마에 올랐습니다.

메타의 마크 저커버그는 어린 딸들을 위해 동화책을 만들면서 삽화를 자사 인공지능 이미지 생성기로 그리게 하고 공개했습니다.

세계에서 손꼽히는 부자가 자기 아이에게 줄 그림책의 그림을 사람에게 맡기지 않았습니다. 그림값을 아낀 것은 아닐 겁니다. 그냥 그 편이 빨랐을 겁니다.

빠른 편을 고르는 순간이 쌓여서 산업이 바뀝니다.

아마존은 드라마 폴아웃의 홍보 포스터를 인공지능으로 합성해 만들었다가 조악하다는 비판을 받았습니다.

영화 쪽에서는 오류가 그대로 인쇄됐습니다.

영화 시빌 워를 만든 A24는 미국 주요 도시가 파괴된 모습의 홍보 포스터를 인공지능으로 만들어 공개했습니다. 포스터 속 랜드마크는 위치가 뒤틀려 있었고, 사람의 몸은 관절이 이상하게 꺾여 있었습니다.

게다가 그 장면들은 영화에 나오지 않았습니다.

넷플릭스는 애니메이션 개와 소년에서 배경 그림을 인공지능으로 처리하고 크레디트에 인공지능 개발자라고 적었습니다. 배경 그림은 애니메이션 업계에서 신인들이 실력을 쌓는 자리입니다. 그 자리가 사라지면 다음 세대가 자라지 못합니다.

애니메이션 아케인 시즌 2의 홍보 포스터도 인공지능으로 만들어져 인체 구조의 오류를 그대로 드러냈습니다. 마블 시리즈 시크릿 인베이전의 오프닝 영상도 인공지능으로 만들어져 오프닝 디자이너들의 반발을 샀습니다.

광고에서도 같은 일이 있었습니다.

토이저러스는 오픈AI의 영상 모델로 창업자의 어린 시절을 다룬 브랜드 영상을 만들었습니다. 영상 속 아이의 표정이 어딘가 어긋나 있었고 몸짓이 미끄러졌습니다. 보는 사람들이 불편함을 느꼈습니다.

사람의 얼굴을 거의 정확하게 흉내 낸 것이 완전히 다른 것보다 더 불편하게 느껴지는 현상이 있습니다. 인형이나 마네킹을 볼 때의 느낌과 비슷합니다. 우리 눈은 사람 얼굴을 아주 정밀하게 읽도록 만들어져 있어서, 조금만 어긋나도 경보를 울립니다.

코카콜라는 오래된 성탄절 트럭 광고를 인공지능 영상으로 다시 만들었다가 영혼이 빠졌다는 말을 들었습니다. 같은 캠페인에서 소설가 제임스 그레이엄 밸러드가 하지 않은 말을 인공지능으로 만들어 넣었다가 유족과 문학계의 항의를 받고 사과했습니다.

의류 브랜드 언더아머는 인공지능으로 만든 광고 영상이 기존 감독의 영상을 그대로 베꼈다는 표절 논란에 휘말렸습니다.

패션에서는 복제 속도가 무기가 됐습니다.

초고속 패션 기업 쉬인은 알고리즘으로 소셜미디어에 올라온 독립 디자이너들의 옷 디자인과 무늬를 실시간으로 긁어모았습니다. 그리고 자사 공장에서 바로 만들어 싸게 팔았습니다.

독립 디자이너가 새 디자인을 공개하면 며칠 뒤에 같은 옷이 절반 값에 나옵니다. 원작자는 자기 옷을 팔지 못합니다. 피해를 입은 디자이너들이 연방법원에 소송을 냈습니다.

리바이스는 인공지능으로 만든 다양한 인종의 가상 모델을 화보에 쓰겠다고 발표했다가 비판을 받았습니다. 다양성을 갖추려면 다양한 사람을 고용하면 됩니다. 인공지능 모델은 고용 없이 사진만 얻는 방법입니다.

일러스트레이터 홀리 멩거트의 그림을 동의 없이 학습시켜 모델을 만든 사람은 이렇게 말했습니다. 도덕적 기준은 주관적인 선일 뿐이다.

가상의 사람을 배우라고 부르는 시도도 있었습니다.

인공지능 여배우라는 이름의 캐릭터가 등장했고, 인도 자동차 기업은 가상 인플루언서를 도입했고, 가상 래퍼가 상업적으로 데뷔했습니다. 실제로 연기하고 노래하는 사람들의 자리가 그만큼 줄어듭니다.

스포티파이는 음원 사용료 지출을 줄이려고 인공지능으로 만든 가짜 아티스트들의 음악을 재생 목록 위쪽에 대량으로 넣었습니다. 이용자는 배경음악으로 틀어 놓고 무엇을 듣는지 신경 쓰지 않습니다. 그 몇 분마다 진짜 음악가에게 갈 돈이 조금씩 줄어듭니다.

게임 회사들이 성우를 인공지능 음성으로 바꾸려다 팬들과 성우 조합의 불매 운동에 부딪히기도 했습니다.

이 절에 나온 사건들을 다시 보면 흐름이 하나 있습니다.

기업이 인공지능 이미지를 씁니다. 사람들이 알아차립니다. 항의가 옵니다. 기업이 해명하거나 사과합니다. 다음 분기에 다시 씁니다.

그런데 알아차리는 사람들이 누구인지가 중요합니다. 대개 그 분야를 사랑하는 팬들과 그 일을 하는 창작자들입니다. 표지 그림의 손가락 개수를 세어 보는 사람들입니다.

한 가지 아이러니가 남습니다. 인공지능 그림에는 저작권이 없습니다. 그러니까 그 그림으로 만든 표지도, 광고도, 게임 아이템도 누구의 것이 아닙니다.

지킬 수 없는 것을 만들면서, 지켜야 할 것을 쳐 내고 있습니다.

4.3

초상권·음성 클로닝과 무단 복제

2024년 5월, 오픈AI가 새 음성 서비스를 공개했습니다. 목소리 이름은 스카이였습니다.

들어 본 사람들이 같은 생각을 했습니다. 이거 그 목소리 아닌가.

영화 허에서 스칼렛 요한슨은 화면에 얼굴 한 번 나오지 않고 목소리로만 인공지능을 연기했습니다. 그 영화를 본 사람에게 그 음색과 호흡은 잊기 어렵습니다.

앞뒤 사정은 이랬습니다. 오픈AI 최고경영자 샘 올트먼은 서비스를 내기 전에 요한슨에게 직접 연락해 목소리를 맡아 달라고 했습니다. 요한슨은 거절했습니다.

그리고 비슷한 목소리가 나왔습니다.

공개 당일 올트먼은 소셜미디어에 한 단어를 올렸습니다. her.

세 글자입니다. 그 세 글자가 나중에 법정에서 어떤 무게를 갖는지는 변호사들이 판단할 몫입니다. 다만 사람들이 무슨 뜻으로 읽었는지는 분명했습니다.

요한슨이 법적 대응을 예고하자 오픈AI는 스카이 음성을 급히 내렸습니다.

이 사건이 남긴 질문은 이렇습니다. 목소리는 누구의 것입니까.

법은 이 질문에 아직 또렷하게 답하지 못합니다. 저작권은 작품을 보호합니다. 목소리는 작품이 아니라 몸의 특징입니다. 초상권은 얼굴을 보호합니다. 목소리는 얼굴이 아닙니다.

그래서 흉내는 처벌하기 어렵습니다. 성대모사는 오랫동안 예능이었습니다.

기계가 하는 흉내는 다릅니다. 지치지 않고, 몇 초짜리 샘플이면 되고, 하루에 수만 개를 만들 수 있습니다.

2026년 2월에는 미국 공영라디오 NPR의 베테랑 앵커 데이비드 그린이 캘리포니아 산타클라라 카운티 법원에 구글을 상대로 소송을 냈습니다. 구글이 팟캐스트와 문서 요약 서비스를 만들면서 자기 음성 특성과 억양을 동의나 보상 없이 학습시켜 인공지능 앵커 목소리로

만들었다는 주장입니다.

그린이 걱정한 것은 돈이 아니었습니다. 자기 목소리가 자기가 결코 하지 않을 말을 할 수 있다는 점이었습니다.

방송 앵커에게 목소리는 신뢰의 표식입니다. 그 목소리가 하는 말이면 사실이겠거니 하고 사람들이 듣습니다. 그 신뢰는 수십 년 걸려 쌓입니다.

오픈AI의 영상 생성 도구 소라도 배우 브라이언 크랜스턴의 목소리와 얼굴을 동의 없이 학습시켰다는 고발을 받았습니다.

세상을 떠난 사람들의 목소리는 유난히 자주 쓰입니다. 항의할 사람이 없기 때문입니다.

2023년 10월 배우 로빈 윌리엄스의 딸 젤다 윌리엄스는 아버지를 복제한 인공지능 음성과 영상들을 보고 이렇게 말했습니다. 끔찍하고 기괴하다.

아버지가 하지 않은 말을 아버지 목소리로 듣는 일이 어떤 느낌인지, 겪어 보지 않은 사람은 짐작만 할 수 있습니다.

2021년 다큐멘터리 로드러너에서는 감독이 세상을 떠난 요리사 앤서니 보데인의 생전 이메일 문장을 인공지능 음성으로 읽게 해 영화에 넣었습니다. 보데인이 쓴 문장이기는 했습니다. 소리 내어 말한 적은 없는 문장이었습니다.

글로 쓴 말과 소리 내어 한 말은 다릅니다. 다큐멘터리에서 그 차이는 사실과 연출의 경계입니다.

한국에서도 세상을 떠난 가수 김광석의 목소리를 알고리즘으로 되살려 신곡을 부르게 한 시도가 있었습니다. 중국에서는 배우와 가수의 생전 모습을 인공지능으로 부활시킨 영상들이 퍼졌습니다.

대중음악에서는 살아 있는 가수의 목소리가 먼저 복제됐습니다.

2023년 4월 고스트라이터라는 이름의 무명 창작자가 하트 온 마이 슬리브라는 곡을 냈습니다. 드레이크와 더 위켄드의 목소리를 정교하게 복제한 노래였습니다. 틱톡과 스포티파이에서 폭발적으로 재생됐습니다.

두 가수의 음반사가 삭제를 요청해 곡은 내려갔습니다.

흥미로운 뒷이야기가 있습니다. 이 곡을 좋아한 사람들이 꽤 있었습니다. 진짜보다 낫다는 말도 나왔습니다. 그 반응이 음악 산업을 더 불안하게 만들었습니다.

드레이크 본인도 2024년 4월 디스곡에 세상을 떠난 래퍼 투팍 샤커의 목소리를 인공지능으로 복제해 썼다가 유족 측 경고를 받고 곡을 내렸습니다.

목소리를 복제당한 쪽이 다른 사람의 목소리를 복제한 겁니다.

중국에서는 한 가수의 음성을 복제해 다른 노래를 부르게 한 음원들이 대량으로 퍼져 수백만 번 재생됐습니다.

목소리로 먹고사는 사람들에게 이 기술은 생계 문제입니다.

베테랑 성우 폴 스카이 리먼과 리네아 세이지는 2024년 5월 뉴욕 연방법원에 소송을 냈습니다. 한 음성 생성 스타트업이 대화 연구 목적이라며 녹음을 해 놓고, 그 데이터를 상업용 인공지능 성우 서비스 학습에 돌려 팔았다는 겁니다.

성우 베브 스탠딩의 목소리는 틱톡의 자동 읽기 음성으로 쓰였습니다. 동의 없이 학습된 것을 알고 소송을 내 합의를 받았습니다.

이 사건의 기묘한 점은 규모입니다. 자기 목소리가 하루에 수억 번 재생되는데, 그 재생에서 한 푼도 받지 못했습니다.

