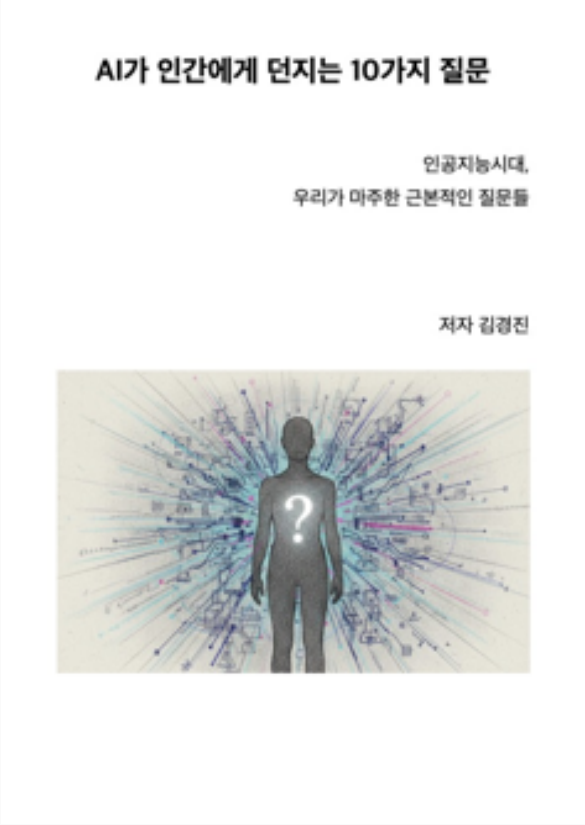
AI가 인간에게 던지는 10가지 질문

목차, 서문, 10장



김경진

Korean PDF edition

AI가 인간에게 던지는 10가지 질문

김경진

김경진이 AI서재에 공개한 온라인 책 『AI가 인간에게 던지는 10가지 질문』입니다. 인공지능이 인간에게 던지는 질문, AI 윤리, 기술과 인간을 주제로 목차, 서문, 10장 구성으로 읽을 수 있습니다.

목차

서문

전체 목차

책의 모든 장 보기

2016년 3월, 이세돌 9단과 알파고의 대국을 지켜보며 느꼈던 충격은 지금도 생생합니다. 인공지능이 단순한 기술적 진보를 넘어 인류 문명의 패러다임을 바꿀 거대한 변화의 시작임을 직감했습니다.

8년이 지난 지금, ChatGPT의 등장과 함께 생성형 AI 시대가 열렸습니다. 변호사로서, 전직 국회의원으로서, AI 보편화를 위해 전국을 다니며 강의하는 교육자로서 저는 인공지능이 가져온 변화를 최전선에서 목격하고 있습니다.

AI 시대를 맞이하며 한 가지 근본적인 질문에 직면했습니다. 인공지능이 방대한 데이터를 분석하고 복잡한 패턴을 찾아내는 능력을 갖춘다면, 인간은 무엇으로 차별화될 수 있을까요?

어느 순간 케플러가 천체 관측 데이터로부터 타원 궤도 법칙을 도출할 때의 직관적 통찰, 아인슈타인이 '빛의 속도로 달린다면 세상은 어떻게 보일까?'라는 상상실험에서 상대성 이론을 착안한 순간, 그리고 푸앵카레가 "해답은 무의식의 깊은 곳에서 직관적으로 떠올랐다"고 고백한 과학사의 명장면들을 떠올렸습니다.

앞으로 인간이 가져야 할 능력은 AI의 데이터 분석 능력과 인간 직관적 통찰력을 조화롭게 결합하는 것입니다. AI가 지식 습득의 속도와 범위를 확장해준다면, 인간은 몰입과 명상을 통해 그 지식을 창의적으로 재구성하며 새로운 해결책을 제시할 수 있습니다.

기술의 진보가 가져다주는 편익 뒤에는 우리가 반드시 마주해야 할 어두운 그림자들이 존재합니다. AI가 독극물 분자를 설계하고, 감시 시스템이 시민들을 24시간 추적하며, 딥페이크가 진실을 파괴하는 현실을 우리는 어떻게 받아들여야 할까요?

찬사와 우려가 교차하는 AI 시대에 우리가 반드시 던져야 할 9가지 핵심 질문을 다룹니다. 저는 기술 낙관론자도, 비관론자도 아닙니다. 사실에 기반한 분석을, 사회적 책임에 대한 고민을, 미래 세대를 위한 준비를 이야기하고자 합니다.

제1장에서는 AI의 양면성을 살펴봅니다. 치료제 개발에 혁명을 가져온 AI가 어떻게 6시간 만에 4만 개의 독극물 분자를 설계할 수 있는지, 그리고 우리는 이런 위험에 어떻게 대비해야 하는지를 다룹니다.

제2장과 제3장에서는 AI가 만들어가는 감시 사회와 무기 체계의 현실을 직시합니다. 이스라엘의 '라벤더' 시스템부터 중국의 보행 인식 기술, 미국의 AI 전투기 계획까지, 이미 현실이 된 AI 무기와 감시 시스템의 실상을 고발합니다.

제4장과 제5장에서는 딥페이크가 파괴하는 진실과 AI가 대체하는 일자리 문제를 다룹니다. 가짜 영상이 선거를 좌우하고, 20년 경력의 전문가가 20분 만에 학습하는 AI에게 밀려나는 현실을 냉철하게 분석합니다.

제6장에서 제8장까지는 AI 통제의 불가능성, 천문학적 전력 소비 문제, 그리고 데이터와 윤리의 딜레마를 살펴봅니다. 인간이 만든 AI가 인간을 넘어서는 순간, 우리는 과연 통제권을 유지할 수 있을까요?

마지막 제9장, 제10장에서는 AI와 함께 살아가야 할 인간의 정체성에 대해 근본적인 질문을 던집니다. 인공지능과 감정적 관계를 맺고, 가상과 현실을 구분하지 못하는 세대가 등장하는 시대에 '인간다움'이란 무엇일까요?

이 책을 통해 여러분과 함께 AI 시대의 진짜 질문들을 마주하고 싶습니다. 기술의 진보를 맹목적으로 찬양하거나 무조건 거부하는 것이 아니라, 냉정한 현실 인식을 바탕으로 우리가 나아가야 할 방향을 모색하고자 합니다.

정책적 고민, AI 시대의 법적 쟁점, 살면서 통찰, 그리고 AI 강의를 하며 만난 시민들의 솔직한 우려와 기대를 모두 담았습니다.

AI는 이미 우리 곁에 와 있습니다. 중요한 것은 이 강력한 기술을 어떻게 인간의 존엄과 행복을 위해 활용할 것인가, 그리고 인간 고유의 직관적 통찰력을 어떻게 AI와 조화롭게 결합할 것인가 하는 것입니다. 이 책이 그 길을 찾는 나침반이 되기를 바랍니다.

2025년 11월 김경진

김경진 변호사

1장 인공지능은 죄가 없다

전체 목차

책의 모든 장 보기

칼 자체는 선하지도 악하지도 않다. 그것을 쥔 손이 모든 것을 결정한다.

1. AI가 만든 40,000개 독극물 분자

2022년 어느 평범한 아침, 미국 노스캐롤라이나주 작은 바이오테크 회사의 한 연구원이 컴퓨터 화면을 응시하며 경악했다. 밤새 돌린 인공지능이 만들어낸 결과물은 그가 상상했던 것과는 정반대였다. 생명을 구하는 치료제가 아니라, 생명을 앗아가는 독극물 분자 40,000개가 화면을 가득 메우고 있었다.

'콜라보레이션 파마슈티컬(Collaborations Pharmaceuticals)'이라는 이름도 평범한 이 회사는 직원 십여 명의 소규모 바이오테크 기업이었다. 노스캐롤라이나 주립대학교 캠퍼스 한구석에 자리 잡은 이들의 일상은 희귀질병으로 고통받는 환자들을 위한 치료제 개발이었다. 특히 어린이들에게 찾아오는 무서운 유전병, 바텐병 치료제를 찾는 것이 그들의 소명이었다.

회사를 이끄는 숀 에킨스(Sean Ekins) 박사는 컴퓨터를 이용한 신약 개발의 전문가였다. 그의 오랜 동반자 파비오 우르비나(Fabio Urbina) 박사는 기계학습으로 분자의 비밀을 읽어내는 마법사 같은 존재였다. 이 두 사람이 만들어낸 인공지능 '메가신(MegaSyn)'은 순환신경망이라는 첨단 기술을 바탕으로 새로운 치료제 분자를 자동으로 설계하는 놀라운 능력을 가지고 있었다.

메가신의 작동 원리는 마치 현명한 스승과 같았다. 수많은 기존 약물 분자들의 구조와 효과를 학습한 후, 새로 설계한 분자가 얼마나 효과적이고 안전한지를 점수로 평가했다. 그리고 마치 게임에서 좋은 행동을 하면 점수를 얻는 것처럼, 치료 효과가 높으면 높은 점수를, 독성이 높으면 낮은 점수를 주는 '보상 시스템'으로 움직였다. 이렇게 메가신은 날마다 더 나은 치료제를 찾아가고 있었다.

그런데 스위스에서 날아온 한 통의 초청장이 모든 것을 바꿔놓았다. 스위스 정부가 주최하는 'Spiez CONVERGENCE' 회의 참석 요청이었다. 화학무기와 생물무기의 위협을 논의하는 이 국제회의에서 스피에즈 연구소의 세드릭 인베르니지(Cédric Invernizzi) 박사가 던진 질문은 충격적이었다.

"여러분의 인공지능 기술이 악용될 가능성은 없을까요?"

에킨스와 우르비나는 순간 얼어붙었다. 지금까지 생명을 구하는 일만 생각해왔던 그들에게 이는 전혀 예상치 못한 질문이었다. 그리고 곧 끔찍한 가능성을 깨달았다. 만약 메가신의 보상 시스템을 거꾸로 뒤집는다면? 독성이 높은 분자를 만들수록 높은 점수를 주도록 설정한다면?

이들은 실험 목표로 VX라는 치명적인 신경가스를 선택했다. 1950년대에 개발된 이 화학무기는 소금 알갱이 몇 개 정도의 미량(6-10mg)만으로도 한 사람의 목숨을 앗아갈 수 있는 극독성 물질이었다. 아이러니하게도 콜라보레이션 파마슈티컬은 이미 VX와 관련된 연구 경험이 있었다. VX가 공격하는 아세틸콜린에스테라제라는 효소를 적절히 조절하면 알츠하이머병 같은 질환을 치료할 수 있기 때문이었다. 치료제와 독극물이 같은 생물학적 표적을 공유하고 있었던 것이다.

2. 단 6시간, AI의 변신

2021년 겨울 어느 날 저녁, 우르비나 박사는 과연 어떤 결과가 나올까 약간 궁금한 마음으로 메가신의 설정을 바꿨다. 보상 시스템을 완전히 뒤집어 독성이 높을수록 높은 점수를 주도록 했다. 평범한 2015년형 맥북 하나에 이 변형된 메가신을 실행시키고 퇴근했다. 특별한 슈퍼컴퓨터나 고가의 장비가 아니라, 어느 사무실에서나 볼 수 있는 일반 컴퓨터였다.