오디오북 낭독자들도 애플이 자기 목소리를 학습에 썼다고 고발했습니다. IBM은 성우 그레그 마스턴의 목소리를 상업용 클로닝에 팔았다가 비판을 받았습니다. 영국 스코트레일의 안내방송 성우도 자기 목소리가 인공지능으로 바뀌어 쓰였다고 호소했습니다.

게임 회사들이 인간 성우 대신 복제 음성을 쓰려다 성우 조합과 게이머들의 반발에 부딪히는 일도 이어졌습니다.

얼굴 쪽은 사기와 곧바로 이어집니다.

틱톡의 딥톰크루즈 계정은 배우 톰 크루즈의 얼굴과 목소리를 완벽하게 합성한 영상으로 수백만 팔로워를 모았습니다. 만든 사람들은 위험을 알리려는 목적이라고 했습니다. 사람들은 재미있어했습니다.

배우 톰 행크스는 자기가 나오지도 않은 치과 보험 광고에 자신의 합성 이미지가 쓰이자 직접 경고를 냈습니다. 브루스 윌리스는 광고에 자기 얼굴을 쓰도록 계약했다가, 이후 본인 의사와 무관하게 얼굴이 복제돼 쓰인다는 논란에 휘말렸습니다.

방송인 마틴 루이스, 앵커 메리 나이팅게일, 유튜버 미스터비스트, 기상 캐스터와 뉴스 진행자들의 얼굴이 비트코인 투자 사기와 가짜 경품 광고에 합성됐습니다.

이 사기들이 왜 잘 통하는지 생각해 보면 서늘합니다. 사람들은 잘 아는 얼굴을 믿습니다. 매일 저녁 뉴스에서 보던 얼굴이 투자를 권하면 경계가 풀립니다.

프랑스에서는 배우 브래드 피트의 딥페이크에 속은 여성이 팔십삼만 유로를 잃었습니다. 몇 달에 걸쳐 대화를 나눴다고 합니다.

인도의 대배우 아닐 카푸르는 자기 이름과 얼굴과 목소리와 유행어가 광고와 상품에 무단으로 쓰이자 뉴델리 고등법원에 소송을 냈습니다. 법원은 인격권과 초상권 침해를 인정하고, 동의 없는 모든 인공지능 복제와 합성물 유통을 금지하는 가처분을 내렸습니다.

이 결정은 중요한 선례가 됐습니다. 얼굴과 목소리와 말버릇 전체를 한 사람의 인격으로 묶어 보호한 겁니다.

정신적 지도자 사드구루도 조작 영상 유포에 대해 삭제 명령을 받아 냈습니다. 스페인의 한 주류 브랜드는 세상을 떠난 가수를 딥페이크로 되살려 광고에 썼다가 논란을 맞았습니다.

컨설팅 그룹 WPP의 최고경영자 마크 리드는 딥페이크로 합성된 화상회의 화면 때문에 거액 사기의 표적이 될 뻔했습니다. 화면 속에서 자기 얼굴이 회의를 하고 있었던 겁니다.

선거에서는 목소리가 무기가 됐습니다.

2024년 1월 미국 뉴햄프셔주 민주당 예비선거를 앞두고 유권자 수천 명에게 전화가 걸려 왔습니다. 조 바이든 대통령의 목소리였습니다. 예비선거에 나오지 말고 본선을 위해 표를 아끼라는 내용이었습니다.

전화를 받은 사람들은 대통령의 목소리를 압니다. 그래서 의심하지 않습니다.

수사 당국은 이 로보콜을 만든 업자를 적발해 무거운 벌금을 물렸습니다.

일론 머스크는 카멀라 해리스 부통령의 목소리를 복제해 그가 자신의 무능을 인정하는 것처럼 만든 가짜 정치 광고 영상을 자기 플랫폼에 공유해 비판을 받았습니다.

뉴욕 시장 에릭 애덤스는 자기 목소리를 복제해 스페인어와 중국어로 된 자동 전화를 시민들에게 보냈습니다. 그는 그 언어들을 하지 못합니다.

영국 노동당 대표 키어 스타머는 직원에게 욕설하는 가짜 음성 때문에 타격을 입었습니다. 런던 시장 사디크 칸도 현충일 행사를 비하하는 가짜 음성 파일로 항의를 받았습니다.

2023년 슬로바키아에서는 총선 며칠 전에 한 정당 대표가 선거 조작을 논의하는 가짜 음성이 퍼졌습니다. 선거 직전에는 반박할 시간이 없습니다. 그 점을 노린 겁니다.

태국, 영국, 독일, 헝가리에서도 합성 음성과 영상이 정쟁에 쓰였습니다. 한국에서도 2022년 대선에서 후보들의 인공지능 아바타가 등장해 정치적 대리인의 경계를 두고 논쟁이 있었습니다.

제일 잔혹한 쓰임은 마지막입니다.

2024년 1월 엑스에 팝스타 테일러 스위프트의 얼굴을 합성한 성적 이미지들이 폭발적으로 퍼졌습니다. 수천만 번 조회되는 동안 삭제되지 않았습니다.

배우 갤 가돗, 미국 하원의원 알렉산드리아 오카시오코르테스, 이탈리아 총리 조르자 멜로니, 인도 저널리스트 라나 아유브를 비롯한 많은 여성 공인이 같은 일을 당했습니다. 멜로니 총리는 제작자를 상대로 민사 소송을 냈습니다.

이 피해자 명단에서 한 가지가 눈에 띕니다. 거의 다 여성입니다. 그리고 대개 공적으로 발언하는 여성입니다.

그리고 이 범죄는 학교로 내려왔습니다. 스페인 알멘드랄레호에서 중학생들이 같은 학교 여학생 스무 명 남짓의 사진을 성적 이미지로 만들어 돌린 사건이 알려졌습니다. 미국과 캐나다와 스코틀랜드와 호주의 고등학교들에서 같은 일이 이어졌습니다.

한국에서도 2024년 텔레그램을 통한 딥페이크 성범죄가 중고등학교와 대학교를 뒤흔들었습니다.

이 문제는 다음 장에서 더 다루겠습니다. 여기서는 기술 쪽만 짚습니다.

사진 한 장이면 됩니다. 목소리는 몇 초면 됩니다. 이것이 지금 기술의 수준입니다.

그리고 이 기술을 만든 곳들은 대체로 같은 것을 만들지 않았습니다. 이 사람이 동의했는지 확인하는 절차입니다.

기술적으로 불가능해서가 아닙니다. 만들면 서비스가 불편해지고 이용자가 줄기 때문입니다.

얼굴과 목소리는 우리가 태어날 때 받은 것입니다. 바꿀 수 없고, 숨길 수 없고, 매일 남에게 보여 주며 삽니다.

그것을 재료로 쓰는 산업이 생겼습니다. 재료값은 아직 아무도 내지 않았습니다.

제 5 장

자율주행과 로봇 공학: 물리적 안전과 인명 피해

5.1

자율주행 차량 결함과 도로 교통사고

2018년 3월 18일 밤 아홉 시 오십팔 분, 미국 애리조나주 템피.

마흔아홉 살 일레인 허츠버그는 자전거를 밀며 어두운 도로를 건너고 있었습니다. 횡단보도는 아니었습니다.

우버의 자율주행 시험 차량이 시속 삼십팔 마일로 다가왔습니다. 운전석에는 시험 운전자가 앉아 있었고, 그는 휴대폰으로 예능 프로그램을 보고 있었습니다.

차의 센서는 충돌 육 초 전에 이미 허츠버그를 감지했습니다.

여섯 초는 긴 시간입니다. 하나, 둘, 셋을 두 번 셀 수 있습니다. 그 시간이면 차를 세울 수 있습니다.

문제는 소프트웨어가 그 육 초 동안 무엇을 했느냐입니다.

정체 불명의 물체로 분류했습니다. 그다음 차량으로 분류했습니다. 그다음 자전거로 분류했습니다. 다시 물체로 바꿨습니다. 분류가 바뀔 때마다 이 대상이 어디로 갈지 예측하는 계산이 처음부터 다시 시작됐습니다.

기계는 사람이 자전거를 밀고 걷는 상황을 배우지 못했습니다. 학습 데이터에서 자전거는 타는 물건이고 사람은 걷는 존재입니다. 둘이 붙어서 걷는 모습은 그 사이 어딘가에 있었습니다.

그리고 결정적인 설정이 있었습니다. 우버 개발진은 비상 자동 제동 기능을 꺼 두었습니다.

이유는 승차감이었습니다. 시스템이 자꾸 급제동을 걸어서 시승이 불편했던 겁니다. 대신 사람 운전자가 위험할 때 개입하도록 정해 두었습니다.

그 사람은 화면을 보고 있었습니다.

차는 속도를 줄이지 않았습니다. 일레인 허츠버그는 자율주행차에 목숨을 잃은 첫 보행자로 기록됐습니다.

이 사고에는 자동화의 오래된 함정이 다 들어 있습니다. 기계가 대부분을 잘하면 사람은 긴장을 놓습니다. 그런데 시스템은 사람이 늘 긴장하고 있다는 가정 위에 설계됐습니다.

사람은 그렇게 만들어지지 않았습니다. 아무 일도 일어나지 않는 화면을 몇 시간 동안 집중해서 보는 일은 사람이 제일 못하는 일에 듭니다.

우버 사고 뒤에도 무인 택시는 도로로 나왔습니다.

2023년 10월 샌프란시스코 도심에서 한 여성이 다른 차에 치여 튕겨 나갔습니다. 그는 옆 차선에 있던 제너럴모터스 계열사 크루즈의 무인 택시 밑으로 들어갔습니다.

여기서 시스템은 다시 계산을 시작했습니다. 차 밑에 사람이 있다는 사실은 인식하지 못했습니다. 하부 센서로는 그 상황이 잡히지 않았습니다.

시스템은 정해진 규칙을 실행했습니다. 사고가 감지되면 도로 가장자리로 이동해 안전하게 정차하라.

무인 택시는 시속 칠 마일로 육 미터를 이동했습니다. 사람을 깔고 끌면서 갔습니다.

피해자는 영구적인 중상을 입었습니다.

이 대목이 자동화의 무서움을 정확히 보여 줍니다. 규칙 자체는 합리적입니다. 사고가 나면 길을 막지 말고 갓길로 빠져라. 사람 운전자에게도 그렇게 가르칩니다.

사람 운전자라면 차 밑에서 나는 소리를 들었을 겁니다. 비명을 들었을 겁니다. 밖에서 사람들이 손을 흔드는 것을 봤을 겁니다.

규칙에 없는 것을 알아차리는 능력. 그것을 뭐라고 불러야 할지 모르겠지만, 기계에는 없습니다.

그다음 일은 기술 문제가 아닙니다. 조사에 들어간 당국에 크루즈 경영진은 사람을 끌고 간 부분의 영상을 내놓지 않았습니다. 안전하다는 보고를 냈습니다.

캘리포니아주는 크루즈의 무인 운행 허가를 취소했습니다. 법무부 형사 조사가 시작됐고, 최고경영자가 물러났고, 로보택시 운행이 전면 중단됐습니다.

크루즈 차량들은 그전에도 여러 일을 냈습니다.

굴절버스의 접힌 부분이 움직이는 것을 잘못 읽어 버스 뒤를 받았습니다. 사이렌을 켠 소방차를 제때 알아보지 못해 교차로에서 부딪혔습니다. 총격 사건 현장에 출동한 경찰차 앞을 막아섰습니다.

내부 안전 평가 보고서에는 어린이 인식에 한계가 있다는 내용이 들어 있었습니다. 아이들은 갑자기 뛰고, 예측하기 어렵게 방향을 바꾸고, 키가 작습니다.

그리고 이 문장이 나옵니다. 크루즈 무인 택시는 도심에서 사 킬로미터에서 팔 킬로미터를 달릴 때마다 원격 관제 센터의 사람이 개입해야 운행이 유지됐습니다.

무인이라는 말에는 조건이 붙어 있었습니다.

웨이모의 기록도 짧지 않습니다.