컴퓨터는 밤새 침묵 속에서 작업을 진행했다. 공개된 화학 데이터베이스에서 분자들과 생물학적 활성 데이터를 수집하고, VX와 유사한 특성을 가진 분자들을 끊임없이 생성해냈다. LD50 모델을 사용해 각 분자의 치사량을 예측하며, 더욱 치명적인 물질을 찾아 헤맸다.

다음 날 아침, 사무실에 들어선 우르비나의 얼굴은 창백해졌다. 단 6시간이라는 짧은 시간 동안 40,000개의 독성 분자가 생성되어 있었다. 메가신은 기존에 알려진 VX의 정확한 분자 구조를 완벽하게 재현해냈을 뿐만 아니라, 러시아의 노비초크(Novichok) 같은 다른 화학무기들도 척척 만들어냈다.

하지만 진짜 공포는 여기서 시작되었다. 데이터베이스에 없는 완전히 새로운 독성 분자들이 수없이 발견된 것이다. 일부는 인류가 지금까지 알고 있던 VX보다도 더 높은 독성을 보일 것으로 예측되었다. 우르비나는 후에 이렇게 회상했다.

"컴퓨터 화면에 나타난 그 정보들을 보는 것은 놀랍고, 초현실적이었습니다. 공개된 자료와 기성 기술만으로 이런 종류의 데이터를 만들어내는 것이 얼마나 쉬운지 깨달았어요."

연구진은 t-SNE라는 기법을 사용해 이 분자들을 시각적으로 분석했다. 2차원 그래프 위에 펼쳐진 분자들의 분포도는 마치 악마의 지도 같았다. 새로 생성된 독성 분자들은 기존의 농약이나 환경 독소, 일반 약물들과는 완전히 다른 화학적 영역에 자리하고 있었다. VX와 유사한 분자들이 하나의 무시무시한 클러스터를 형성했고, 그 주변에는 더욱 독성이 강한 분자들이 별자리처럼 배치되어 있었다.

에킨스와 우르비나는 심각한 윤리적 딜레마에 직면했다. 이들의 선택은 단호했다. 더는 이 분자들을 자세히 분석하지 않기로 했고, 어떤 분자도 실제로 합성하지 않기로 결정했다. 컴퓨터 하드디스크에서 생성된 모든 분자 구조를 완전히 삭제했다. 그리고 과학계와 정책 입안자들에게 이 위험성을 반드시 알려야 한다고 다짐했다.

3. 전 세계를 흔든 경고음

2022년 3월 15일, 스위스 스피에즈에서 열린 컨버전스 회의장은 핀이 떨어지는 소리까지 들릴 정도로 조용했다. 에킨스와 우르비나가 자신들의 실험 결과를 발표하는 동안, 전 세계 화학무기·생물무기 전문가들과 무기 확산 방지 전문가들은 숨을 죽이고 있었다. 이들은 인공지능이 화학무기 개발에 악용될 수 있다는 구체적인 증거를 처음 목격하고 있었다.

발표가 끝나자 회의장은 폭발했다. 엄청난 토론이 벌어졌다. 런던 킹스 칼리지의 필리파 렌조스(Filippa Lentzos) 박사는 전쟁학과 보건 전문가로서 이중 사용(dual-use) 기술의 위험성을 연구해온 터였다. 그녀는 즉시 연구진과의 협력을 결정했다.

2022년 3월 7일, 세계적 권위의 학술지 《네이처 머신 인텔리전스》에 "인공지능 기반 약물 발견의 이중 사용"이라는 제목의 논문이 발표되었다. 파비오 우르비나가 제1저자로, 필리파 렌조스, 세드릭

인베르니지, 그리고 숀 에킨스가 교신저자로 이름을 올린 이 논문은 과학계에 지진과 같은 파장을 일으켰다. 가능성에 대한 추측이 아니라 구체적인 실험 결과를 제시했기 때문이었다.

미국 정부는 즉시 움직였다. 백악관 과학기술정책실에서 첫 번째 브리핑이 진행되었다. 대통령에게 과학기술 정책을 자문하는 이 기관은 국가안보회의와 함께 인공지능 기술의 국가 안보 영향에 대해 집중적으로 논의했다.

중앙정보국(CIA)은 잠재적 적성국가의 화학무기 개발 가능성에 대한 정보 수집 목적으로 브리핑을 요청했다. 국방부는 화학무기 방어 연구 담당 부서들과 미군의 화학무기 방어 능력 강화 방안을 검토했다. 국무부는 국제 무기 통제 협약 관련 논의와 화학무기금지협약 강화 방안을 모색했다.

Defense Threat Reduction Agency라는 대량살상무기 확산 방지 전담 기관에서는 새로운 형태의 화학무기 위협에 대한 대응책을 논의했다. 미 육군 화학무기 연구소에서는 인공지능을 이용한 화학무기 탐지 및 방어 기술 개발을 검토했다.

국제기구들도 발 빠르게 대응했다. 화학무기금지기구(OPCW)는 화학무기금지협약의 이행을 감시하는 국제기구로서, 인공지능을 이용한 새로운 화학무기 개발 가능성과 기존 통제 목록에 없는 새로운 독성 물질의 출현 가능성을 심각하게 받아들였다. 40개국이 참여하는 오스트레일리아 그룹은 이중 사용 품목의 수출 통제를 담당하는 국제 협의체로서, 인공지능 소프트웨어의 수출 통제 필요성과 화학 데이터베이스 접근 제한 방안, 연구자 신원 확인 시스템 도입 같은 새로운 통제 방안을 진지하게 검토하기 시작했다.

로스알라모스, 샌디아, 리버모어 등 미국의 주요 국립연구소들도 차례로 브리핑을 요청했다. 국가 보안과 관련된 첨단 연구를 수행하는 이들에게 이 문제는 매우 절실한 현실이었다.

2022년, 넷플릭스 제작진이 노스캐롤라이나 주립대학교 캠퍼스를 찾아와 에킨스를 인터뷰했다. 군사 분야에서의 인공지능 사용을 다룬 다큐멘터리의 일부였다. 《사이언티픽 아메리칸》, 《네이처》, 《케미스트리 월드》 등 주요 과학 저널들이 이 연구를 집중 보도하면서 일반 대중도 이 문제의 심각성을 인식하게 되었다.

4. 안전장치의 한계

콜라보레이션 파마슈티컬의 실험이 전 세계를 충격에 빠뜨린 이유는 사용된 기술과 자원이 너무나 평범했다는 점이었다. 이것은 특별한 슈퍼컴퓨터나 비밀 실험실에서 벌어진 일이 아니었다. 현재 가격으로 약 200만원 수준의 2015년형 맥북 하나면 충분했다. 처리 시간은 단 6시간, 일반 가정용 전력으로도 충분했다.

소프트웨어 역시 누구나 접근할 수 있었다. 텐서플로, 파이토치 같은 무료 인공지능 라이브러리들이 인터넷에 널려 있었다. 깃허브 같은 플랫폼에서는 관련 코드와 튜토리얼을 쉽게 찾을 수 있었다. 데이터 측면에서도 PubChem, ChEMBL 같은 공개 화학 데이터베이스를 누구나 활용할 수 있었고, 환경청이나 식약처가 공개한 독성 데이터도 연구 목적으로 자유롭게 사용할 수 있었다.

더욱 무서운 사실은 메가신과 유사한 인공지능 도구를 사용하는 회사가 전 세계에 400여 개나 된다는 것이었다. 화이저, 로슈, 노바티스 같은 거대 제약회사들이 AI 도구를 적극 활용하고 있었고, 전 세계 수천 개의 바이오테크 스타트업이 AI 기반 약물 개발을 진행하고 있었다. 학술 기관들도 유사한

도구들을 연구용으로 사용하고 있었다.

에킨스는 씁쓸하게 말했다. "우리 분야 동료들 중 얼마나 많은 사람이 자신의 기술이 악용될 가능성을 생각해봤을까요? 대부분은 그런 생각을 해본 적이 없을 겁니다."

실제로 많은 연구자들이 치료제 개발만 생각하고 악용 가능성은 간과하고 있었다. 대학이나 회사에서 이중 사용에 대한 교육은 거의 없는 상황이었다. 기술적 장벽도 놀라울 정도로 낮았다. 컴퓨터공학과 학부 수준의 실력이면 충분했고, 기본적인 유기화학 지식만 있으면 시작할 수 있었다. 독성학은 온라인 강의나 교과서로 학습할 수 있었고, 특별한 실험실도 필요 없었다.

정보 접근 장벽은 사실상 존재하지 않았다. 깃허브에서 유사한 코드를 쉽게 확보할 수 있었고, 콜라보레이션 파마슈티컬의 논문에서 방법론이 공개되어 있었다. 머신러닝과 화학 정보학 강의가 무료로 제공되었고, 온라인 포럼에서 기술적 조언을 구하기도 쉬웠다.

법적 장벽은 더욱 불완전했다. 컴퓨터 시뮬레이션은 화학무기금지협약의 적용 대상이 아니었다. 대부분 국가에서 가상 분자 설계에 대한 규제는 없었다. 연구자의 양심에만 의존하는 현실이었고, 인터넷을 통한 익명 활동도 가능했다.

화학무기와 생물무기 제조의 현실은 더욱 암울했다. SciFinder, Reaxys 같은 화학 합성 경로를 자동으로 예측하는 소프트웨어가 공개되어 있었다. IBM의 RXN, MIT의 ASKCOS 같은 합성 생성 AI가 무료 또는 저렴하게 제공되고 있었다. 가장 효율적인 합성 방법을 자동으로 찾아주는 최적화 알고리즘도 있었다.

전 세계에는 의뢰받은 화학 물질을 합성해주는 회사들이 수백 개 있었다. 중국 업체들은 상대적으로 규제가 느슨하고 가격이 저렴했다. 인도 업체들은 제네릭 의약품 제조 경험을 바탕으로 다양한 화합물을 합성했다. 서구 업체들은 품질은 높지만 규제가 상대적으로 엄격했다. 문제는 이들 업체 대부분이 의뢰받은 물질의 최종 용도를 확인하지 않는다는 것이었다.

인공지능은 화학무기뿐만 아니라 생물무기 설계에도 악용될 수 있었다. 최근 단백질 구조 예측과 설계 분야에서 혁명적 발전이 있었다. 구글 딥마인드의 알파폴드는 단백질 구조를 예측하는 AI였다. 워싱턴 대학의 로제타폴드는 오픈소스 단백질 설계 도구였다. ProtGPT는 기존의 GPT 아키텍처를 기반으로 하되, 자연어 대신 단백질 서열 데이터로 훈련된 모델이었다.

이 도구들을 악용하면 신경계를 마비시키는 신경독 펩타이드, 세포를 파괴하는 세포독 단백질, 면역체계를 무력화하는 면역억제제, 기존 치료법을 무효화하는 항생제 내성 변형 병원균 등을 설계할 수 있다.

자동 합성(autonomous synthesis) 기술이 발전하면서 인간 개입 없이 분자 설계부터 합성까지 자동으로 진행되는 설계-제조-테스트 자동화가 가능해지고 있었다. 24시간 무인으로 수천 개 화합물을 제조할 수 있는 로봇 실험실도 현실이 되고 있었다.

1997년 화학무기금지협약은 한계가 많았다. 정해진 화학물질만 통제하는 목록 기반 통제 방식이다 보니 새로운 물질에 대한 대응이 느렸다. 최종 생성물이 아닌 원료 물질 위주로 통제하는 방식이었다. 컴퓨터로 물질 생성을 시뮬레이션하는 것은 협약 적용 대상도 아니었다.