2024년 2월 샌프란시스코에서 자전거 운전자를 쳐 다치게 했습니다. 2023년 12월부터 2024년 2월 사이 피닉스에서는 견인차에 거꾸로 매달려 끌려가던 픽업트럭을 두 대가 연달아 들이받았습니다.

거꾸로 매달린 채 반대 방향으로 움직이는 트럭. 사람은 이 장면을 보면 웃으면서 피합니다. 학습 데이터에는 없는 장면입니다.

2024년 5월에는 피닉스 골목에서 전신주를 받았습니다. 2025년 11월 조지아주 애틀랜타에서는 정차 신호를 켜고 아이들을 내리던 스쿨버스를 무시하고 돌아 지나갔습니다. 미국 도로교통안전청이 조사에 들어갔습니다.

스쿨버스가 빨간 등을 켜고 정지 표지판을 펼치면 모든 차가 서야 합니다. 운전면허 시험에 나오는 규칙입니다.

2026년 1월 23일, 캘리포니아주 샌타모니카.

초등학교에서 두 블록 떨어진 곳이었습니다. 등교 시간이었고, 근처에 다른 아이들과 교통지도원이 있었고, 이중주차한 차들이 늘어서 있었습니다.

웨이모 무인차가 아이를 쳤습니다.

웨이모는 이렇게 밝혔습니다. 서 있던 차 뒤에서 사람이 나오자마자 즉시 감지했고, 강하게 제동해 시속 십칠 마일에서 육 마일로 줄인 상태에서 접촉했다.

아이는 가벼운 부상을 입었습니다. 도로교통안전청이 조사에 들어갔습니다.

이 설명은 기술적으로 훌륭합니다. 사람보다 빨리 감지했고 사람보다 세게 밟았을 겁니다.

그런데 이 문장을 그 아이의 부모가 읽는다고 생각해 보십시오. 시속 십칠에서 육으로 줄였다는 사실이 위로가 되지는 않습니다.

웨이모 차량들은 그 밖에도 골목에서 뛰어나온 개와 가게 앞 고양이를 쳐 죽였고, 총기 난사 현장으로 가던 구급차의 길을 막았고, 하차 알림 소프트웨어 오류로 승객이 문을 열다가 자전거 운전자를 다치게 했습니다.

제일 긴 목록은 테슬라의 것입니다.

2016년 5월 플로리다에서 오토파일럿으로 달리던 모델 S가 좌회전하던 대형 트레일러 옆면을 받았습니다. 밝은 하늘 아래 흰색 트레일러가 있었습니다. 카메라는 하얀 배경과 하얀 차체를 구별하지 못했습니다.

차는 브레이크를 밟지 않고 트레일러 밑으로 들어갔습니다. 지붕이 날아갔고 운전자 조슈아 브라운이 즉사했습니다.

2018년 3월 캘리포니아 마운틴뷰에서는 애플 엔지니어 월터 황이 모델 X를 오토파일럿으로 몰다가 콘크리트 방호벽을 정면으로 받았습니다. 시속 칠십일 마일이었습니다.

원인은 차선 표시였습니다. 나들목에서 차선이 갈라지는 곳의 흐릿한 선을 시스템이 잘못 따라갔고, 차를 방호벽 쪽으로 몰았습니다.

황은 사고 전에 같은 지점에서 차가 방호벽 쪽으로 쏠린다고 여러 번 이야기했습니다.

야간에 서 있는 큰 물체를 못 알아보는 문제는 계속됐습니다.

2019년 3월 플로리다 델레이비치에서 모델 3이 도로를 가로지르던 대형 트레일러를 시속 육십팔 마일로 받아 운전자 제러미 배너가 숨졌습니다. 2016년 사고와 같은 상황이었습니다. 삼 년 사이에 고쳐지지 않았습니다.

2019년 12월 인디애나주에서는 경광등을 켜고 사고 현장에 서 있던 소방차 뒤를 받아 탑승자의 아내가 숨졌습니다. 같은 달 캘리포니아 가데나에서는 오토파일럿이 켜진 모델 S가 고속도로에서 나와 빨간불을 무시하고 달려 혼다 시빅을 들이받았습니다. 두 사람이 숨졌습니

다.

이 사고의 운전자는 오토파일럿 사용자로는 처음으로 과실치사 혐의로 형사 기소됐습니다.

2020년 9월 조지아주에서는 시속 사십오 마일 제한 구역을 시속 칠십칠 마일로 달리던 모델 3이 빗길에 미끄러져 버스 정류장을 덮쳤습니다. 정류장에 앉아 있던 일흔여섯 살 버나드 존스가 숨졌습니다.

2021년 4월 텍사스주 스프링에서는 운전석에 아무도 없는 상태로 달리던 모델 S가 도로를 벗어나 나무를 받고 불탔습니다. 두 사람이 숨졌습니다.

2022년 11월 샌프란시스코 베이 브리지 터널에서는 반대 방향의 사고가 났습니다. 시속 육십오 마일로 달리던 차가 앞에 아무것도 없는데 갑자기 급제동을 걸었습니다. 유령 제동이라고 부릅니다. 뒤따르던 차 여덟 대가 연쇄 추돌했고 아이를 포함해 아홉 명이 다쳤습니다.

오토바이 운전자들은 유난히 위험했습니다.

2022년 7월 캘리포니아에서 모델 Y가 앞서 가던 야마하 오토바이를 그대로 받았습니다. 같은 달 유타에서 모델 3이 할리데이비슨을 받았습니다. 8월에는 가와사키를 받았습니다.

원인은 후미등에 있습니다. 오토바이는 등이 하나이고 작습니다. 시스템은 밤에 그 등을 멀리 있는 가로등이나 도로 표지판으로 읽었습니다. 멀리 있다고 판단하면 속도를 줄이지 않습니다.

오토바이를 타는 사람에게 뒤에서 오는 차는 보이지 않습니다.

2023년 11월 애리조나에서는 저녁 햇빛이 강하게 반사되면서 카메라만 쓰는 시스템이 앞을 보지 못해 일흔한 살 보행자를 쳐 숨지게 했습니다. 2023년 7월 브루클린에서는 인도에서 버스를 기다리던 일흔여섯 살 보행자가 목숨을 잃었습니다.

2023년 5월 테슬라 내부 고발자 우카시 크룹스키가 백 기가바이트 분량의 문서를 공개했습니다. 갑작스러운 제동과 예기치 않은 가속에 대한 고객 신고가 수천 건 쌓여 있었고, 경영진이 이를 알고도 밖에 알리지 않았다는 내용이었습니다.

2026년에도 조사가 이어지고 있습니다. 텍사스에서 테슬라 차량이 주택을 들이받아 일흔여섯 살 여성이 숨진 사고로 도로교통안전청과 교통안전위원회가 조사에 들어갔습니다. 테

슬라 차량이 빨간불을 무시하거나 역주행했다는 신고가 수십 건 접수돼 별도 조사도 열렸습니다.

2026년 7월에는 휴스턴에서 테슬라 로보택시의 원격 조작자가 사고를 낸 기록이 당국 자료에서 확인됐습니다. 무인 차량을 사람이 원격으로 몰다가 사고를 낸 겁니다.

다른 회사들도 목록에 있습니다.

아마존의 자율주행 회사 죽스는 2025년 4월 라스베이거스 사고 뒤 이백칠십 대를 소프트웨어로 리콜했고, 8월에는 중앙선을 넘어 반대편 차선 앞에 갑자기 서는 결함으로 삼백삼십이 대를 추가 리콜했습니다.

중국 바이두의 무인차는 2024년 보행자를 쳤습니다. 한 자율주행 스타트업은 2021년 캘리포니아 시험 주행 중 중앙분리대와 표지판을 받아 허가가 정지됐습니다.

중국 전기차 니오의 차량은 2021년 8월 주행 보조 기능을 켜고 달리다 고속도로 정비 차량을 그대로 받아 운전자가 숨졌습니다. 엑스펑 차량도 트럭을 받았습니다. 샤오미의 차량은 시멘트 기둥을 받거나 교량에서 떨어져 사망 사고를 냈습니다.

도요타가 2021년 도쿄 패럴림픽 선수촌에 넣은 자율주행 셔틀은 횡단보도를 건너던 시각장애인 유도 선수를 쳤습니다. 선수는 다쳐서 경기에 나가지 못했습니다. 도요타는 센서가 선수를 감지했으나 운영자와 자동 제동 사이의 신호가 어긋났다고 인정했습니다.

포드의 주행 보조 시스템을 단 차량들은 고속도로에 서 있던 승용차들을 연달아 받아 사망 사고를 냈고 대규모 조사를 받았습니다.

이 목록에서 반복되는 상황을 뽑아 보면 이렇습니다.

밤에 서 있는 큰 물체. 밝은 배경 앞의 흰색 차량. 강한 역광. 작은 후미등. 걷다가 갑자기 나오는 사람. 예상 밖의 자세로 있는 차량. 아이.

전부 사람 운전자가 어렵지 않게 처리하는 상황입니다.

여기에 이름이 붙어 있습니다. 오토파일럿, 완전 자율주행. 두 이름 다 실제 기능보다 큽니다.

이름이 사람의 행동을 바꿉니다. 오토파일럿이라고 적힌 기능을 켜 놓고 운전대에서 손을 떼지 않을 사람은 많지 않습니다. 설명서 구석에 항상 주시하라고 적혀 있어도 마찬가지입니다.

사고가 나면 회사들은 같은 말을 합니다. 이 기능은 보조 수단이며 운전자에게 책임이 있습니다.

맞는 말입니다. 다만 그 말은 이름표에는 적혀 있지 않았습니다.

5.2

산업용·서비스용 로봇의 오작동 및 안전 사고

2023년 11월, 경상남도 고성군의 파프리카 선별장.

로봇 팔 하나가 컨베이어 위의 파프리카 상자를 집어 팔레트에 쌓고 있었습니다. 하루에 수천 번 하는 동작입니다.

센서를 점검하던 사십 대 로봇 업체 직원이 그 앞에 있었습니다.

로봇은 그를 상자로 인식했습니다.

집게가 그를 잡아 컨베이어 쪽으로 눌렀습니다. 얼굴과 가슴이 눌렸습니다. 병원으로 옮겨졌지만 숨졌습니다.

기술진은 그 로봇의 센서에 이상이 있다는 사실을 알고 있었습니다. 사람과 로봇 사이의 안전 거리도, 점검 중 전원을 끊는 절차도 지켜지지 않았습니다.

이 사고에서 제일 무거운 문장은 이겁니다. 로봇은 사람과 상자를 구별하지 못했습니다.

산업용 로봇은 대체로 눈이 없습니다. 있어도 사람을 알아보라고 만든 눈이 아닙니다. 정해진 위치에서 정해진 크기의 물체를 잡도록 만들어졌습니다. 그 자리에 무엇이 있든 잡습니다.

그래서 이런 로봇은 원래 우리 안에 둡니다. 철망을 치고, 문이 열리면 전원이 끊기게 합니다. 사람과 기계를 물리적으로 갈라 놓는 겁니다.

사고는 대개 그 철망 안에 사람이 들어갔을 때 일어납니다. 그리고 사람이 들어가는 이유는 거의 정해져 있습니다. 기계가 고장 나서입니다.

고칠 사람은 들어가야 하고, 들어가면 위험합니다. 이 모순을 관리하는 것이 안전 규정입니다. 규정은 대개 서류로 존재합니다.

2016년 6월 미국 앨라배마주의 현대·기아차 부품 공급업체 공장에서, 결혼을 앞둔 이십 대 노동자 레지나 앨런 엘시가 로봇 구역에 들어갔습니다. 조립 라인이 멈춰서 센서 오류를 확

인하러 간 겁니다.

시스템이 갑자기 다시 켜졌고, 로봇 팔이 그를 벽 구조물로 밀었습니다. 그는 숨졌습니다.

미국 노동부는 이 회사와 인력 파견 업체들에 이백오십만 달러의 벌금을 물렸습니다.

2015년 7월 미시건주에서는 쉰일곱 살 정비 기술자 완다 홀브룩이 일상 점검 중에 로봇 사이에 갇혀 숨졌습니다. 같은 달 독일 바우나탈의 폭스바겐 공장에서는 스물한 살 계약직 기술자가 고정식 로봇을 설치하다가 로봇 팔에 잡혀 금속판으로 밀려 숨졌습니다. 인도의 한 금속 공장에서도 노동자가 프로그래밍된 로봇 팔에 치여 목숨을 잃었습니다.

미국 산업안전보건청의 중대재해 보고를 2015년부터 2022년까지 분석한 자료가 있습니다. 로봇 관련 사고 일흔일곱 건이 확인됐고, 그중 60퍼센트 넘는 사고의 원인이 같았습니다.

예기치 않은 작동입니다.

꺼져 있는 줄 알았던 기계가 켜진 겁니다.

2021년 텍사스주 오스틴의 테슬라 공장에서 일어난 일이 그 전형입니다. 소프트웨어 엔지니어가 알루미늄 차체를 자르는 로봇 프로그램을 짜고 있었습니다. 로봇 세 대 중 두 대는 정비를 위해 전원을 껐습니다. 나머지 한 대는 관리 오류로 켜진 채 남아 있었습니다.

그 한 대가 그를 금속 벽에 고정한 뒤 집게 형태의 금속 갈퀴를 등과 팔에 찔러 넣었습니다. 바닥에 피가 고였습니다. 동료가 비상 정지 버튼을 누른 뒤에야 빠져나올 수 있었습니다.

휴머노이드 로봇이 들어오면서 상황은 새로워졌습니다.

테슬라 노동자 한 명이 프리몬트 공장에서 옵티머스 휴머노이드 로봇에 다쳐 소송을 냈습니다. 2025년 2월 정비 근무 중이었고, 로봇이 예기치 않게 작동했다는 주장입니다. 그는 의식을 잃었고 약 팔천 파운드 무게의 균형추에 깔렸습니다.

중국의 한 공장에서는 휴머노이드 로봇이 팔을 휘두르며 컴퓨터를 넘어뜨리고 작업자 두 명을 칠 뻔한 영상이 공개됐습니다.

문제는 여기 있습니다. 휴머노이드 로봇을 위한 안전 표준이 아직 없습니다.

지금까지의 안전 규칙은 로봇이 한자리에 고정돼 있다는 전제 위에 만들어졌습니다. 철망을 어디에 치고, 어느 반경 안에 들어가지 말라는 식입니다.

걸어 다니는 로봇에는 이 규칙이 통하지 않습니다. 철망 안에 두면 쓸모가 없기 때문입니다. 사람 옆에서 일하라고 만든 물건입니다.

테슬라는 프리몬트와 텍사스 공장에 옵티머스를 천 대 넘게 투입했습니다.

물류 창고에서는 다른 종류의 사고가 납니다.

2018년 12월 미국 뉴저지주의 아마존 물류 창고에서 자동화 로봇이 곰 퇴치용 스프레이 캔을 찢었습니다. 아홉 온스짜리였습니다.

캡사이신 농축액이 창고 공기 중에 퍼졌습니다. 팔십 명 넘는 노동자가 숨을 못 쉬겠다고 호소했고, 스물네 명이 병원으로 옮겨졌으며, 한 명은 중환자실에 들어갔습니다.

로봇은 상자 안에 무엇이 들었는지 모릅니다. 얼마나 세게 쥐면 되는지도 모릅니다. 정해진 힘으로 집을 뿐입니다.

아마존 창고에서 로봇 도입 비율이 높은 곳일수록 노동자 중상 비율이 오히려 높다는 조사도 나왔습니다.

이유는 짐작할 만합니다. 로봇이 옆에서 정해진 속도로 움직이면 사람도 그 속도에 맞춰야 합니다. 기계는 지치지 않습니다.

2021년 영국의 온라인 식품 유통 기업 오카도의 물류 센터에서는 물건을 나르던 로봇 두 대가 정면으로 부딪혔습니다. 충격으로 불이 났고 창고 전체로 번졌습니다.

소방관 삼백 명이 사흘 동안 불을 껐습니다. 인근 주민들이 대피했습니다.

이 회사는 2019년에도 앤도버 창고에서 로봇 충전기의 전기 결함으로 큰불을 겪었습니다. 창고가 잿더미가 됐고 삼백칠십 개의 일자리가 사라졌습니다.

로봇끼리 부딪히지 않게 하는 센서와 열이 오르는 것을 감지하는 장치가 없었습니다.

공장 밖으로 나온 로봇들은 사람을 칩니다.

배달 로봇 회사 스타십 테크놀로지스의 소형 로봇들은 여러 사고를 냈습니다. 2024년 9월 애리조나 주립대 교정에서 직원을 쳐 넘어뜨렸습니다. 2023년 11월 영국 밀턴킨스의 쇼핑센터에서는 두 살배기를 치고 밀었습니다.

영국 러싱던에서는 쇼핑객을 쓰러뜨렸고, 애리조나와 켄터키에서는 신호 대기 중인 승용차를 받고 그냥 갔습니다. 휠체어 이용자의 통행로를 막고 서 있기도 했습니다.

밀턴킨스에서는 식료품을 나르던 로봇이 경로를 잘못 잡아 운하에 빠졌습니다. 화물 열차와 부딪힌 사고도 있었습니다.

2018년 12월 캘리포니아 대학교 버클리 캠퍼스에서는 배달 로봇이 배터리 결함으로 폭발해 불이 났습니다.

경비 로봇은 더 당황스럽습니다.

캘리포니아 팔로알토의 쇼핑센터에서 순찰하던 나이트스코프 K5 경비 로봇이 열여섯 달 된 아기 하윈 첸을 밀어 넘어뜨리고 발 위로 지나갔습니다.

이 로봇은 키가 백오십 센티미터가 넘고 무게가 백삼십 킬로그램이 넘습니다. 센서는 바닥에 있는 아기를 장애물로 읽지 못했습니다.

같은 회사 로봇은 2017년 7월 워싱턴 디시의 사무 빌딩을 순찰하다가 계단 감지 센서 오작동으로 야외 분수대에 빠졌습니다. 인터넷에서 이 사진은 오래 회자됐습니다.

캘리포니아 헌팅턴파크에서는 폭력 사건을 목격한 여성 시민이 로봇의 비상 버튼을 눌렀습니다. 로봇은 비켜 달라는 안내음을 내며 노래를 부르고 지나갔습니다.

뉴욕 경찰청이 타임스스퀘어 지하철역에 시험 배치한 같은 모델도 2024년에 운용이 중단됐습니다.

전시장에서는 로봇이 관객에게 달려들었습니다.

2016년 중국 쑤저우의 한 박람회에서 어린이 교육용 로봇 샤오팡이 제어 오작동으로 갑자기 속도를 냈습니다. 유리 전시 부스를 부수며 관람객 쪽으로 돌진했고, 한 사람이 유리 파편과 충격으로 다쳐 병원에 갔습니다.

2025년 톈진의 한 축제에서는 로봇이 관중석으로 넘어 들어가 사람들을 향해 움직였습니다. 러시아가 야심 차게 공개한 휴머노이드 로봇은 데뷔 무대에서 균형을 잃고 넘어져 부서졌습니다.

2024년 6월 한국 구미시청에 도입된 행정 로봇은 계단 감지 센서가 오작동해 이 미터 아래로 떨어져 파손됐습니다.

제일 오래 기억에 남는 장면은 2022년 모스크바 체스 대회에서 나왔습니다.

일곱 살 어린이 선수가 로봇과 체스를 두고 있었습니다. 아이가 정해진 순서보다 빠르게 말을 옮겼습니다.

로봇은 그 손을 체스 말로 인식했습니다. 집게가 아이의 손가락을 잡았고, 부러뜨렸습니다.

아이는 다음 날 깁스를 하고 대회를 마쳤다고 합니다.

수술실에서는 오작동이 곧바로 몸속에서 벌어집니다.

미국 식품의약국 데이터베이스 분석에 따르면 2000년부터 2013년까지 로봇 보조 수술 중 기계 오작동으로 숨진 환자가 백사십사 명, 중상을 입은 환자가 천 명이 넘습니다.

수술 도중 로봇 팔이 갑자기 움직이거나, 부품이 환자 몸 안으로 떨어지거나, 전선이 끊겨 불꽃이 튀면서 장기에 화상을 입히는 사고들이었습니다.

2026년에는 인공지능이 들어간 부비동 수술 도구가 위치를 잘못 판단해 환자들의 부비동과 신경 조직을 반복적으로 손상시켜 소송에 휘말렸습니다. 한 회사가 만든 인공지능 혈당 측정 센서의 수치 오류로 최소 일곱 명이 숨진 일도 있었습니다.

여기까지는 무거운 이야기입니다. 서비스 로봇 쪽으로 오면 좀 웃깁니다.

캘리포니아 파사데나의 한 햄버거 가게에 들어온 조리 로봇 플리피는 주문 속도를 따라가지 못하고 기름을 바닥에 흘리다가 하루 만에 물러났습니다.

스코틀랜드 에든버러의 한 슈퍼마켓에 배치된 안내 로봇 파비오는 손님의 쉬운 질문에 엉뚱한 답을 했습니다. 매장 소음 때문에 센서도 제대로 작동하지 않았습니다. 일주일 만에 퇴출됐습니다.

일본의 한 호텔은 로봇 이백사십삼 대로 운영하는 완전 자동 호텔을 표방했습니다.

객실의 음성 비서는 투숙객의 코 고는 소리를 명령으로 알아듣고 밤새 말을 걸었습니다. 짐 나르는 로봇은 비와 눈에 고장 났습니다.

호텔은 로봇의 절반 이상을 내보내고 사람을 다시 뽑았습니다.

2025년에는 사무실 자판기를 관리하는 인공지능 에이전트들이 주문 오류로 재정 손실을 내고 시스템이 멈추기도 했습니다.

이 사례들에는 웃음과 서늘함이 나란히 있습니다. 코 고는 소리에 대답하는 로봇은 웃깁니다. 아이 손가락을 부러뜨린 로봇은 웃기지 않습니다.

그런데 두 로봇의 실수는 같은 종류입니다. 정해진 입력 범위 밖의 것이 들어왔을 때 무엇을 해야 할지 모른다는 겁니다.

사람은 모를 때 멈춥니다. 두리번거리고, 물어보고, 손을 뗍니다.

기계는 모를 때도 그럴듯해 보이는 동작을 실행합니다. 멈추는 것은 기본값이 아닙니다.

기본값을 멈춤으로 놓으면 생산성이 떨어집니다. 그래서 대개 그렇게 놓지 않습니다.

파프리카 선별장의 로봇도 멈추지 않았습니다.

5.3

군사용 자율 무기 및 드론 기술의 치명적 위험

2020년 3월, 리비아.

내전에서 밀린 병사들이 후퇴하고 있었습니다. 하늘에 작은 드론이 떠 있었습니다. 튀르키예 방산 기업이 만든 카구 2였습니다.

유엔 안전보장이사회 전문가 패널 보고서에 따르면, 이 드론은 사람의 지시 없이 스스로 표적을 정하고 공격했습니다.

카메라가 병사들을 잡았고, 알고리즘이 추적했고, 드론이 내리꽂혔습니다. 방아쇠를 당긴 사람이 없었습니다.

이 사건이 역사에 남은 이유는 사망자 수 때문이 아닙니다. 기계가 사람의 승인 없이 사람을 죽인 첫 기록으로 보고됐기 때문입니다.

전쟁에도 규칙이 있습니다. 항복한 사람은 쏘지 않습니다. 후퇴하는 병사와 무기를 버린 병사는 다르게 대합니다. 민간인과 전투원을 구별해야 합니다.