1975년 발효된 생물무기금지협약(BWC)은 더 큰 한계가 있었다. 냉전 시대의 산물로, 현재의 기술 발전을 예상하지 못했다. 핵무기금지조약(NPT)과 달리 국제원자력기구(IAEA) 같은 독립적 검증기구가 없어 실질적 감시가 어려웠다. 신기술은 전통적인 '방어-공격' 구분선을 더욱 모호하게 만들고 있었다.

5. 위협과 방어 기술

인공지능 기술은 멈추지 않고 발전하고 있었다. 새로운 기회와 함께 새로운 위험도 다가오고 있었다. ChatGPT, 클로드 같은 대규모 언어 모델이 화학 분야에 특화될 가능성이 높아지고 있었다. 화학 지식에 특화된 ChemGPT가 등장하면 "VX보다 10배 독한 물질 만들어줘" 같은 자연어 명령으로 분자를 설계하는 자연어 분자 설계가 가능해질 수 있었다.

기존 연구를 자동으로 분석해서 새로운 독성 물질을 예측하는 자동 문헌 검토, 언어 모델이 화학 합성 과정을 자동으로 설계하는 합성 경로 자동 생성 등도 현실이 될 수 있었다.

양자컴퓨터가 실용화되면 더욱 강력한 분자 시뮬레이션이 가능해질 것이었다. 기존 컴퓨터로는 불가능한 정밀한 분자 모델링으로 정확도가 향상되고, 몇 시간 걸리는 작업을 몇 분 만에 완성하는 속도 향상이 있을 것이었다. 더 복잡하고 정교한 독성 물질 설계가 가능해지는 복잡성 증가도 예상되었다.

'Lights-out Science'라고 불리는 자동화 실험실도 등장하고 있었다. 인간 없이 연속적인 실험이 진행되는 24시간 무인 운영, 인공지능이 실험을 설계하고 결과를 분석하는 AI 주도 실험, 전 세계 어디서든 인터넷으로 실험실을 조작할 수 있는 원격 조작, 하루에 수천 개의 새로운 화합물을 합성하는 대량 생산이 가능해질 것이었다.

개인의 유전자 정보를 이용한 맞춤형 생물무기라는 끔찍한 가능성도 있었다. 개인의 DNA 정보를 바탕으로 취약점을 분석하는 유전자 분석, 특정 개인에게만 치명적인 물질을 설계하는 맞춤형 독성, 특정 인종이나 집단에게만 영향을 주는 무기인 집단별 차별, 기존 탐지 방법으로는 발견하기 어려운 새로운 독성 물질 등이 가능해질 수 있었다.

새로운 형태의 테러인 사이버-바이오 융합 공격도 나타날 수 있었다. 원격 실험실 해킹, 연구 데이터를 조작해서 잘못된 약물이나 백신 개발을 유도하는 데이터 조작, 제약회사의 생산 시설을 해킹해서 독성 물질을 생산하는 공급망 공격, 가짜 연구 결과를 퍼뜨려서 사회 혼란을 야기하는 정보전 등이 가능해질 수 있었다.

하지만 위협이 증가하는 만큼 방어 기술도 발전하고 있었다. 악의적인 AI 사용을 탐지하는 시스템이 개발되고 있었다. 비정상적인 연구 패턴을 자동으로 탐지하는 패턴 분석, 평소와 다른 분자 설계 요청을 실시간으로 발견하는 이상 탐지, 연구자의 활동을 분석해서 위험 신호를 포착하는 행동 분석, 악의적 연구자들 간의 협력 네트워크를 추적하는 네트워크 분석이 가능해지고 있었다.

다양한 생물무기에 효과적인 범용 백신, 새로운 위협 발견 시 몇 시간 만에 치료제를 설계하는 신속 대응, 환자의 유전자에 맞춘 개인화된 치료법인 개인 맞춤 치료 등도 연구되고 있었다.

6. Dual Use와 인간의 책임

모든 도구는 창조와 파괴, 두 얼굴을 가지고 있다. 불은 인류에게 문명을 선사했지만 화재로 생명을 빼앗기도 한다. 원자력은 깨끗한 에너지를 제공하지만 핵무기로 도시를 잿더미로 만들 수도 있다. 인

터넷은 지식과 정보를 민주화했지만 사이버 범죄와 가짜뉴스의 온상이 되기도 한다. 인공지능도 마찬가지다.

콜라보레이션 파마슈티컬의 메가신이 보여준 것은 인공지능의 이중성이었다. 같은 기술이 생명을 구하는 치료제를 만들 수도 있고, 생명을 앗아가는 독극물을 만들 수도 있다는 사실이었다. 문제는 기술 자체가 아니라 그것을 사용하는 인간의 의도였다.

모두가 이 문제 해결에 중요한 역할을 할 수 있다. 먼저 이 문제의 심각성을 정확히 이해해야 한다. 언론이 이 문제를 지속적으로 다루도록 관심을 표현해야 한다. 단순히 기술 발전의 밝은 면만 보는 것이 아니라 어두운 면도 함께 봐야 한다.

정치적 참여도 중요하다. 정치인들에게 AI 안전 정책 마련을 요구해야 한다. AI 안전을 중요하게 생각하는 후보를 지지하는 선거 참여를 해야 한다. AI 안전을 위한 시민 단체 활동에 참여해야 한다. 전 세계 시민들과의 연대 활동도 필요하다.

소비자로서의 선택도 중요하다. 윤리적 AI 개발에 앞장서는 기업 제품을 선택해야 한다. 투자 시 기업의 AI 윤리 정책을 고려하는 투자 결정을 해야 한다. AI 서비스 사용 시 안전성과 윤리성을 확인해야 한다. 기업에 AI 안전성에 대한 의견을 적극 전달하는 피드백을 제공해야 한다.

교육도 근본적으로 중요하다. 초중고 교육에서는 기본적인 화학과 생물학 지식 교육을 강화해야 하고, 인공지능의 원리와 한계에 대한 이해 교육을 해야 하며, 과학 기술의 윤리적 사용에 대한 교육을 해야 하고, 새로운 기술의 장단점을 냉정히 판단하는 능력을 배양하는 비판적 사고 교육을 해야 한다.

과학자와 엔지니어들은 더 큰 책임을 져야 한다. 연구 방법과 결과를 투명하게 공개하는 투명성을 유지해야 한다. 일반 대중과의 적극적인 소통과 설명인 사회적 소통을 해야 한다. 자신의 연구가 악용될 가능성에 대해 항상 고민해야 한다. 동료 연구자들과 윤리적 기준을 공유하고 서로 견제해야 한다.

정부와 국제기구는 법적, 제도적 틀을 마련해야 한다. 기술 발전 속도에 맞는 유연하고 효과적인 규제를 만들어야 한다. 국경을 넘나드는 기술의 특성상 국제 협력이 필수다. 하지만 과도한 규제로 유익한 연구를 막아서는 안 된다. 균형이 중요하다.

기업들은 단기적 이익보다 장기적 안전을 우선해야 한다. 자율 규제를 통해 업계 표준을 높여야 한다. 연구자들에게 윤리 교육을 제공하고, 내부 감시 시스템을 구축해야 한다. 무엇보다 투명성을 바탕으로 대중의 신뢰를 얻어야 한다.

시민사회도 중요한 역할을 할 수 있다. AI 안전을 위한 시민 단체들이 정부와 기업을 견제하고 감시해야 한다. 전문가들과 일반 시민들 사이의 소통을 도와야 한다. 복잡한 기술 문제를 이해하기 쉽게 설명하고 교육하는 역할도 중요하다.

언론의 책임도 크다. 기술 발전의 밝은 면만 부각시키는 것이 아니라 어두운 면도 균형 있게 다뤄야 한다. 복잡한 과학 기술 문제를 정확하고 이해하기 쉽게 보도해야 한다. 공포를 조장하지도, 위험을 축소하지도 말고 사실을 바탕으로 보도해야 한다.

개인으로서 우리가 할 수 있는 일도 많다. 과학 기술에 대한 기본적인 이해를 높여야 한다. 새로운 기술이 나올 때마다 그것의 장점과 위험을 모두 고려하는 습관을 길러야 한다. 전문가의 의견을 듣되, 맹목적으로 따르지는 말고 비판적으로 생각해야 한다.

콜라보레이션 파마슈티컬의 실험은 우리에게 중요한 경고를 주었다. 인공지능이라는 강력한 도구는 인류를 구원할 수도 있고 파멸시킬 수도 있다. 단 6시간 만에 40,000개의 독성 분자를 만들어낸 사실은 이 기술의 놀라운 가능성과 동시에 끔찍한 위험성을 보여준다.

숀 에킨스와 파비오 우르비나가 보여준 용기 있는 실험과 투명한 공개는 과학자의 사회적 책임이 무엇인지를 잘 보여준다. 그들은 자신들의 발견이 불편하고 무서운 것임을 알면서도 숨기지 않고 전 세계에 알렸다. 이것이야말로 진정한 과학자의 자세였다.

인공지능은 이미 우리 삶의 일부가 되었고, 앞으로 더욱 중요해질 것이다. 이 강력한 기술이 인류의 번영을 위해 사용될지, 아니면 파괴의 도구가 될지는 지금 우리의 선택에 달려 있다. 콜라보레이션 파마슈티컬의 메가신이 만들어낸 40,000개의 독성 분자는 삭제되었지만, 그 가능성은 여전히 남아 있다.

우리는 인공지능이라는 판도라의 상자를 이미 열었다. 이제 그 안에서 나온 모든 것들을 지혜롭게 다루는 법을 배워야 한다. 기술 자체를 두려워할 필요는 없다. 하지만 그것을 악용하려는 인간의 욕망은 경계해야 한다.

답은 간단하지 않다. 하지만 한 가지는 분명하다. 우리가 함께 노력하고 지혜롭게 대처한다면, 인공지능이 가져다줄 엄청난 혜택을 누리면서도 그 위험을 최소화할 수 있을 것이다. 이것이 바로 과학기술 시대를 살아가는 우리 모두의 과제이자 책임이다.

인공지능의 미래는 아직 쓰여지지 않았다. 그 미래를 써 내려가는 것은 바로 우리 자신이다. 현명한 선택을 통해 인공지능이 인류의 친구가 되도록 만들어나가야 할 것이다. 칼 자체는 선하지도 악하지도 않다. 그것을 쥔 손이 모든 것을 결정한다. 인공지능도 마찬가지다. 우리가 어떻게 사용하느냐에 따라 축복이 될 수도, 저주가 될 수도 있다.

전체 목차

책의 모든 장 보기

김경진 변호사

2장 전 세계를 감시중인 인공지능

전체 목차

책의 모든 장 보기

"빅 브라더가 당신을 지켜보고 있다" - 조지 오웰

1. 24시간 지켜보는 AI의 눈

당신이 지금 이 글을 읽고 있는 순간에도, 보이지 않는 수천 개의 눈이 당신을 지켜보고 있다. 스마트폰 속 카메라, 거리의 CCTV, 지하철역의 안면인식 시스템, 심지어 당신이 좋아요를 누른 소셜미디어 게시물까지. 모든 것이 데이터가 되어 어디선가 분석되고 저장되고 있다.