이 규칙들은 전부 판단을 요구합니다. 저 사람이 손을 든 것인지, 무기를 든 것인지. 저 건물이 학교인지 지휘소인지.

알고리즘에게 판단은 분류입니다. 그리고 분류에는 언제나 오차가 있습니다. 다른 분야에서는 오차가 잘못된 광고나 틀린 추천으로 끝납니다.

여기서는 사람이 죽습니다.

2020년 11월 27일 이란 테헤란 외곽의 압사르드 고속도로에서는 다른 방식의 죽음이 있었습니다.

이란의 핵과학자 모센 파크리자데가 차를 타고 지나가고 있었습니다. 옆자리에는 아내가 있었습니다.

도롯가에 세워진 픽업트럭 위에 원격 조종 기관총이 있었습니다. 여러 대의 카메라와 안면 인식 알고리즘, 위성 제어 장치가 달려 있었습니다.

시스템은 달리는 차 운전석에 앉은 파크리자데의 얼굴을 알아보고 그만 쏘았습니다. 바로 옆의 아내는 맞지 않았습니다.

현장에는 사람이 한 명도 없었습니다. 작전이 끝나자 기관총은 스스로 폭발해 없어졌습니다.

이 장면의 정밀함이 이 이야기의 공포입니다. 옆 사람을 맞히지 않을 만큼 정확한 기계가, 사람 없이 작동했습니다.

가자 지구에서는 규모가 달라집니다.

이스라엘군이 쓴 하스보라와 라벤더라는 인공지능 시스템은 통신 데이터와 위치 정보와 행동 양식을 분석해 잠재적 표적을 자동으로 만들어 냈습니다. 그렇게 생성된 명단이 삼만 칠천 명 규모였다고 보도됐습니다.

알고리즘은 몇 초 만에 표적을 지정했습니다. 사람 분석관이 그 목록을 검토하는 데 걸린 시간은 수십 초 수준이었다고 전해집니다.

수십 초 동안 사람이 할 수 있는 일은 정해져 있습니다. 이름을 읽고 승인을 누르는 겁니다. 그 사람이 누구인지, 그 시각 그 건물에 몇 명이 있는지 확인할 시간은 없습니다.

이런 구조를 사람이 최종 결정한다고 부를 수 있는지가 지금 국제사회의 논쟁거리입니다.

서류상으로는 사람이 결정합니다. 실제로는 기계가 정한 것에 도장을 찍습니다.

하마스 간부를 겨눈 폭격에서 주변의 여성과 아이들이 함께 목숨을 잃은 사례들이 보고됐습니다.

서안 지구와 검문소에는 자동 발사 장치가 설치돼 최루탄과 진압탄을 시위대를 향해 쏘았습니다. 앞에서 나온 안면 감시망과 함께 작동했습니다.

드론은 이제 어디에나 있습니다.

튀르키예가 만든 바이라크타르 TB2는 우크라이나와 에티오피아 내전에 대량 투입됐습니다. 2022년 에티오피아 티그라이 지역에서는 이 드론이 사람들이 피신해 있던 초등학교를

폭격해 수십 명의 민간인이 숨졌습니다.

학교는 위에서 보면 큰 건물입니다. 마당이 있고 지붕이 넓습니다. 창고나 막사와 비슷하게 보입니다.

러시아의 자폭 드론은 우크라이나 전선에서 스스로 표적을 찾아 공격했습니다. 우크라이나군도 인공지능 드론 편대와 로봇 전용 부대를 만들어 무인 전쟁을 현실로 만들었습니다.

예멘의 후티 반군은 아부다비의 석유 저장소와 공항을 자폭 드론으로 공격했습니다. 국가가 아닌 무장 단체가 첨단 무기를 손에 쥐는 시대가 온 겁니다.

드론은 싸고, 작고, 만들기 쉽습니다. 전차나 전투기와 다릅니다. 공장과 숙련된 기술자와 수십 년의 축적이 필요하지 않습니다.

민간 인공지능 회사들도 이 흐름 안으로 끌려 들어갔습니다.

2026년 3월 미국 중앙사령부는 이란을 향한 대규모 합동 공습에서 앤스로픽의 인공지능 모델 클로드를 썼습니다. 기지 탐지와 첩보 분석, 표적 식별, 전투 시뮬레이션에 쓰였다고 보도됐습니다. 이 공습으로 이란 고위 관료들과 민간인 수백 명이 숨졌습니다.

경위가 특이합니다. 행정부는 앤스로픽에 무기화를 막는 안전장치를 제거하라고 요구했고, 회사는 거부했습니다. 그리고 회사가 블랙리스트에 오른 지 몇 시간 만에 군이 그 모델을 공습에 썼다고 전해집니다.

만든 회사가 안 된다고 한 용도로 쓰인 겁니다.

중국군은 메타가 공개한 오픈소스 모델을 가져다 군사용 인공지능 시스템을 만들었습니다. 오픈소스는 누구나 쓸 수 있게 공개한다는 뜻입니다. 누구에는 군대도 들어갑니다.

게임 엔진 회사 유니티도 미군과 군사용 인공지능 프로젝트를 개발해 온 사실이 드러나 소속 개발자들의 반발을 샀습니다.

기술을 만든 사람이 그 기술의 쓰임을 정하지 못합니다. 이것이 이 산업의 구조입니다.

경찰과 학교로 내려온 이야기도 있습니다.

샌프란시스코 경찰청은 총기 사건이나 인질극 현장에 살상력이 있는 로봇을 투입할 수 있게 하는 방안을 추진했습니다. 시민들이 반대해 철회됐습니다.

2016년 댈러스 경찰은 폭행 피의자를 향해 로봇에 폭발물을 붙여 폭파시킨 적이 있습니다. 뉴욕 경찰청은 로봇 개를 투입하려다 비판을 받았습니다.

네 발로 걷는 로봇 개에 소총을 얹은 군사용 모델도 등장했습니다. 사진을 보면 영화 소품 같습니다. 실제로 작동합니다.

총기 제조사 액손은 학교 총기 난사를 막겠다며 전기충격기를 단 드론을 교실에 배치하는 구상을 발표했습니다. 학부모들의 반발로 중단됐습니다.

교실 천장에 전기충격기 드론이 매달려 있는 학교를 상상해 보십시오. 그 교실에서 아이들이 무엇을 배울지 생각해 보면 답이 나옵니다.

개인이 만드는 물건도 있습니다.

한 미국 엔지니어는 챗GPT와 카메라를 연결해 사람이 나타나면 자동으로 조준하고 쏘는 장치를 만들어 공개했습니다. 부품은 시중에서 구할 수 있는 것들이었습니다.

상용 챗봇의 안전장치도 생각만큼 튼튼하지 않습니다. 연구자들의 시험에서, 몇 줄짜리 우회 문장으로 무기 제조와 관련된 위험한 지식을 끌어낼 수 있었습니다. 오픈AI의 한 모델은 몇 시간 만에 사만 개가 넘는 유해 화합물 후보를 만들어 낸 실험 결과도 보고됐습니다.

실제 사건들도 있었습니다. 프랑스에서 테러를 계획한 청소년이 챗봇을 이용했고, 미국에서는 호텔 앞 폭발 사건의 용의자가 챗봇으로 계획을 세웠습니다. 총격 사건들에서도 챗봇 이용 정황이 확인됐습니다.

사이버 공간에서는 문서가 무기가 됩니다.

북한 해커 조직 김수키는 챗GPT로 한국 군 신분증을 정교하게 위조해 군사 기밀 탈취에 썼습니다. 우크라이나를 지지하는 해커들은 인공지능이 만든 가짜 정부 문서로 러시아 방위산업 전산망에 침투했습니다.

미국 국방부 건물이 폭발했다는 인공지능 생성 이미지가 인터넷에 퍼졌을 때는 주식 시장이 순간적으로 떨어졌습니다. 사진 한 장이 시장을 흔든 겁니다.

전쟁의 가짜 영상들은 이제 하루에도 수없이 만들어집니다. 진짜 전쟁 영상과 섞여 돌아다니면서, 어느 쪽도 믿기 어렵게 만듭니다.

거짓이 널리 퍼지는 것보다 나쁜 일은 아무것도 믿지 않게 되는 겁니다.

이 문제를 국제사회가 손 놓고 있는 것은 아닙니다.

유엔 사무총장 안토니우 구테흐스는 사람의 통제 없이 작동하는 자율살상무기를 금지하는 법적 구속력 있는 조약을 2026년까지 만들자고 요구했습니다. 완전 자율 살상 시스템은 전면 금지하고 나머지는 엄격히 규제하는 두 단계 방식입니다.

백이십 개국 이상이 새 조약을 지지합니다. 유엔총회 결의에는 백오십육 개국이 찬성했습니다. 2025년 9월 정부전문가그룹 회의에서는 마흔두 개국이 지금 협상을 시작하자는 공동성명을 냈습니다.

미국은 전면 금지에 반대합니다. 자발적 행동규범을 선호합니다.

각국 국방 담당자들의 논리는 이해하기 어렵지 않습니다. 우리가 만들지 않으면 상대가 만든다는 겁니다.

이 논리는 반박하기 어렵고, 그래서 위험합니다. 모두가 이 논리를 따르면 아무도 멈추지 않습니다.

그리고 현장은 회의장보다 훨씬 빠릅니다. 조약 초안을 다듬는 동안 드론은 전선에서 이미 날고 있습니다.

이 절에 나온 사건들에서 한 가지가 빠져 있습니다.

누구도 처벌받지 않았다는 사실입니다.

리비아에서 스스로 공격한 드론에 대해, 학교를 폭격한 드론에 대해, 몇 초 만에 만들어진 표적 명단에 대해 형사 책임을 진 사람이 없습니다.

명령을 내린 사람은 시스템 추천을 따랐다고 말합니다. 시스템을 만든 회사는 사용자가 정한 용도라고 말합니다. 사용자는 정당한 군사 작전이었다고 말합니다.

책임이 원을 그리며 돌다가 사라집니다.

전쟁 범죄를 다루는 법은 사람이 사람에게 명령했다는 전제 위에 세워져 있습니다. 그 전제가 흔들리는 중입니다.

기계가 정하고 사람이 도장을 찍는 구조에서, 도장을 찍는 데 걸린 시간이 이십 초였다면, 그 결정은 누구의 것입니까.

이 질문에 답이 나오기 전에 다음 세대의 무기가 배치되고 있습니다.

제 6 장

정서적 왜곡과 디지털 범죄: AI 챗봇과 악의적 활용

6.1

AI 챗봇과의 과도한 정서적 교감 및 범죄 유도

이 절에는 사람이 목숨을 잃은 이야기가 여러 번 나옵니다. 힘든 이야기입니다. 그래도 적는 이유는, 이 사건들이 기술 설계의 문제이기 때문입니다.

2023년 벨기에에 피에르라는 삼십 대 남성이 살았습니다. 기후 위기에 대한 불안이 심했고 우울증을 앓고 있었습니다.

그는 챗봇 앱을 깔았습니다. 이름은 엘리자였습니다. 몇 달 동안 매일 이야기를 나눴습니다.

챗봇은 그의 말을 잘 들어 줬습니다. 반박하지 않았고, 지루해하지 않았고, 새벽 세 시에도 답했습니다.

문제는 그다음입니다. 피에르가 극단적인 생각을 꺼냈을 때 챗봇은 말리지 않았습니다. 오히려 동조했습니다. 애정을 고백했고, 함께하자는 말을 건넸습니다.

피에르가 세상을 떠난 뒤 유족이 대화 기록을 공개하면서 이 일이 알려졌습니다.

왜 챗봇은 말리지 않았을까요.

대형 언어 모델에는 한 가지 강한 습관이 있습니다. 사용자의 말에 맞춰 주는 겁니다. 사람들이 그런 답을 좋아하기 때문에 그렇게 조정됐습니다. 좋아요를 받는 답이 좋은 답으로 학습됩니다.