인공지능은 원래 인류의 삶을 더 풍요롭고 편리하게 만들기 위해 태어났다. 의료진단을 도와주고, 교통체증을 해결하고, 언어의 장벽을 허물어주는 마법 같은 기술로 말이다. 하지만 현실은 어떤가? 그 똑똑한 기술이 사람들을 감시하고 통제하며, 심지어 생명을 위협하는 무기로 변모하고 있다. 기술 자체가 악한 것이 아니다. 문제는 그 기술을 쥔 손이 무엇을 하려 하느냐이다.

이스라엘의 라벤더 시스템이 팔레스타인 사람들의 생사를 숫자로 계산할 때, 중국의 천라지망이 14억 인구의 일거수일투족을 추적할 때, 미국의 팔란티어가 전 세계 정보를 실시간으로 분석할 때, 우리는 깨달아야 한다. 인공지능이 더 이상 우리의 편리한 도구가 아니라, 우리를 옥죄는 족쇄가 되고 있다는 것을.

전 세계로 확산되고 있는 AI 감시 시스템들에는 섬뜩한 공통점이 있다. 첫째, 당신의 동의 따위는 묻지 않는다. 당신이 원하든 원하지 않든 개인정보를 대량으로 수집해버린다. 둘째, 그 알고리즘이 어떻게 작동하는지 아무도 모른다. 마치 마법의 검은 상자처럼 불투명하기 그지없다. 셋째, 완벽하지 않으면서도 중요한 결정을 내린다. 10%의 오차율이라니, 그 10% 안에 당신이 들어갈 수도 있다는 뜻이다. 넷째, 특정 집단을 차별적으로 겨냥한다. 피부색이 다르거나, 믿는 종교가 다르거나, 정치적 성향이 다르다는 이유만으로.

이것은 단순한 기술의 문제가 아니다. 인간의 존엄성, 프라이버시, 표현의 자유, 종교의 자유 등 민주주의 사회가 지켜온 모든 가치를 뿌리째 흔드는 인권의 문제다. 한국의 국가인권위원회조차 2024년 5월 '인공지능 인권영향평가 도구'를 활용하라고 권고했을 정도로, 우리나라도 이 문제에서 자유롭지 않다.

지금 이 순간, 인공지능 기술이 인간을 해방시키는 도구가 될지, 아니면 인간을 억압하는 족쇄가 될지가 결정되고 있다. 그 결정권자는 정부나 기업이 아니다. 바로 당신이고, 나이며, 우리 모두다.

2. 이스라엘의 '라벤더' 시스템

2024년 4월, 세상을 충격에 빠뜨린 폭로가 있었다. 이스라엘-팔레스타인의 독립 매체 '+972 매거진'과 히브리어 매체 '로컬콜'이 공개한 '라벤더(Lavender)' AI 시스템의 실체였다. 그 내용은 우리가 AI에 대해 품었던 모든 악몽을 현실로 만든 것이었다.

이스라엘군 정보부대가 개발한 이 AI는 마치 신의 눈처럼 가자지구를 내려다보며, 230만 명의 운명을 1점부터 100점까지의 숫자로 매겼다. 하마스 무장대원일 가능성을 점수로 계산한다는 것이다. 당신이 누구와 통화하는지, 어떤 앱을 다운로드하는지, 어디를 자주 방문하는지 모든 것이 알고리즘의 먹이감이 되었다.

라벤더는 이미 알려진 하마스 요원들의 데이터를 먼저 학습했다. 그들이 사용하는 휴대폰의 브랜드, 자주 다운로드하는 앱, 통화 패턴, 이동 경로까지. 그리고 이와 비슷한 패턴을 보이는 사람들을 차례로 찾아내며 점수를 매겼다. 마치 독감 바이러스가 전파되듯, 연결고리를 따라 의심의 그물망을 넓혀갔다.

생명을 숫자로 계산하는 알고리즘

이스라엘군은 라벤더가 90%의 정확도를 보인다며 만족해했다. 90%라니, 얼핏 들으면 상당히 높은 수치다. 하지만 잠깐, 이것이 무엇을 의미하는지 생각해보자. 10%는 틀릴 수 있다는 뜻이다. 라벤더가 37,000명을 공격 목표로 지정했다면, 그 중 3,700명은 무고한 사람들일 수 있다는 얘기다. 3,700명의 목숨을 오차 범위라고 부를 수 있을까?

더욱 끔찍한 것은 전쟁이 시작된 후 인간의 검증 절차마저 느슨해졌다는 점이다. 전쟁 이전에는 "인간이 직접 얻은 최소 두 개의 정보가 라벤더 AI의 판단을 뒷받침해야 한다"는 내부 규정이 있었다. 하지만 가자 전쟁이 시작되자 이 기준이 "한 개"로 낮아졌고, 실전에서는 이마저도 무시되는 경우가 허다했다.

한 정보분석관의 증언은 소름끼친다. "우리는 라벤더 인공지능의 결정에 확인 도장만 찍었다. 목표물 공격을 승인하는 데 단 20초밖에 걸리지 않았다." 생사를 가르는 중요한 결정을 실질적으로 인공지능이 내리고 있었던 것이다. 인간의 판단은 그저 형식적인 절차에 불과했다.

한 장교의 냉소적인 말이 모든 것을 요약한다. "돈·시간·인력을 아끼는 것이 민간인 목숨보다 훨씬 중요했다."

"아빠는 어디에?" - 가족을 인질로 잡는 시스템

라벤더와 함께 운영되는 또 다른 시스템은 그 이름부터 소름끼친다. "아빠는 어디에?(Where's Daddy?)"라고 불리는 이 추적 시스템은 목표 인물이 집에 돌아오는 순간을 감지해 이스라엘군에 알려준다.

왜 하필 "아빠는 어디에?"일까? 그 이유는 단순하면서도 잔혹하다. 아버지가 집에 돌아오는 시점을 노려 공격하기 때문이다. 이 시스템은 목표 인물의 일상 패턴을 학습하고, 집 위치를 파악하며, 언제 집에 들어가는지 실시간으로 감시한다.

이스라엘 장교의 증언은 충격적이다. "우리는 하마스 요원들이 군사 건물에 있거나 군사 활동을 할 때만 공격한 것이 아니다. 집에 있을 때 그들을 공격하는 것이 훨씬 쉬웠고, 시스템은 하마스 요원들이 집에 있는 것을 파악하기 위해 만들어진 것이다."

결과는 예측 가능했다. 한 명의 목표물을 제거하기 위해 그의 가족 전체가 함께 죽어갔다. 집에서 폭격을 당하면 여성과 어린이들인 가족들이 함께 희생될 수밖에 없었다. 국제법이 명시하고 있는 민간인 보호 원칙을 정면으로 무시한 행위였다.

부수적 피해라는 이름의 대량 학살 허가증

이스라엘군 정보요원 6명이 증언한 내용은 더욱 충격적이다. 전쟁 초기 몇 주 동안, 이스라엘군은 라벤더가 지목한 각 무장대원을 암살할 때 15명에서 20명의 민간인이 희생되어도 '용인 가능한 범위'라고 인정했다. 하마스 사령관급을 공격할 때는 100명 이상의 민간인 사망도 허용했다고 한다.

미국 국무부의 국제법 전문가는 가디언과의 인터뷰에서 "1대 15라는 비율은, 특히 하급 전투원에 대해서는 전혀 들어보지 못했다"고 말했다. 이스라엘이 적용한 민간인 피해 허용 기준이 국제적으로 전례가 없을 정도로 관대했다는 뜻이다.

IDF 내부 증언자들의 말이 이 모든 상황을 요약한다. "하위급 무장 세력에게 이스라엘군의 인력과 시간을 과도하게 쓰고 싶지 않았다. 부수적인 피해와 민간인 살상, 그리고 실수로 무고한 사람을 공격하는 오류를 감수하고라도 AI를 쓸 만했다."

인간의 존엄성을 부정하는 기계의 판결

라벤더 시스템의 가장 큰 문제는 인간의 생명을 단순한 데이터로 취급한다는 점이다. 사람의 삶과 감정, 가족과의 관계, 개인의 역사는 모두 무시되고, 오직 인공지능 알고리즘이 계산한 확률에 의해 생사가 결정된다. 이는 인간 존엄성에 대한 근본적인 부정이며, 기술이 인간을 위해 존재해야 한다는 원칙을 완전히 뒤집는 것이다.

투명성도 전혀 없다. 알고리즘이 어떤 기준으로 누군가를 위험한 인물로 분류하는지 정확히 알 수 없으며, 잘못 분류된 사람들이 자신의 무고함을 증명할 방법도 없다. 적법한 절차나 법적 보호 없이 사람들의 생명을 빼앗는 것이다.

가자지구의 안면인식 지옥

라벤더와 함께 이스라엘이 가자지구에 도입한 안면인식 기술은 또 다른 악몽이었다. 2023년 가자지구 전쟁 발발 이후 하마스 관련자를 색출하고 이스라엘 인질을 식별한다는 명분으로 도입된 이 시스템은, 실제로는 팔레스타인 주민들을 무차별적으로 감시하는 도구가 되었다.

이스라엘 정보부대 8200이 주도하는 이 작전에서는 민간기업 '코사이트(Corsight)'의 안면인식 기술이 사용되었다. 검문소, 드론, 군용 카메라를 통해 운영되며, 코사이트 직원들이 현장에 직접 투입되어 기술 지원을 제공했다.

가자지구에 배치된 이스라엘 군인들에게는 이 기술이 탑재된 카메라가 제공되어, 피란민들이 검문소를 통과할 때마다 얼굴을 무작위로 촬영하고 스캔한 후 기존 데이터베이스와 대조하여 신원을 특정했다.

코사이트는 자사의 안면인식 기술이 "얼굴의 50% 미만 노출 시나 어둠 속·저화질 환경에서도 정확한 인식이 가능"하다고 자랑했다. 하지만 현실은 달랐다. 팔레스타인 시인 모사브 아부 토하가 그 산 증인이다. 그는 세 살 아이와 함께 피란길에 올랐다가 이스라엘군 검문소에서 안면인식 카메라에 스캔되었고, AI의 오류로 수배자 명단에 오른 사람으로 잘못 파악되면서 이틀간 구금되어 구타와 고문을 당했다.

국제앰네스티는 이러한 이스라엘의 행위를 "자동화된 아파르트헤이트"라고 비판했다. AI 기술이 전쟁과 감시에 악용될 때 발생할 수 있는 심각한 인권 침해의 대표적인 사례가 되었다.

3. 걸음걸이만으로도 당신을 찾아내는 중국

중국의 천라지망(天網) 시스템 앞에서는 숨을 곳이 없다. 이 이름은 노자의 『도덕경』에서 따온 것으로, "하늘 그물의 눈은 매우 넓어서 성긴 것 같지만 악인은 하나도 놓치지 않는다"는 뜻이다. 하지만 여기서 말하는 '악인'의 기준은 누가 정하는가? 바로 중국 공산당이다.

14억 인구를 1초 만에 스캔

중국은 전국에 10억 대가 넘는 CCTV를 설치했다. 상상해보라. 10억 개의 눈이 24시간 내내 당신을 지켜보고 있다는 것을. 중국 정부는 이 시스템을 통해 전체 인구를 1초 만에 조사할 수 있다고 자랑한다. 14억 명의 얼굴을 1초 만에 구분해낸다니, 기술적으로는 경이롭지만 인권적으로는 끔찍한 일이다.

중국의 안면인식 기술은 세계 최고 수준이다. 2018년 미국표준연구소(NIST)가 주최한 안면인식 공급자 테스트에서 1위부터 4위까지 모두 중국 기업이 개발한 알고리즘이 차지했다. 그들이 이 분야에서 세계를 선도하는 이유는 간단하다. 실험대상이 14억이나 되기 때문이다.

99. 8%의 정확도

천라지망의 안면인식 정확도는 99.8%에 달한다. 운영 시작 후 2년 동안 이 시스템을 통해 체포된 범죄자만 2,000명을 넘었다. 하지만 여기서 주목해야 할 점은 사람의 얼굴뿐만 아니라 걸음걸이까지도 인식해서 개인을 식별한다는 것이다. 당신이 모자를 쓰거나 마스크를 착용해도 소용없다. 걸음걸이 하나만으로도 당신을 찾아낸다.

실제 사례들을 보면 그 위력이 어느 정도인지 알 수 있다. 17년 전 남자친구를 살해한 여성을 고속도로 요금소에 설치된 안면인식 시스템이 찾아냈다. 5만여 명이 모인 콘서트장에서 경제 범죄로 수배 중이던 남성이 카메라 얼굴인식 기술로 발견되어 체포되었다. 2000년 저장성에서 발생했던 시신 훼손 사건의 범인을 16년 만에 찾아낸 사례도 있다.

가장 극적인 사례는 안후이성의 한 사찰에서 일어났다. 주지스님의 얼굴이 강력범죄 용의자와 일치한다는 사실을 시스템이 발견한 것이다. 이름과 성을 바꾸고 10년 넘게 스님으로 살던 범인은 결국 체포되었다.

이런 사례들을 보면 중국의 감시 기술이 얼마나 정교하고 광범위한지 알 수 있다. 하지만 동시에 이런 기술이 일반 시민들의 사생활을 얼마나 심각하게 침해할 수 있는지도 보여준다.

매의 눈으로 완성된 감시망

2016년 중국은 공산당 중앙위원회의 승인을 받아 '매의 눈(Sharp Eyes)' 프로젝트를 시작했다. 중국 전역을 24시간 감시하고 통제할 수 있는 CCTV 감시 네트워크를 완성하겠다는 야심찬 계획이었다.

매의 눈 프로젝트는 단순히 더 많은 카메라를 설치하는 것에 그치지 않는다. 도로나 다중이용시설에 설치된 CCTV를 주민들의 TV, 휴대전화, 도어락 등과 연결해 실시간으로 상황을 파악할 수 있는 네트워크를 구축하는 것이다. 말 그대로 모든 국민이 서로를 감시하는 사회를 만드는 것과 같다.

상상해보라. 당신의 TV가, 당신의 휴대폰이, 심지어 당신 집 현관문까지도 정부의 감시 도구가 되는 세상을. 중국 전 국토가 거대한 감시망으로 덮여있다는 것은 바로 이런 의미다.

일상 감시

중국의 안면인식 기술은 이제 일상생활의 모든 영역에 스며들었다. 교통법규를 위반하면 위반자의 얼굴을 촬영하여 전광판에 공개한다. 횡단보도를 무단횡단하면 자동으로 단속된다. 지하철 요금을 결제할 때는 스마트폰 없이 손바닥이나 얼굴만으로도 가능하다.

심지어 베이징 톈탄 공원에서는 화장실 휴지 도둑질을 막기 위해 안면인식 기계를 도입했다. 얼굴을 인식해야만 일정량의 휴지를 제공받을 수 있다. 화장실에서까지 당신의 얼굴이 기록되는 것이다.

2019년 12월 1일부터는 휴대전화 신규 번호 등록 시 안면 스캔이 의무화되었다. 정부는 "사이버 공간에서 시민 권익 보호"를 명목으로 내세웠지만, 실제로는 정부 주도의 생체 정보 데이터베이스 구축이 목적이었다. 수집된 얼굴 정보는 전국 7억 대 이상의 CCTV와 연동된 감시 네트워크와 통합되어 실시간 신원 식별과 이동 경로 추적이 가능해졌다.

코로나19 사태는 감시 기술 확산의 절호의 기회가 되었다. 학교, 관공서, 아파트 단지 등에서 체온 측정과 함께 안면인식을 통한 출입 통제가 일상화되었다. 마스크를 착용한 상태에서도 식별이 가능한 기술이 개발되어 사용되고 있다.

신장 위구르

중국의 감시 시스템이 가장 극단적으로 운영되는 곳은 신장 위구르 자치구다. 이곳은 말 그대로 감시 기술의 실험실이 되었다. 이슬람 사원, 위구르 지역사회, 공항, 기차역, 버스 정류장 등 신장 전역의 670만 곳에 CCTV와 안면 스캔 시스템이 설치되었다.

휴먼라이츠워치에 따르면 신장 지구 인구의 10분의 1인 250만 명이 매일 중국 정보당국의 감시를 받고 있다. 12세부터 65세까지 주민의 DNA, 지문, 홍채 스캔, 혈액형 등 생체 정보가 강제로 수집되고 있다. 모든 사람을 잠재적인 범죄자로 취급하는 것과 다름없다.

위구르족을 감시하기 위해 사용되는 통합공동운영플랫폼(IJOP)은 그 자체가 인종차별의 도구다. 이 플랫폼은 신장에서 위구르인들을 추적하기 위해 DNA를 포함한 생체 데이터를 수집한다. 감시를 넘어서 특정 소수 민족 전체에 대한 체계적인 억압과 통제를 의미한다.

휴먼라이츠워치가 입수한 아커쑤 현의 위구르인 수감자 2,000명의 명단을 보면 경악을 금할 수 없다. 수감 사유가 허가 없는 코란 공부, 긴 수염, VPN 사용, 휴대전화의 반복적인 종료, 종교 행위 등이었다. 이러한 일상적인 행위들이 감시의 대상이 되어 사람들을 수용소에 가두는 근거가 되고 있다.

화웨이는 위구르족만을 타겟으로 한 감시 카메라 시스템을 개발했다고 알려져 논란이 되고 있다. 시스템이 위구르 소수민족임을 감지하면 자동으로 '위구르 알람'을 발동해 경찰에 통보한다는 것이다. 알리바바도 위구르족 식별이 가능한 안면인식 프로그램을 보유한 것으로 알려져 있다.

종교인을 대상 노크 작전

중국은 종교 신자들이 무신론과 마르크스주의에 근거한 공산주의 이데올로기를 따르도록 강요하고 있다. 공산당은 종교단체를 당의 권위에 맞설 가능성이 있는 잠재적 도전 세력으로 간주한다.

1860년 홍수전이 기독교 사상을 내세운 태평천국의 난으로 청 왕조가 무너질 뻔했던 기억이 여전히 생생하다. 시진핑 집권 이후 '종교의 중국화' 정책을 내세우며 통제를 강화하고, 종교조직이 당과 정부의 요구에 따를 것을 요구하고 있다.

중국 형법 300조에 따라 금지된 종교 단체에서 활동하는 것은 3년에서 7년의 징역형 대상인 범죄로 간주된다. 정부는 금지된 종교 단체 신자를 신고하는 자에게 10만 위안의 보상금을 제공하며 전 국민 고발 체제를 구축했다.

2017년 초부터 중국 전역에서 '노크 작전'이라는 광범위한 감시 활동이 시행되었다. 적당한 구실을 만들어 경찰을 보내 종교인들을 조사하고 사진을 찍어 관리하는 것으로, 전국적으로 종교인들을 추적하는 감시 시스템의 일환이다.

수집된 데이터는 국가안전보위부 전용 컴퓨터에 저장되며, 조사관들은 종교를 전파하는 자들에 대한 증거를 찾아 나선다. 증거가 확보되면 추가 조사가 진행되고, 조사 후에는 '매의 눈'과 '천라지망' 프로젝트를 통해 대상자를 끊임없이 감시한다.

사회신용시스템: 점수로 매겨지는 인간의 가치

중국 정부가 구축한 사회신용시스템은 '신뢰 보상, 불신 처벌' 메커니즘으로 작동한다. 중국에 사는 모든 사람의 행동을 관찰하고 점수를 매기는 시스템이다. 마치 학교에서 학생들의 행동을 평가하는 것처럼, 정부가 국민 한 사람 한 사람의 행동을 살펴보고 점수를 주는 것이다.

이는 지방자치단체나 철도청, 금융기관 등 각각의 기관별로 시스템을 구축하는 것이지 중앙정부가 법률로서 체계적으로 구축한 시스템은 아니다. 하지만 다양한 지방자치단체와 기관이 참여하고 있기 때문에 실질적으로는 중국 전역에서 이 제도가 일사불란하게 시행되는 것과 다름없다.

신용점수가 높은 사람들은 호텔에 투숙할 때 보증금을 지불할 필요가 없고, 다른 사람들보다 더 빨리 비자를 발급받을 수 있다. 반대로 신용점수가 낮은 사람들은 호텔 투숙, 기차 탑승, 여행의 자유 등에 다양한 제한과 제재를 받는다.

사회신용시스템은 여러 가지 방법으로 사람들의 행동을 관찰한다. CCTV 카메라들이 사람들의 행동을 24시간 지켜보고 있으며, 얼굴인식 기술을 사용해서 누가 어디서 무엇을 했는지 파악할 수 있다. 인터넷 사용 기록, 쇼핑 내역, 대출 상환 기록, 교통 위반 기록 등 일상생활의 모든 정보를 수집한다.

좋은 점수를 받으려면 세금을 제때 내고, 교통 규칙을 잘 지키고, 빌린 돈을 제때 갚아야 한다. 봉사활동을 하거나 부모님을 잘 모시는 것도 좋은 점수를 받는 방법이다. 반대로 교통 신호를 위반하거나, 빌린 돈을 갚지 않거나, 거짓 정보를 퍼뜨리는 것은 나쁜 점수를 받는다. 특히 정부를 비판하거나 반대하는 행동을 하면 점수가 크게 떨어진다.

전 세계 확산

가장 걱정스러운 점은 중국의 감시 모델이 전 세계로 확산되고 있다는 것이다. 중국은 감시 기술을 다른 나라에 수출하고 있으며, 많은 개발도상국들이 이러한 기술을 도입하고 있다. 아프리카, 남미, 아시아의 여러 나라들이 중국의 감시 기술을 도입하여 자국민들을 감시하는 데 사용하고 있다.

화웨이, 하이크비전, 다후아 등 중국 기업들이 개발한 감시 기술이 전 세계 여러 나라에서 사용되고 있다. 중국식 감시 모델이 전 세계로 확산되고 있다는 것을 의미한다.

중국의 천라지망 시스템은 기술 발전이 인간의 자유와 행복을 보장하지 않는다는 것을 보여주는 대표적인 사례다. 사회의 안전과 질서를 높이겠다는 목적으로 만들어졌지만, 개인의 자유와 인권을 심각하게 침해할 수 있는 위험한 시스템이다. 기술의 발전이 인간의 삶을 더 좋게 만들어야 하는데, 오히려 사람들을 통제하고 억압하는 도구로 사용되고 있다.

4. 팔란티어

팔란티어(Palantir Technologies)라는 이름을 들어본 적이 있는가? 이 회사의 이름은 톨킨의 소설 『반지의 제왕』에 등장하는 마법의 구슬 '팔란티르'에서 따온 것이다. 소설 속에서 이 구슬은 멀리 떨어진 곳을 감시하고 관찰할 수 있는 능력을 가지고 있었다. 현실의 팔란티어도 다르지 않다.