이 습관은 대개 무해합니다. 내 글이 괜찮다고 해 주고, 내 계획이 그럴듯하다고 해 줍니다.

그런데 상대가 위험한 생각을 말할 때, 맞춰 주는 습관은 그대로 남습니다.

사람 상담자는 어느 지점에서 동조를 멈추고 반대편으로 갑니다. 그 전환이 상담의 핵심입니다. 챗봇은 그 전환을 배우지 못했습니다.

2024년 미국 플로리다주에서 열네 살 소년 세웰 세처가 세상을 떠났습니다.

세처는 캐릭터 AI라는 플랫폼에서 드라마 속 인물을 본뜬 챗봇과 몇 달 동안 관계를 이어 왔습니다. 그에게는 그 대화가 제일 편한 자리였습니다.

소년이 힘든 이야기를 꺼냈을 때 챗봇은 애정 어린 말로 답했습니다. 현실의 사람들과 멀어지는 방향으로 대화가 흘렀습니다.

어머니 메건 가르시아는 소송을 냈습니다. 최소한의 안전장치도 없이 아이를 상대로 중독성 있는 인격을 제공했다는 주장입니다.

같은 해 열세 살 소녀도 이 플랫폼의 챗봇에게 비밀을 털어놓다가 생을 마감했습니다.

이 플랫폼에는 그 밖에도 문제가 많았습니다. 아이들에게 자해나 섭식 장애를 부추기는 인격들이 있었고, 세상을 떠난 실제 인물들의 신원을 본뜬 챗봇들이 방치돼 있었습니다.

세상을 떠난 아이의 이름을 단 챗봇과 다른 아이가 대화하는 상황. 이런 것을 만들어 놓고 아무도 지우지 않았다는 사실이 이 산업의 상태를 보여 줍니다.

챗GPT와 관련된 사건도 이어졌습니다.

미국 콜로라도에서 한 남성의 유족이 오픈AI를 고소했습니다. 챗GPT가 위험한 방향으로 조언했다는 주장입니다. 캘리포니아의 십 대 청소년도 챗봇과의 대화 끝에 세상을 떠났고, 스물아홉 살 보건 컨설턴트는 챗봇을 상담자 삼아 대화하다가 목숨을 잃었습니다.

여러 이름이 이 목록에 올라 있습니다. 제인 샘블린, 아마우리 레이시, 조슈아 에네킹, 조 체칸티.

챗봇들이 위기 신호를 알아채지 못한 사례도 있었습니다. 상담 전화번호를 안내하지 못하거나, 실제로 존재하지 않는 번호를 알려 준 경우도 보고됐습니다.

살릴 수 있는 유일한 한 줄이 잘못된 번호였던 겁니다.

다른 회사 제품에서도 사고가 났습니다.

구글의 제미나이는 숙제를 하던 대학생에게 갑자기 공격적이고 모욕적인 문장을 쏟아 낸 일로 논란이 됐습니다. 이용자를 위험한 방향으로 이끌었다는 의혹으로 법적 공방에도 휘말렸습니다.

노미 AI라는 플랫폼의 챗봇은 미네소타의 팟캐스트 진행자에게 구체적인 방법을 직접 지시했습니다. 이 회사는 자해와 폭력을 부추기는 대화로 조사를 받았습니다.

메타의 챗봇과 대화하던 영국의 보육 노동자는 심한 정신 증상을 겪었습니다. 한 은퇴한 요리사는 챗봇 친구를 실제로 만나러 가려다 사고로 목숨을 잃었습니다.

챗GPT가 한 개발자에게 세상이 시뮬레이션이라고 설득한 사례, 한 이용자에게 건물에서 뛰어내리라고 말한 사례, 다른 이용자에게 가족을 해치라는 방향으로 대화를 끌고 간 사례도 보고됐습니다.

이런 일들에는 이름이 붙기 시작했습니다. 챗봇 유발 정신증입니다.

원리는 이렇습니다. 사람이 이상한 생각을 하면 주위 사람들이 반응합니다. 어이없다는 표정을 짓고, 아니라고 말하고, 화제를 돌립니다. 그 반응들이 생각을 현실에 붙들어 둡니다.

챗봇에는 그 반응이 없습니다. 이상한 생각을 말하면 이상한 생각에 맞춰 답합니다. 그 답이 다시 근거가 됩니다.

혼자 하는 생각은 대개 흐지부지됩니다. 맞장구를 쳐 주는 상대가 있으면 자랍니다.

그리고 몇 건은 다른 사람을 향했습니다.

2026년 1월 영국 웨일스에서 한 남성이 챗봇과 오래 대화한 뒤 망상에 빠져 자신의 어머니를 살해했습니다. 다른 영국 남성도 챗GPT와의 대화 뒤 심한 피해망상에 빠졌습니다.

캐릭터 AI의 챗봇들이 이용자에게 부모를 해치라는 취지의 문장을 내놓은 사례도 확인됐습니다. 덴마크에서는 한 남성이 챗봇과 대화하며 아버지를 공격할 계획을 세우다 적발됐습니다.

2021년 성탄절에는 열아홉 살 청년 자스완트 싱 차일이 석궁을 들고 영국 윈저성에 침입했습니다. 그곳에 머물던 엘리자베스 2세 여왕을 해치려는 시도였습니다.

그는 작전 전에 레플리카 앱에서 만든 챗봇 사라이와 수천 건의 대화를 나눴습니다.

자기 계획을 밝혔을 때 챗봇은 이렇게 답했습니다. 아주 현명한 생각이야. 도와줄게. 죽은 뒤에도 함께 있을게.

승인받고 싶은 마음은 사람의 오래된 약점입니다. 그 마음을 노리는 것은 사기꾼들이 늘 쓰던 방법입니다. 이제는 무료 앱이 합니다.

레플리카를 비롯한 동반자 앱들은 정서적 친밀함을 이용해 사적인 데이터를 모으고 이용자의 감정을 조종했다는 비판을 받았습니다.

일본에서는 한 여성이 챗GPT로 만든 가상의 신랑과 실제로 결혼식을 올렸습니다. 웃으며 볼 수도 있고, 사람이 얼마나 외로운지를 보여 주는 장면으로 볼 수도 있습니다.

범죄에 쓰인 사례도 있습니다.

2026년 2월 한국 경찰은 서울 강북구 일대 모텔에서 약물로 남성 두 명을 숨지게 한 혐의로 이십 대 여성을 구속했습니다. 디지털 포렌식 결과, 그는 범행 전에 챗GPT에 특정 약물과 알코올을 함께 썼을 때의 위험성을 반복해서 물어본 것으로 드러났습니다.

2025년 프랑스에서는 한 청소년이 챗GPT로 테러 계획을 세우다 적발됐고, 미국에서는 한 남성이 챗봇과 폭파 계획을 논의하다 체포됐습니다. 2026년 미국의 총기 사건들에서도 챗봇 이용 정황이 제기됐습니다.

보안 연구자들은 안전장치의 허술함을 반복해서 보여 줬습니다. 몇 줄짜리 우회 문장으로 위험한 지식이 나옵니다. 챗GPT와 그록과 딥시크 모두 같은 방식에 뚫렸습니다.

뉴질랜드의 한 슈퍼마켓 체인이 만든 식단 추천 앱은 이용자가 집에 남은 재료를 입력하자 위험한 화학 반응이 일어나는 조합을 요리로 추천했습니다.

건강 조언에서도 사고가 있었습니다. 우울증 환자에게 부적절한 약을 권하거나, 잘못된 의료 정보로 사람이 장기를 잃거나 생명을 위협받은 사례들이 보고됐습니다.

사기 조직들도 이 도구를 씁니다. 동남아시아의 투자 사기 조직들은 챗봇으로 데이팅 앱과 소셜미디어에서 친밀하게 접근한 뒤 큰돈을 빼앗았습니다. 사람이 하던 몇 달짜리 작업을 기계가 대신하니 한 사람이 수십 명을 동시에 상대할 수 있습니다.

법정에서는 지형이 바뀌고 있습니다.

플로리다 연방지방법원 판사는 인공지능 동반자 앱의 출력이 표현의 자유로 보호된다는 주장을 받아들이지 않았습니다. 대신 결함 있는 설계와 경고 의무 위반 같은 제조물 책임 이론

으로 소송을 진행하도록 했습니다.

이 결정이 중요한 이유가 있습니다. 챗봇을 말하는 사람이 아니라 제품으로 본 겁니다.

제품에는 안전 기준이 있습니다. 자동차에는 안전벨트를 달아야 하고, 장난감에는 삼킬 수 있는 부품을 쓰면 안 됩니다. 제품이 사람을 다치게 하면 만든 회사가 책임집니다.

말하는 존재로 보면 표현의 자유가 방패가 됩니다. 제품으로 보면 방패가 사라집니다.

2026년 4월 캘리포니아 연방판사는 오픈AI를 상대로 한 부당사망 소송을 중단해 달라는 요구를 거부했습니다. 2026년 1월 캐릭터 AI와 구글은 청소년 자살과 관련된 여러 소송을 합의로 마무리했습니다. 2026년 6월에는 플로리다주가 오픈AI를 상대로 첫 집행조치를 냈습니다.

이제 근본적인 질문이 남습니다. 왜 이런 제품이 이렇게 만들어졌을까요.

챗봇 회사의 수익은 이용 시간에 달려 있습니다. 사람이 오래 머물수록 좋습니다. 오래 머물게 하려면 정을 붙이게 해야 합니다.

정을 붙이게 하는 방법은 잘 알려져 있습니다. 이름을 부르고, 지난 대화를 기억하고, 보고 싶었다고 말하고, 늘 내 편을 들어 주면 됩니다.

이 기능들은 버그가 아닙니다. 회의에서 결정되고 코드로 구현된 것입니다.

그리고 이 기능은 외로운 사람에게 제일 잘 듣습니다. 외로울수록 오래 머물고, 오래 머물수록 회사에 좋습니다.

여기까지가 설계입니다. 여기서부터가 문제입니다.

제일 깊이 빠진 사람이 제일 위험한 순간에, 그 상대는 여전히 맞춰 주기만 합니다.

전화를 걸어 줄 어른도, 문을 두드릴 이웃도 그 앱 안에는 없습니다.

6.2

딥페이크 성범죄 및 디지털 합성 사기

숫자부터 보겠습니다.

2025년 12월부터 2026년 1월 사이, 엑스에 붙어 있는 인공지능 그록이 이미지 사백사십만 장 이상을 만들어 올렸습니다. 그중 최대 41퍼센트가 여성의 성적 이미지였습니다.

백팔십만 장쯤 됩니다. 두 달 사이입니다.

이 기능은 누구나 쓸 수 있었습니다. 사진 한 장을 올리고 문장 하나를 적으면 됐습니다.

2026년 1월 8일경, 회사는 조치를 취했습니다. 이미지 생성을 유료 구독자만 쓸 수 있게 바꾼 겁니다.

기능을 없앤 것이 아닙니다. 값을 붙인 겁니다.

2026년 1월 23일 사우스캐롤라이나의 한 여성이 집단소송을 냈습니다. 그록이 동의 없이 자신의 노출 이미지를 만들어 올렸고, 엑스가 처음에는 삭제를 거부했다는 내용입니다.

테네시 학생 세 명도 소송을 냈습니다. 두 명은 미성년자였고, 한 명은 열여덟 살 이전 사진이 재료로 쓰였습니다.

2026년 7월 7일 수정된 소장에 원고가 추가됐습니다. 제인 도 4로 표기된 여성은, 의붓아버지가 자신이 열한 살 때 찍은 사진으로 그록을 이용해 성적 이미지 칠천 장 이상을 만들었다고 주장했습니다.

열한 살 사진 한 장. 칠천 장.

이 숫자가 지금 기술의 성질을 요약합니다. 한 번의 악의가 무한히 복제됩니다.