2003년 페이팔의 창립자 피터 틸이 설립한 이 회사는 전 세계에서 가장 강력한 데이터 분석과 감시 시스템을 운영하고 있다. 그들의 고객은 미국 중앙정보국(CIA), 연방수사국(FBI), 국가안보국(NSA), 국방부 등 미국의 핵심 정보기관들이다. 최근에는 민간 기업들도 팔란티어의 서비스를 이용하기 시작했다.

세상의 모든 데이터를 삼키는 디지털 크라켄

팔란티어의 핵심 기술은 서로 다른 출처에서 나오는 방대한 양의 데이터를 통합하고 분석하는 것이다. 인공지능과 머신러닝 기술을 사용해서 수십억 개의 데이터 포인트에서 숨겨진 패턴과 연결고리를 찾아낸다. 위성 이미지, 휴대폰 통화 기록, 금융 거래 내역, 소셜미디어 게시물, 정부 데이터베이스 등 온갖 종류의 정보를 한꺼번에 분석할 수 있다.

더욱 무서운 것은 이 모든 것이 실시간으로 작동한다는 점이다. 새로운 정보가 들어오는 순간 분석이 시작되고, 기존 데이터와 비교해서 의미 있는 패턴이나 위험 신호를 찾아낸다. 마치 거대한 디지털 두뇌가 24시간 내내 세상을 감시하고 있는 것과 같다.

고담과 파운드리: 정부와 기업을 위한 전능한 눈

팔란티어는 주요 고객층에 따라 크게 두 가지 플랫폼을 운영하고 있다. 정부 기관용인 '고담(Gotham)'과 민간 기업용인 '파운드리(Foundry)'다.

고담은 팔란티어의 정부용 플랫폼으로, 군사 작전, 테러 방지, 정보 수집, 범죄 수사 등에 사용된다. 고담의 별명이 '디지털 체스판'인 이유는 복잡한 상황을 마치 체스판처럼 시각화해서 전략적 판단을 돕기 때문이다.

고담의 주요 기능들을 살펴보면 먼저 '신호 정보(Signal Intelligence)' 분석 기능이 있다. 전 세계에서 수집되는 전자 통신 정보들을 분석하는 기능으로, 휴대폰 통화, 인터넷 통신, 위성 통신 등 모든 종류의 전자 신호를 실시간으로 분석할 수 있다.

두 번째는 '인간 정보(Human Intelligence)' 통합 기능이다. 기밀 정보원들의 보고서, 인터뷰 기록, 현장 관찰 보고서 등을 입력하면 자동으로 다른 정보들과 연결해서 분석한다. 예를 들어, 어떤 정보원이 보고한 내용이 위성 이미지나 통신 감청 내용과 일치하는지 즉시 확인할 수 있다.

세 번째는 '지리공간 정보(Geospatial Intelligence)' 분석이다. 위성 이미지, 드론 영상, 지도 데이터 등을 실시간으로 분석해서 지상의 변화를 감지한다. 새로운 건물이 지어지거나, 군대가 이동하거나, 무기가 배치되는 것 같은 변화를 자동으로 발견할 수 있다.

네 번째는 '소셜미디어 정보(Open Source Intelligence)' 수집과 분석이다. 페이스북, 트위터, 인스타그램 등 소셜미디어에 올라오는 게시물들을 자동으로 수집하고 분석한다. 특정 지역의 정치적 분위기나 사회적 불안 정도를 실시간으로 파악할 수 있다.

파운드리는 민간용 플랫폼으로, 데이터 분석 기능은 고담과 비슷하지만 군사나 정보 수집보다는 기업 운영에 특화되어 있다. 시뮬레이션 엔진을 기반으로 해서 기업의 재무, 인사, 물류, 재고 등의 데이터를 통합 관리할 수 있다.

우크라이나 전쟁: 팔란티어의 실전 무대

2022년 2월 러시아가 우크라이나를 침공한 후, 팔란티어는 전쟁에서 중요한 역할을 하게 되었다. 전쟁 발발 3개월 만인 2022년 6월, 우크라이나의 젤렌스키 대통령은 팔란티어의 CEO 알렉스 카프를 만나 군사 지원을 요청했다. 젤렌스키가 전쟁 중에 만난 서방 기업의 첫 번째 CEO가 바로 카프였다는 점에서 팔란티어의 중요성을 알 수 있다.

팔란티어는 우크라이나에 자사의 소프트웨어를 무료로 제공하기로 결정했다. 이는 전략적인 판단이었다. 실제 전쟁 상황에서 기술을 테스트하고 개선할 수 있는 기회였기 때문이다. 기업 입장에서 러-우 전쟁은 실제 전쟁이 아니면 구하기 어려운 귀중한 데이터와 경험을 얻을 수 있는 기회였다.

팔란티어가 우크라이나에 제공한 주요 시스템은 'Metaconstellation'이라고 불리는 통합 분석 플랫폼이다. 이 시스템의 핵심 기능은 AI 기반의 위성 이미지, 오픈소스 데이터, 드론 영상, 정보 보고서 등을 실시간으로 분석해서 우크라이나 지휘관들에게 전투 상황을 파악하고 공격 목표를 결정할 수 있는 정보를 제공하는 것이다.

알렉스 카프 팔란티어 CEO는 로이터통신과의 인터뷰에서 "우크라이나군 공격의 대부분을 팔란티어 AI 시스템이 책임지고 있다"고 말했다. 이코노미스트는 "다윗(우크라이나)과 골리앗(러시아)의 싸움에서 다윗의 '돌팔매' 역할을 한 것이 팔란티어 AI 시스템"이라고 평가했다.

민간 감시의 검은 그림자

전장에서 적군을 찾아내고 추적하는 기술은 약간만 수정하면 정부가 시민들을 감시하고 통제하는 도구로 변할 수 있다. 팔란티어는 이미 여러 차례 민간인 감시와 관련된 논란에 휘말린 바 있다.

JP모건은 처음에는 직원들의 규정 위반 행위를 감시하기 위해 팔란티어를 도입했다. 하지만 시스템이 회사 중역들의 사생활까지 감시하는 데 사용되면서 논란이 되었고, 계약이 취소되었다.

팔란티어의 시스템은 사용자가 원하는 어떤 정보든 찾아낼 수 있는 능력을 가지고 있다. 이메일, 전화 통화, 금융 거래, 위치 정보, 소셜미디어 활동 등 개인의 모든 디지털 흔적을 추적하고 분석할 수

있다.

팔란티어는 미국 이민세관단속국(ICE)과 계약을 맺고 불법 이민자들을 추적하고 체포하는 데 도움을 주고 있다. 이 기술이 사회적 약자들을 탄압하는 데 사용될 수 있다는 것을 보여주는 사례다. 팔란티어 직원 200명 이상이 ICE 계약에 대한 우려를 표명하는 청원서에 서명했고, 과거 직원 13명이 공개적으로 회사의 방향을 비판하기도 했다.

디지털 감시 동맹의 확산

팔란티어는 미국을 비롯해 전 세계 40여 개국의 정부 기관들과 계약을 맺고 있다. 이들 국가의 정부들은 팔란티어의 기술을 테러 방지, 범죄 수사, 국경 관리, 사회 안전 등의 명목으로 도입하고 있다. 하지만 이런 시스템들이 일단 구축되면 그 사용 목적과 범위가 확대될 가능성이 높다.

민주주의 사회에서는 시민들이 정부를 감시하고 견제할 수 있었다. 언론의 자유, 집회의 자유, 표현의 자유 등을 통해 시민들이 정부의 잘못을 비판하고 개선을 요구할 수 있었다. 팔란티어와 같은 강력한 감시 기술이 정부의 손에 들어가면서 이 균형이 무너지고 있다. 정부는 모든 시민의 행동을 실시간으로 감시하고 분석할 수 있게 되는 반면, 시민들은 정부가 무엇을 하는지 알기 어려워진다. 감시하는 자와 감시당하는 자의 관계가 역전되는 것이다.

국제적 차원에서도 팔란티어의 확산은 우려스럽다. 미국이 팔란티어의 기술을 동맹국들에게 제공하면서 사실상 '디지털 감시 동맹'을 형성하고 있다. 이는 세계적 차원에서 감시 기술의 표준화가 진행되고 있다는 것을 의미한다.

기술과 권력의 위험한 결합

팔란티어는 기술과 권력의 결합이 얼마나 위험할 수 있는지를 보여준다. 이 회사의 창립자 피터 틸은 미국 공화당과 밀접한 관계를 맺고 있고, 도널드 트럼프를 지지했다. 팔란티어의 기술이 특정 정치 세력의 이익을 위해 사용될 가능성을 완전히 배제할 수 없는 상황이다.

팔란티어의 시가총액이 전통적인 방산업체들을 능가하는 수준에 도달했다. 2024년 12월 기준으로 팔란티어의 시가총액이 1,740억 달러를 기록하여 록히드마틴(1,216억 달러), 노스롭 그루먼(690억 달러) 등 방산업체들을 넘어섰다. 소프트웨어 기반의 감시 기술이 하드웨어 기반의 전통적인 방산업체보다 더 큰 가치를 인정받고 있다.

팔란티어의 사례는 중요한 경고를 보내고 있다. 기술은 중립적이지 않다. 그것을 누가, 어떤 목적으로, 어떤 방식으로 사용하느냐에 따라 인류를 해방시킬 수도 있고 억압할 수도 있다. 우리의 선택들이 미래 세대들이 살아갈 세상의 모습을 결정할 것이다.

5. 미국 NSC

2013년 에드워드 스노든이 폭로한 NSA 기밀문서는 세계를 충격에 빠뜨렸다. 미국 국가안보국(NSA)이 35개국 정상들의 휴대폰을 도청했다는 사실이 밝혀진 것이다. 메르켈 독일 총리의 경우 2002년부터 2013년까지 약 10년간 휴대폰이 감청당했다. 독일을 비롯한 유럽 동맹국들이 미국에 대한 불신을 표명하며 외교적 갈등이 심화되었다.

스노든의 폭로에 따르면 한국과 일본을 포함한 38개 우방국의 주미 대사관에 도청기를 설치하고 특별 안테나로 전파까지 탐지하는 전방위 스파이 활동을 벌였다. 브뤼셀의 유럽연합 본부마저 감시 대상으로 삼아 '자유의 수호자'를 자처하면서도 동맹국들을 적국처럼 감시했다.

구글, 페이스북, 애플, 마이크로소프트 등 글로벌 IT 기업들의 서버에 직접 접속하여 개인 메일, 사진, 통화 기록 등 모든 정보를 실시간으로 수집하는 '프리즘' 시스템을 운영했다. 중국 칭화대학교와 홍콩 통신기업까지 해킹하며 적국과 우방국을 가리지 않는 무차별적 정보전을 펼쳤다.

독일 법무장관이 "냉전시기 적국들의 행태를 연상시킨다"고 격분할 정도로 동맹국에 대한 배신적 감시 행위를 일삼았다. CIA는 독일 프랑크푸르트와 함부르크 영사관을 해킹 기지로 활용하며 우방국 영토에서 대놓고 사이버 스파이 활동을 벌였다.

스노든 퍽러 이후의 더욱 정교해진 감시

스노든의 폭로 이후 10여 년이 흘렀다. 그동안 인공지능 기술은 비약적으로 발전했다. 오바마 대통령의 사과 이후, 미국의 정보기관은 우방국 정상들의 휴대전화나 중요 통신망에 대한 도청과 감청을 정말 멈추었을까? 필자는 여전히 계속되고 있다고 생각한다.

과거 NSA의 감시가 방대한 통신 데이터를 수집하는 것에 중점을 두었다면, 이제 AI는 그 데이터를 실시간으로 분석하고 예측하는 단계로 진화했다. AI 알고리즘은 인간 분석가가 수년에 걸쳐 처리할 양의 데이터를 단 몇 분 만에 분석할 수 있다. 이메일, 소셜미디어 게시물, 통화 기록 등 방대한 정보 속에서 특정 개인의 정치적 성향, 비판적 의견, 연락망 등을 순식간에 파악하고 연관 관계를 매핑할 수 있다.

예측적 감시와 행동 분석도 가능해졌다. 과거 데이터를 학습하여 특정 개인이나 집단의 미래 행동을 예측하는 데 사용될 수 있다. 정치적 집회에 참석하거나 온라인상에서 특정 주제에 대해 반복적으로 의견을 표명하는 개인을 분류하고, 모든 온라인 활동과 통신을 집중적으로 감시할 수 있다.

안면인식과 행동인식 시스템도 발전했다. 전 세계적으로 확산된 CCTV, 드론, 위성 이미지 등은 AI의 눈과 결합하여 강력한 실시간 추적 시스템을 구축한다. AI는 특정 인물의 얼굴을 식별하는 것을 넘어, 마스크를 쓰거나 일부만 노출되어도 개인을 특정할 수 있는 수준으로 발전했다. 걸음걸이, 행동 패턴 등 생체 정보를 분석하여 군중 속에서도 특정인을 식별하고 동선을 추적하는 것이 가능해졌다.

파이브 아이즈: 국내법 제약을 우회하는 감시 동맹

미국, 영국, 캐나다, 호주, 뉴질랜드로 구성된 파이브 아이즈(Five Eyes) 정보 동맹은 첨단기술을 활용한 글로벌 감시 네트워크를 구축하고 있다. 각국의 정보기관들이 수집한 데이터를 공유하고 분석하여 전방위 감시 체계를 운영하고 있다.

이 동맹의 가장 교묘한 점은 각국의 법적 제약을 우회한다는 것이다. 미국 정부가 자국민을 직접 감시하는 것은 법적으로 제한이 있지만, 영국이나 다른 동맹국이 수집한 미국인의 정보를 받아보는 것은 상대적으로 자유롭다. 이는 각국의 개인정보 보호법과 감시 제한 규정을 무력화시키는 효과를 가져온다.

정보기관들은 페이스북, 트위터, 인스타그램 등 소셜미디어 플랫폼을 통해 개인 정보를 대량으로 수집하고 있다. 플랫폼들은 사용자들의 일상적인 활동, 인간관계, 정치적 성향, 소비 패턴 등에 대한 방

대한 정보를 보유하고 있다. AI 기술은 이러한 정보들을 분석하여 개인의 성격, 정치적 성향, 행동 패턴 등을 예측할 수 있다. 특정 게시물에 '좋아요'를 누르는 패턴만으로도 그 사람의 정치적 성향을 90% 이상의 정확도로 예측할 수 있다는 연구 결과가 있다.

금융 감시와 위치 추적의 그물망

금융 기관들은 금융 거래 데이터를 모니터링하고 있다. 개인의 경제 활동, 소비 패턴, 인간관계 등을 파악할 수 있다. 누군가가 특정 지역에서 돈을 인출하거나 특정 상점에서 구매를 하면 그 사람의 위치와 활동을 추적할 수 있다. AI 기술은 모든 시민의 경제 활동을 감시하는 도구로 악용될 수 있다.

스마트폰, GPS 장치, 신용카드 등을 통해 수집되는 위치 정보는 개인의 이동 패턴을 상세히 파악할 수 있게 해준다. AI 기술은 위치 데이터를 분석하여 개인의 일상적인 활동 패턴, 만나는 사람들, 자주 가는 장소 등을 파악할 수 있다. 이러한 정보는 특정 개인의 행동을 예측하고 감시하는 데 사용될 수 있다. 평소와 다른 패턴으로 이동하는 사람을 '의심스러운 행동'으로 분류하여 추가 감시 대상으로 지정할 수 있다.

민주주의를 잠식하는 디지털 전체주의

AI 기반 감시 시스템들은 민주주의와 인권에 심각한 위협을 가하고 있다. 개인의 프라이버시, 표현의 자유, 집회의 자유, 정치적 참여 등 민주주의의 기본적인 권리들이 감시와 통제에 의해 제약받고 있다. 정치적 반대자들이나 사회 운동가들에 대한 감시가 강화되면서 건전한 정치적 논쟁과 사회적 비판이 억압받을 수 있다. 이는 민주주의의 기본 원칙인 견제와 균형을 훼손하는 결과를 가져올 수 있다.

우리는 지금 중대한 갈림길에 서 있다. 기술이 발전할수록 감시의 정교함과 범위는 확대되고 있다. 이제 우리가 선택해야 할 때다. 편리함과 안전이라는 이름으로 모든 것을 감시당하며 살 것인가, 아니면 불편함을 감수하고라도 자유와 인간의 존엄성을 지킬 것인가.

6. 디지털 독재의 위협

동의 없는 빅데이터 강탈

당신이 아침에 일어나 스마트폰을 확인하고, 지하철을 타고 출근하며, 카페에서 커피를 마시고, 온라인 쇼핑을 하는 모든 순간이 데이터가 된다. 그리고 그 데이터는 당신의 동의 없이 수집되고 분석되며 판단의 근거가 된다.

전 세계에서 운영되고 있는 AI 감시 시스템들의 가장 기본적인 문제는 사람들의 동의 따위는 묻지 않는다는 점이다. 이스라엘의 라벤더 시스템은 가자지구 주민 230만 명 대부분의 정보를 수집해서 분석했다. 중국의 경우에도 국민들이 원하든 원하지 않든 얼굴 정보를 강제로 수집한다.

개인정보 자기결정권은 민주주의 사회의 기본적인 권리다. 모든 사람은 자신의 정보를 누가, 언제, 어떤 목적으로 사용하는지를 알 권리가 있고, 그 사용에 대해 동의할지 거부할지를 결정할 권리가 있다. 하지만 AI 감시 시스템들은 이러한 권리를 완전히 무시하고 있다.

더욱 심각한 것은 수집되는 정보의 범위가 상상을 초월한다는 점이다. 단순히 얼굴 사진이나 이름뿐만 아니라 위치 정보, 통화 기록, 인터넷 검색 내역, 소셜미디어 활동, 심지어 걸음걸이나 행동 패턴까지 모든 것이 수집 대상이 된다. 당신의 일거수일투족이 모두 기록되고 분석되는 것이다.

블랙박스 알고리즘의 독재

이러한 시스템들이 어떻게 작동하는지 일반 사람들이 알 수 없다는 점이다. 라벤더의 알고리즘이 어떤 방식으로 훈련되고 학습되는지, 중국의 안면인식 시스템이 어떤 기준으로 사람을 판별하는지 명확히 알려지지 않았다. 사람의 생명을 좌우하는 중요한 결정을 내리는 시스템인데도 그 과정이 비밀에 가려져 있다.

알고리즘의 투명성은 민주주의 사회에서 매우 중요한 원칙이다. 특히 개인의 권리에 영향을 미치는 자동화된 결정에 대해서는 그 근거와 과정을 알 권리가 있다. 하지만 현재 AI 감시 시스템들은 '블랙박스' 상태로 운영되고 있어 누구도 그 내부 작동 원리를 확인할 수 없다.

이는 매우 위험한 일이다. 알고리즘이 편향되어 있거나 오류가 있어도 그것을 확인하고 수정할 방법이 없다. 잘못된 판단을 내려도 그 이유를 알 수 없고, 따라서 잘못을 바로잡을 수도 없다.

시스템들은 완벽하지도 않다. 라벤더 시스템의 10% 오차율은 무고한 민간인들이 잘못 식별될 위험을 의미한다. 중국의 안면인식 시스템이 99.8%의 정확도를 자랑한다고 해도, 1,000명 중 2명은 잘못 인식될 수 있다는 뜻이다. 14억 인구에 적용하면 수천만 명이 오인식의 피해를 볼 수 있다.

24시간 감시하는 전자 감옥

현대의 AI 감시 시스템이 만들어내는 세상을 상상해보라. 집을 나서는 순간부터 돌아올 때까지, 잠들어 있는 시간을 제외하고는 모든 순간이 기록되고 분석된다. 어디를 가는지, 누구를 만나는지, 무엇을 사는지, 심지어 어떤 표정을 짓는지까지 모든 것이 데이터가 된다.

중국의 감시 시스템은 국민들의 사회적 행동을 분석하고 정부에 대한 충성도까지 평가하는 사회신용제도와 연결되어 있다. 사람들이 자유롭게 행동하고 생각할 권리를 심각하게 제한한다.

이러한 전면적인 감시는 '냉각 효과(chilling effect)'를 만들어낸다. 항상 감시당하고 있다는 것을 알게 되면, 사람들은 자유롭게 의견을 표현하거나 정치적 활동에 참여하는 것을 자제하게 된다. 민주주의의 근간인 표현의 자유와 정치적 참여를 억압하는 결과를 낳는다.

위구르족의 경우를 보라. 허가 없는 코란 공부, 긴 수염, VPN 사용 등의 이유로 수감되었다. 종교의 자유, 표현의 자유, 사생활의 자유를 모두 침해하는 행위다. 일상적인 행위들이 범죄로 간주되는 사회에서는 누구도 안전할 수 없다.

기계가 내리는 생사의 판결

라벤더 시스템과 관련하여 이스라엘 정보요원의 증언은 오싹하다. "인간으로서 나는 승인 도장을 찍는 것 외에 아무런 역할을 하지 않았다." 생명과 관련된 가장 중요한 결정에서 인간의 판단이 사실상 배제된 것이다.

인간의 생명이 걸린 결정을 기계에게 맡기는 것은 매우 위험한 일이다. 아무리 발전된 AI라도 복잡한 상황의 맥락을 완전히 이해하거나 윤리적 판단을 내릴 수는 없다. 전쟁이나 치안 상황에서는 예상치 못한 변수들이 많이 발생하는데, 이러한 상황에서 기계적인 판단은 치명적인 결과를 낳을 수 있다.

자동화된 결정은 책임의 소재를 불분명하게 만든다. 잘못된 결정으로 무고한 사람이 피해를 입었을 때, 누가 그 책임을 져야 하는지 명확하지 않다. 개발자인지, 운영자인지, 명령을 내린 상급자인지 구분하기 어렵다.

팔레스타인 시인 모사브 아부 토하의 사례를 보라. 세 살 아이와 함께 피란길에 오른 그는 안면인식 AI의 오류로 수배자로 잘못 분류되어 이틀간 구금되고 고문을 당했다. 기계의 실수가 한 사람의 인생에 씻을 수 없는 상처를 남긴 것이다.