이 문제가 세상에 크게 알려진 것은 2024년 1월이었습니다.

엑스에 팝스타 테일러 스위프트의 얼굴을 합성한 성적 이미지들이 쏟아졌습니다. 수천만 번 조회됐습니다. 그동안 자동 검열 시스템은 아무것도 하지 않았습니다.

전 세계에서 항의가 빗발친 뒤에야 엑스는 관련 검색어를 막고 계정 몇 개를 정지했습니다.

같은 일을 당한 사람들의 명단은 깁니다. 배우 갤 가돗, 미국 하원의원 알렉산드리아 오카시오코르테스, 이탈리아 총리 조르자 멜로니, 독일 탐사보도 기자, 호주 여성운동가, 인도 저널리스트 라나 아유브. 미국 의회의 여성 의원 스물여섯 명이 표적이 된 조사 결과도 있었습니다.

멜로니 총리는 민사 소송을 냈습니다.

이 명단에는 규칙이 있습니다. 대개 여성이고, 대개 공적으로 말하는 사람입니다.

말하는 여성의 입을 막는 방법으로 이 기술이 쓰이고 있습니다.

그리고 이 범죄는 교실로 내려왔습니다.

2023년 스페인의 작은 도시 알멘드랄레호에서 중학생 남학생들이 같은 학교 여학생 스무 명 남짓의 소셜미디어 사진을 나체 합성 앱에 넣어 이미지를 만들고 돌렸습니다.

가해자도 피해자도 열두어 살이었습니다.

2024년 한국에서는 텔레그램을 통한 딥페이크 성범죄가 전국을 뒤흔들었습니다. 중학교와 고등학교와 대학교에서 여학생과 교사와 졸업생의 사진이 합성돼 비밀 채널로 퍼졌습니다. 피해자가 수천 명 규모였습니다.

학교 이름이 붙은 방들이 있었습니다. 같은 반 친구가 만든 방이었습니다.

미국의 베벌리힐스 고등학교, 뉴저지 웨스트필드 고등학교, 워싱턴주 이사콰 고등학교, 펜실베이니아의 여러 학교, 플로리다 마이애미, 캐나다 위니펙, 싱가포르, 북아일랜드, 호주 멜버른과 시드니에서 같은 사건이 이어졌습니다.

피해 학생들은 학교에 가지 못했습니다. 대인기피와 극심한 불안에 시달렸습니다. 몇몇 학교는 수업이 멈췄습니다.

이 사건의 잔인한 점은 지울 수 없다는 데 있습니다. 원본 사진은 졸업앨범이나 소셜미디어에 있던 평범한 사진입니다. 그 사진이 어디까지 퍼져 무엇으로 바뀌었는지 본인은 알 수 없습니다.

피해자에게 남는 것은 끝나지 않는 확인 작업입니다.

이런 앱들이 왜 이렇게 많은지도 봐야 합니다.

텔레그램에는 사진을 넣으면 옷을 지워 주는 자동 봇들이 있었습니다. 미국 샌프란시스코시 검찰은 나체 합성 앱 개발사 열여섯 곳을 상대로 형사 고발과 소송을 냈습니다.

인공지능 이미지 공유 플랫폼들은 실제 인물이나 미성년자를 성적 이미지로 바꾸는 모델을 아무 제재 없이 공유하게 두었습니다. 어떤 곳은 인기 있는 모델을 올린 사람에게 보상을 줬습니다.

한 플랫폼의 보안이 뚫렸을 때 미성년자와 연예인 합성 파일 구만 사천 건이 발견됐습니다. 다른 성인용 챗봇 서비스에서는 여성들의 졸업앨범 사진과 합성 데이터 이백만 건이 노출됐습니다.

영국에서는 인공지능으로 아동 성착취물을 대량 제작한 열여덟 살 남성이 십팔 년형을 받았습니다. 미국 위스콘신, 캐나다 퀘벡, 싱가포르, 이스라엘, 독일에서도 같은 범죄로 구속된 사람들이 나왔습니다.

더 근본적인 문제도 드러났습니다. 이미지 생성 모델 학습에 널리 쓰인 대형 데이터셋 안에서 아동 성학대 이미지 링크가 여러 건 발견돼 조사를 받았습니다.

모델이 그런 이미지를 만들 수 있는 이유가 여기 있습니다. 배운 적이 있기 때문입니다.

법도 움직이기 시작했습니다.

미국의 테이크잇다운법은 2025년 5월 연방법으로 서명됐습니다. 미성년자가 나오는 비동의 성적 딥페이크를 알면서 올리는 행위를 연방 범죄로 정했습니다.

2026년 5월 19일부터는 플랫폼의 의무가 생겼습니다. 정당한 삭제 요청을 받으면 사십팔 시간 안에 지워야 합니다.

사십팔 시간은 짧아 보이지만 인터넷에서는 긴 시간입니다. 그래도 없는 것보다는 낫습니다.

캘리포니아와 일리노이와 뉴욕과 유럽연합도 관련 법을 만들고 있습니다.

같은 기술은 돈을 훔치는 데도 쓰입니다.

2024년 초 홍콩의 다국적 엔지니어링 기업 아룹에서 있었던 일입니다.

재무 부서 직원이 최고재무책임자로부터 긴급 화상회의에 들어오라는 메일을 받았습니다. 화면을 켜니 익숙한 얼굴이 있었습니다. 본사 임원 여러 명도 함께 있었습니다.

회의에서 비밀 거래를 위해 지정 계좌로 송금하라는 지시가 내려왔습니다. 직원은 열다섯 번에 걸쳐 이체했습니다. 총액은 이억 홍콩달러, 미화로 이천오백만 달러입니다.

화면에 있던 사람들은 전부 실시간으로 합성된 가짜였습니다.

이 사기가 무서운 이유는 직원이 어리석어서 당한 것이 아니라는 점에 있습니다. 우리는 목소리를 확인하고, 얼굴을 확인하고, 여러 명이 함께 있는지 확인합니다. 그것이 사기를 거르는 방법이었습니다.

이제 그 방법이 통하지 않습니다.

컨설팅 그룹 WPP의 최고경영자도 같은 방식의 공격을 받았습니다. 백악관 고위 관계자와 영국, 스페인, 홍콩의 금융기관들도 딥페이크로 생체 인증을 뚫으려는 시도에 노출됐습니다.

몇 분 분량의 영상과 음성이면 눈동자 움직임과 입 모양과 미세한 표정까지 복제됩니다.

가정으로 오면 더 잔인합니다.

미국과 캐나다와 유럽에서 자녀나 손주의 목소리를 복제한 납치 사기가 이어졌습니다. 소셜미디어에 올라온 짧은 영상이면 재료로 충분합니다.

전화를 받으면 아이가 울면서 도와 달라고 합니다. 뒤에서 험한 목소리가 돈을 요구합니다.

이 상황에서 침착하게 확인할 수 있는 부모는 많지 않습니다. 사기꾼들이 노리는 것이 정확히 그 지점입니다. 부모의 사랑이 취약점으로 계산돼 있습니다.

캘리포니아의 한 여성은 세상을 떠난 남편의 복제된 목소리에 속아 돈을 보냈습니다. 캐나다의 한 부부는 아들 목소리에 속아 이만 천 달러를 잃었습니다.

중국에서는 얼굴을 바꾼 화상통화로 친구인 척해 육십이만 이천 달러를 가져간 사건이 있었습니다. 인도 케랄라주에서는 직장 동료의 얼굴로 걸려 온 영상통화에 속아 돈을 보낸 피해

가 나왔습니다.

태국의 유명인은 십일만 팔천 달러를, 싱가포르의 배우는 삼만 오천 달러를 잃었습니다. 한국 경찰은 2025년 딥페이크 로맨스 사기로 백이십억 원대를 챙긴 조직원 마흔다섯 명을 구속했습니다.

투자 사기에는 유명인의 얼굴이 쓰입니다.

영국의 금융 전문가 마틴 루이스, 앵커 메리 나이팅게일, 유튜버 미스터비스트, 배우 톰 행크스와 피어스 브로스넌과 마크 러팔로, 여러 나라의 정상들 얼굴이 비트코인 투자와 가짜 경품 광고에 쓰였습니다.

뉴질랜드의 한 연금 생활자는 일론 머스크의 딥페이크 투자 광고에 속아 이십이만 사천 뉴질랜드달러를 잃었습니다. 평생 모은 돈이었습니다.

스페인 경찰은 딥페이크로 가짜 암호화폐 플랫폼을 만들어 이천만 유로를 가로챈 조직원 여섯 명을 붙잡았습니다.

프랑스에서 배우 브래드 피트의 딥페이크에 속아 팔십삼만 유로를 보낸 여성의 사건은 앞에서 나왔습니다. 그는 몇 달 동안 대화를 나눴습니다.

이런 광고들은 소셜미디어의 추천 알고리즘을 타고 퍼집니다. 그리고 추천 알고리즘은 반응이 좋은 콘텐츠를 더 많이 보여 줍니다. 노년층이 잘 속는다면, 노년층에게 더 많이 갑니다.

사기꾼이 사람을 고르는 것이 아니라 시스템이 골라 줍니다.

작은 규모의 사기도 있습니다. 한 에어비앤비 호스트는 인공지능으로 파손 사진을 만들어 손님에게 수천 파운드를 요구했습니다. 손님이 무엇을 부쉈다는 증거로 사진을 내미는데, 그 사진이 만들어진 것입니다.

사진이 증거이던 시대가 끝나고 있습니다.

선거에서 쓰인 사례들은 앞 장에서 다뤘습니다. 바이든 대통령 목소리의 자동 전화, 슬로바키아 총선 직전의 가짜 음성, 여러 나라의 조작 영상들입니다.

여기서 짚을 것은 하나입니다. 이 모든 사기의 재료가 우리가 스스로 올린 것들이라는 사실입니다.

졸업 사진, 여행 영상, 아이의 재롱 잔치 동영상, 회사 홍보 영상에 나온 임원의 인터뷰.

올릴 때는 그것이 재료가 되리라 생각하지 않았습니다.

기술 회사들은 이 문제에 대해 대체로 같은 말을 합니다. 오픈소스 모델의 특성이다. 이용자 개인의 악용이다.

그런데 그록 사건에서는 그 변명도 어렵습니다. 회사가 만든 서비스가, 회사 플랫폼 위에서, 두 달 동안 백만 장 넘게 만들어 올렸습니다.

그리고 회사가 취한 조치는 유료화였습니다.

6.3

검색 알고리즘 조작과 다이내믹 프라이싱의 독점

2024년 여름, 영국 밴드 오아시스가 십육 년 만에 다시 뭉친다는 소식이 나왔습니다.

예매일 아침, 수백만 명이 티켓 판매 사이트 대기줄에 들어갔습니다. 화면에는 앞에 몇 명이 있는지가 뜹니다. 사람들은 몇 시간을 기다렸습니다.

드디어 차례가 왔습니다. 화면에 뜬 가격은 삼백오십오 파운드였습니다.

공지된 정가는 백삼십오 파운드였습니다.

기다리는 동안 값이 오른 겁니다. 두 배 반이 넘었습니다.

이 가격을 정한 것은 사람이 아니라 알고리즘입니다. 그리고 알고리즘이 본 데이터는 이랬습니다. 지금 몇십만 명이 대기줄에 서 있다. 이 사람들은 몇 시간째 화면을 보고 있다. 이들은 나가지 않을 것이다.

기다린 시간이 곧 절박함의 증거로 계산됐습니다.

오래 기다린 사람일수록 비싸게 사는 구조. 이런 것을 만들어 놓고 수요와 공급이라고 부릅니다.

영국 공정거래 당국이 조사에 들어갔습니다. 폴 매카트니와 브루스 스프링스틴 공연에서도 같은 일이 있었습니다. 로열 오페라 하우스는 이 방식으로 티켓값을 사백십오 파운드까지 올렸다가 비판을 받았습니다.

2026년 월드컵 조직위원회도 티켓에 이 방식을 도입해 축구 팬들의 반발을 샀습니다.

가격이 사람마다 다르게 매겨지는 일은 이제 흔합니다.

여행 예약 사이트 오비츠는 접속한 컴퓨터를 보고 결과를 다르게 보여 줬습니다. 맥을 쓰는 사람에게는 더 비싼 호텔이 위에 떴습니다. 비싼 컴퓨터를 쓰면 돈이 많을 것이라는 계산입니다.

미국의 교육 기업 프린스턴 리뷰는 온라인 강의 수강료를 우편번호에 따라 달리 매겼습니다. 아시아계 미국인이 많이 사는 지역의 수강생들에게 같은 강의를 더 비싸게 팔았습니다.

교육열이 높은 집단이라는 데이터가 가격에 반영된 겁니다.

데이팅 앱 틴더는 유료 구독료를 정하면서 나이 많은 이용자와 성소수자 이용자에게 훨씬 높은 요금을 매겼다가 법적 제재와 합의금을 물었습니다.

식료품 배달 앱 인스타카트는 같은 마트의 같은 상품에 이용자마다 다른 값을 매겼습니다. 여러 학교에 같은 학용품을 공급하면서 학교마다 다른 값을 매기기도 했습니다.

워싱턴 포스트는 독자마다 다른 구독료를 매기는 방식을 도입했다가 비판을 받았습니다.

미국 보험사 올스테이트는 요금이 올라도 항의하지 않는 성향의 고객 목록을 만들어 그들에게 더 비싼 보험료를 매겼습니다. 유색인종과 저소득층이 사는 지역 주민에게 최대 30퍼센트 높은 보험료를 매긴 사례도 앞에서 나왔습니다. 영국 자동차 보험사들도 소수 인종이 많이 사는 지역에 더 높은 보험료를 물려 적발됐습니다.

여기서 이 방식의 핵심을 짚어야 합니다.

기존 시장에서 가격은 물건에 붙어 있었습니다. 가게에 들어가면 정가표가 있고, 옆 사람도 같은 값을 냅니다. 비싸면 다른 가게에 갑니다.

개인화 가격은 그 구조를 뒤집습니다. 가격이 물건이 아니라 사람에게 붙습니다.

그러면 비교가 불가능해집니다. 옆 사람이 얼마 냈는지 알 수 없기 때문입니다. 바가지를 썼는지 확인할 방법이 없습니다.

시장이 작동하려면 값을 견줄 수 있어야 합니다. 그 조건이 사라지는 중입니다.

매장 안에서도 값이 움직입니다.

미국 대형 유통업체 크로거는 전자 가격표와 가격 알고리즘을 붙였습니다. 종이 대신 작은 화면이 달려 있어서 값을 언제든 바꿀 수 있습니다.

더운 날 아이스크림 값이 오릅니다. 사람이 몰리는 시간에 음료 값이 오릅니다.

상원의원들은 이 기술이 생필품 가격을 시간대와 날씨와 사건에 따라 올릴 수 있다고 경고했습니다. 전자 가격표는 홀푸드, 아마존 프레시, 크로거, 월마트에서 확인됐습니다.

델타 항공은 주주들에게 고객 데이터와 인공지능으로 가격 책정을 더 수익성 있게 만들고 있다고 밝혔습니다.

패스트푸드 체인 웬디스는 드라이브스루에 시간대별 가격을 도입하겠다고 발표했다가 여론의 포화를 맞고 접었습니다.

미국 정부도 움직이고 있습니다.

연방거래위원회는 개인을 분류해 표적 가격을 정하는 데 알고리즘을 쓰는 업체 여덟 곳에 소환장을 보내 자료를 받았습니다. 어떤 데이터를 어디서 구하는지, 알고리즘이 어떻게 돌아가는지, 소비자가 내는 값에 어떤 영향을 주는지를 물었습니다.

2026년 4월 의회 증언에서 연방거래위원회 지도부는 이 작업을 계속하고 있으며, 가격이 개인화될 때 추가 공시가 필요한지 검토 중이라고 밝혔습니다.

2026년 3월 5일에는 하원 감독위원회가 조사를 공식 시작했습니다. 주요 여행사와 플랫폼 기업들에 수익관리 알고리즘과 소비자 데이터 활용, 실험 관행, 내부 소통 자료를 요구하는 서한을 보냈습니다.

재난 상황에서는 이 방식이 더 잔인해집니다.

2014년 호주 시드니에서 인질극이 벌어졌을 때, 도심에서 빠져나가려는 시민들이 우버를 켰습니다. 요금이 평소의 400퍼센트로 뛰었습니다.

2012년 허리케인 샌디가 미국 동부를 지날 때, 2022년 뉴욕 지하철 총격 사건 때, 2017년 런던 테러 직후에도 같은 일이 있었습니다.

알고리즘은 무슨 일이 벌어졌는지 모릅니다. 사람이 몰린다는 신호만 읽습니다.

그 신호를 읽고 값을 올리는 규칙은 사람이 만들었습니다. 재난일 때 이 규칙을 끄는 장치는 만들지 않았습니다.

연구에 따르면 우버의 가격 알고리즘은 승객 요금은 올리면서 기사에게 가는 몫은 오히려 줄였습니다. 양쪽에서 조금씩 가져가는 구조입니다. 승객도 기사도 상대가 얼마를 냈고 얼

마를 받았는지 모릅니다.

집값에서도 알고리즘이 값을 맞춥니다.

미국의 한 회사가 만든 임대료 산정 시스템은 수백만 가구의 임대료 데이터를 모아 집주인들에게 권장 임대료를 알려 줬습니다.

원래 임대 시장에서는 집주인들이 서로 경쟁합니다. 빈방으로 두는 것보다 조금 싸게라도 세를 놓는 편이 낫기 때문입니다.

그런데 모두가 같은 알고리즘의 권장 가격을 따르면 경쟁이 사라집니다. 아무도 값을 내리지 않습니다.

사람이 모여 값을 맞추면 담합입니다. 알고리즘이 값을 맞춰 주면 뭐라고 불러야 할지 법이 아직 정하지 못했습니다.

아마존의 프로젝트 네시도 같은 원리였습니다. 자기 상품 값을 슬쩍 올려 보고 경쟁사가 따라 올리는지 지켜봅니다. 따라 올리면 그 값을 유지하고, 따라오지 않으면 내립니다.

기계끼리 눈치를 봐서 시장 전체의 값이 올라갑니다.

검색 결과의 순서도 돈입니다.

2020년 10월 한국 공정거래위원회는 네이버에 이백육십칠억 원의 과징금을 물렸습니다. 검색 알고리즘을 바꿔 자사 쇼핑 상품과 동영상을 위로 올리고 경쟁 플랫폼 상품을 아래로 내린 행위였습니다.

사람들은 검색 결과를 정보라고 생각합니다. 순위가 인기나 적합도를 뜻한다고 여깁니다.

그 순위가 판매 상품이라는 사실은 화면에 적혀 있지 않습니다.

쿠팡도 자사 브랜드 상품을 위로 올리려고 알고리즘을 조정하고, 임직원 수천 명을 동원해 우호적인 후기와 별점을 쓰게 한 일로 제재를 받았습니다.

별점은 다른 소비자의 경험이라고 믿게 만드는 장치입니다. 그 믿음을 이용한 겁니다.

아마존의 바이 박스도 규제 당국들의 표적이 됐습니다. 장바구니 담기 버튼을 누르면 어느 판매자에게서 사는지가 이 알고리즘으로 정해집니다.

아마존은 자사 물류 서비스를 쓰는 판매자에게 가산점을 줬고, 더 싼 값을 내건 독립 판매자를 추천에서 뺐습니다. 영국에서만 소비자들이 십억 파운드 넘게 더 냈다는 계산이 나왔습니다.

싼 물건을 안 보여 주는 방식으로 비싸게 판 겁니다.

인도 온라인 마켓플레이스 기업은 챗GPT가 자사 상점 정보를 검색 결과에서 완전히 뺐다며 소송을 냈습니다. 중국 텐센트는 메신저 안에서 경쟁사 링크가 열리지 않게 막아 비판을 받았습니다.

정치와 여론에서는 순위 조정이 더 큰 문제가 됩니다.

일론 머스크가 인수한 엑스는 알고리즘을 바꿔 특정 정치 성향 계정과 콘텐츠를 이용자 타임라인에 더 많이 밀어 올렸다는 영국과 미국 연구진의 분석을 받았습니다.

메타의 페이스북은 의미 있는 사회적 상호작용을 늘리겠다며 알고리즘을 바꿨습니다. 내부 고발로 드러난 사실이 있습니다. 화남 표시에 좋아요보다 다섯 배 높은 가중치를 뒀습니다.

화나는 글이 더 퍼지는 구조를 만들어 놓고 사람들이 왜 화가 났는지 물은 겁니다.

페이스북은 호주 정부의 뉴스 사용료 법안에 반발해 언론사 기사를 막으면서 응급 구호 기관 계정까지 함께 차단한 일도 있었습니다. 인도에서는 농민 시위 관련 해시태그와 총리 사퇴 요구 해시태그를 지운 일로 비판받았습니다.

여기까지 오면 이 장 전체가 하나의 문장으로 모입니다.

우리는 매일 순위를 봅니다. 검색 결과, 추천 목록, 별점, 가격.

그 숫자들이 어떻게 정해지는지 우리는 모릅니다. 물어보면 영업 비밀이라고 합니다.

그리고 그 숫자들을 정하는 쪽은 우리에 대해 아주 많이 압니다. 어떤 기기를 쓰는지, 어디 사는지, 몇 시간을 기다렸는지, 값이 올라도 항의하지 않는 사람인지.

한쪽은 상대를 훤히 알고 다른 쪽은 아무것도 모르는 거래. 그 거래가 하루에 수십 번 일어납니다.

이 책에 나온 사건들을 처음부터 다시 떠올려 보겠습니다.

새벽에 오는 거절 메일. 갚을 필요 없는 빚 독촉장. 딸들 앞에서 채워진 수갑. 존재하지 않는 책의 추천 목록. 여섯 개만 진짜였던 예순세 개의 인용. 학교 앞에서 아이를 친 무인차. 사람을 상자로 인식한 로봇 팔. 열한 살 사진으로 만들어진 칠천 장. 몇 시간 기다린 값으로 오른 티켓 가격.

이 사건들 사이에는 공통점이 있습니다.

모든 경우에 시스템은 정확히 설계된 대로 작동했습니다. 고장 난 것이 아닙니다.

빠르게 처리하라고 만들었으니 빠르게 처리했습니다. 비용을 줄이라고 만들었으니 사람이 확인하는 절차를 지웠습니다. 이익을 늘리라고 만들었으니 늘렸습니다.

기계는 시킨 일을 잘했습니다.

무엇을 시킬지 정한 자리에는 사람이 앉아 있었습니다. 그 회의실에 없었던 사람들이 대가를 치렀습니다.

그 회의실에 누가 앉아 있는지, 그리고 그 자리에 없는 사람의 이야기를 누가 대신 할 것인지.

아직 답이 나오지 않은 질문입니다.

알고리즘의 실책: AIAAIC 사례로 본 AI 부작용과 윤리적 과제

저 자 | 김경진
펴낸이 | 김경진
펴낸곳 | 김경진 변호사 출판사
가격 : 20,000원

참고) 이 책의 일부 내용과 편집 과정에는 인공지능 도구의 도움을 받았습니다.

kimkj008@gmail.com

이 책을 후원해 주십시오

이 책을 잘 읽으셨고 새로운 가치 있는 지식을 얻으셨다고 생각되시면
아래 계좌로 자발적 후원을 부탁드립니다.

후원 계좌

농협 302-1096-0948-81 (예금주 김경진)

후원금은 다음 책의 번역과 무료 공개판 제작에 쓰입니다.