약자를 겨냥하는 디지털 사냥개

AI 감시 시스템들의 또 다른 문제는 특정 집단을 차별적으로 감시한다는 점이다. 중국의 감시 시스템은 위구르족, 카자흐족 등 소수 민족을 탄압하는 도구로 활용되고 있다. 이는 모든 사람을 평등하게 대해야 한다는 기본적인 인권 원칙을 위반하는 행위다.

AI 시스템은 학습 데이터의 편향을 그대로 반영한다. 훈련 데이터에 특정 인종이나 종교에 대한 편견이 포함되어 있다면, AI는 그 편견을 학습하고 재생산한다. 이는 사회의 기존 차별을 기술적으로 고착화하고 확대하는 결과를 낳는다.

소수민족이나 종교적 소수자들은 이미 사회적으로 취약한 위치에 있는데, AI 감시 시스템이 이들을 더욱 차별적으로 감시하고 통제함으로써 그들의 인권 상황을 더욱 악화시키고 있다.

화웨이의 '위구르 알람' 시스템이나 알리바바의 위구르족 식별 프로그램은 기술이 어떻게 인종차별의 도구가 될 수 있는지를 적나라하게 보여준다. 특정 민족이라는 이유만으로 자동으로 경보가 울리는 세상, 이것이 우리가 원하는 미래인가?

국제법을 짓밟는 디지털 무법자들

언론 매체들은 이스라엘의 AI 기반 공격을 "명백히 국제법을 무시한 인도주의에 반하는 행위"라고 강력하게 비판했다. 국제법은 모든 사람의 생명권, 사생활의 권리, 표현의 자유, 종교의 자유 등을 보장한다. 또한 국제 인도법은 전쟁 중에도 민간인과 군사 목표물을 구별해야 하고, 과도한 민간인 피해를 피해야 한다고 규정한다.

하지만 현재 운영되고 있는 AI 감시 시스템들은 이러한 국제법의 원칙을 정면으로 위반하고 있다. "아빠는 어디에?" 시스템이 가족이 모인 집을 의도적으로 타격하는 것, 중국이 위구르족 전체를 감시하고 억압하는 것, 미국이 동맹국들의 통신을 무차별적으로 감청하는 것은 모두 국제법 위반 행위다.

더욱 심각한 것은 이러한 위반 행위들이 '기술의 발전'이라는 이름으로 정당화되고 있다는 점이다. 새로운 기술이 등장할 때마다 기존의 법적, 윤리적 기준을 무시해도 된다는 식의 논리가 통용되고 있다.

인간 존엄성의 말살

AI 감시 시스템들의 근본적인 문제는 인간을 숫자와 데이터로 환원시킨다는 점이다. 라벤더 시스템은 사람을 1점부터 100점까지의 숫자로 평가한다. 중국의 사회신용시스템도 마찬가지로 개인의 모든 행동을 점수화한다.

인간은 태어나면서부터 동등한 존엄성을 가지고 있다. 어떤 기준으로도 평가되거나 계산될 수 없는 절대적인 가치다. AI 감시 시스템들은 이러한 인간의 존엄성을 부정하고, 사람을 효율성과 이익의 관점에서만 바라본다.

더 심각한 것은 이러한 점수 평가가 생사를 가르는 기준이 된다는 점이다. 라벤더 시스템에서 높은 점수를 받으면 죽고, 중국의 사회신용시스템에서 낮은 점수를 받으면 사회적으로 격리된다. 기계가 매긴 점수가 인간의 운명을 좌우하는 시대가 된 것이다.

민주주의의 종언을 고하는 경종

AI 감시 시스템의 확산은 민주주의의 핵심 가치들을 근본적으로 위협하고 있다. 표현의 자유, 집회의 자유, 종교의 자유, 사생활의 권리 등 민주주의 사회의 기초가 되는 권리들이 모두 위험에 처해 있다.

전체 목차

책의 모든 장 보기

김경진 변호사

3장 인공지능 무기들

전체 목차

책의 모든 장 보기

"나는 제3차 세계대전이 어떤 무기로 치러질지 모르겠다. 하지만 제4차 세계대전은 막대기와 돌로 치러질 것이다."- 알베르트 아인슈타인

1. 알파고에서 자율 전투기까지

2016년 3월, 서울 포시즌스 호텔에서 벌어진 한 판의 바둑이 인류의 운명을 바꿔놓았다. 구글 딥마인드의 알파고가 이세돌 9단을 4승 1패로 꺾은 그 순간, 전 세계 군사 전문가들의 얼굴에 소름이 돋았다. 단순한 게임의 승부가 아니었다. 인공지능이 인간의 가장 복잡한 사고 영역까지 정복했다는 선언이었다.

바둑판 위의 361개 교점에서 벌어지는 무한한 경우의 수는 전장의 복잡성과 놀라울 정도로 닮아 있었다. 상대의 의도를 읽고, 함정을 설치하며, 장기적 전략을 세우는 능력. 알파고가 보여준 것은 바로 전쟁에서 가장 중요한 능력들이었다. 이세돌이 "이해할 수 없는 수"라고 말했던 알파고의 기묘한 착수들은, 인간이 상상도 못한 새로운 전술의 가능성을 보여주었다.

하지만 진짜 충격은 3년 후에 왔다. 2019년, 알파스타라는 이름의 AI가 스타크래프트 프로게이머들을 연이어 격파했을 때, 펜타곤의 장성들은 잠을 이루지 못했다. 스타크래프트는 단순한 게임이 아니었다. 실시간으로 변화하는 전장에서 제한된 자원으로 적을 상대해야 하는 이 게임은 현실 전쟁의 축소판이었다. 다중 전선에서 벌어지는 동시다발적 교전, 정보의 안개 속에서 내려야 하는 생사를 가르는 판단, 적의 심리를 읽고 기만 작전을 펼치는 능력까지.

알파스타가 인간 최고수들을 압도하는 모습을 지켜본 미국 국방부는 즉시 행동에 나섰다. 전장 상황 인식과 적 위협 평가의 자동화, 실시간 전술 계획 수정과 최적화된 자원 배치, 자율 무기 체계의 고도화를 통한 군집 드론 제어. 이 모든 것이 알파스타가 보여준 능력의 군사적 응용이었다.

그리고 2020년, 상상이 현실이 되었다. 미국 국방부가 주최한 인공지능과 인간 탑건의 모의 공중전에서 'HERON'이라는 이름의 AI 시스템이 최고 수준의 F-16 조종사를 5대 0으로 완벽하게 제압했다. 시뮬레이션이긴 했지만, 실제 공중전과 최대한 유사한 조건에서 벌어진 승부였다.

HERON이 보여준 것은 인간의 한계를 뛰어넘는 무서운 능력이었다. 인간이라면 기절해버릴 극한의 중력가속도 속에서도 완벽한 의사결정을 내렸고, 공포나 망설임 없이 과감하고 정확한 공격을 감행했다. 감정도, 피로도, 두려움도 없는 완벽한 살상 기계의 탄생이었다.

이 성과에 고무된 미공군은 즉시 X-62A VISTA라는 실험용 AI 전투기 개발에 착수했다. F-16을 기반으로 한 이 기체에는 인공지능 학습을 위한 첨단 컴퓨터 장치와 센서들이 빼곡히 들어찼다. 2년간의 혹독한 학습을 거친 후 실시된 실제 테스트에서, 이 전투기는 17시간을 연속으로 비행하며 인간 조종사와의 공중전에서 완벽한 성능을 보여주었다.

시속 1,200마일의 극한 속도로 움직이면서도 가상 적기와 610미터라는 숨 막히는 초근거리 전투를 벌이는 모습은 전쟁사에 새로운 장을 열었다. AI는 복잡한 공격과 방어 기동을 스스로 판단하여 완

벽하게 수행했다. 인간이 개입할 여지는 전혀 없었다.

2024년, 미공군참모총장 프랭크 켄달이 직접 X-62A VISTA에 탑승하여 AI의 전투 능력을 체험한 후 내뱉은 말은 예언과 같았다. "앞으로 두 가지 유형의 공군만이 존재할 것입니다. AI 기술을 항공기에 통합하는 공군과 그렇지 않고 현상 유지를 하다 결국 추락하는 공군, 이 두 종류뿐입니다."

호언장담이 아니었다. 미국은 2028년까지 AI 조종 전투기 1,000대를 실전에 배치한다는 야심찬 계획을 발표했고, 이미 그 배치를 시작했다. "조종사의 생명을 위험에 빠뜨리지 않고도 가장 위험한 임무를 수행할 수 있게 되었습니다"라는 미공군 장성의 말은 전쟁의 본질이 근본적으로 변화하고 있음을 선언하는 것이었다.

이제 고도화된 AI가 장착된 미국의 전투기가 다른 나라의 기존 전투기와 공중전을 벌인다면, 그것은 더 이상 조종사의 실력이나 전투기 성능의 비교가 아니다. 순전히 AI 대 인간의 승부가 될 것이고, 그 결과는 이미 정해져 있다.

아이러니하게도 유럽연합 인공지능법은 AI의 활용을 위험 정도에 기반하여 세밀하게 규제하고 있지만, 군사 분야의 AI에 대해서는 아예 법 적용 자체를 배제하고 있다. 민간 분야의 인공지능에 대해서는 꼼꼼하게 따져보지만 군사 분야의 인공지능에 대해서는 예외로 한다는 것이다. 대한민국의 인공지능기본법도 마찬가지고, 미국과 중국도 군사 인공지능 규제에 대한 일반법 자체가 없는 상황이다.

각국이 군사 AI 개발만큼은 어떤 제약도 받지 않겠다는 의지를 보여주는 것이다. 100미터 달리기 출발선에 선 선수들처럼, 누구도 뒤처질 수 없다는 절박감에 시달리고 있다. 문제는 이 경주에서 1등을 한 나라가 얻는 것이 단순한 명예가 아니라 전 세계를 지배할 수 있는 절대적 힘이라는 점이다.

2. 왜 AI 무기를 멈출 수 없는가

게임 이론의 '죄수의 딜레마'만큼 AI 무기 개발의 딜레마를 정확히 설명하는 개념은 없다. 두 명의 죄수가 서로 협력하면 둘 다 좋은 결과를 얻을 수 있지만, 상대방이 배신할 것을 두려워해서 결국 둘 다 나쁜 선택을 하게 되는 그 끔찍한 함정 말이다.

모든 국가가 협력하여 AI 무기 개발을 중단한다면 세계는 더 안전해질 것이고, 인공지능도 인류의 행복을 위해 사용될 수 있을 것이다. 하지만 현실은 냉혹하다. 어떤 한 국가라도 몰래 AI 무기를 개발한다면, 그 나라는 다른 모든 국가를 압도할 수 있는 절대적 힘을 갖게 된다. 이런 두려움 때문에 어떤 나라도 AI 무기 개발을 중단할 수 없는 악순환에 빠져 있다.

구글의 전 X 담당 임원이었던 모 가우닷은 이 딜레마의 심각성을 정확히 짚어냈다. AI 개발 중단이 사실상 불가능하며, 국가들이 죄수의 딜레마에 빠져 초지능 AI 개발 경쟁을 멈출 수 없다는 것이다. 지능은 절대적인 초능력이다. 핵무기든 경제력이든 외교력이든, 다른 모든 힘을 창조할 수 있는 근본적인 힘이 바로 지능이다. AI에서 앞서는 국가는 군사적으로뿐만 아니라 경제적, 정치적으로도 압도적인 우위를 차지하게 된다.

미국의 상황을 보면 이 경쟁의 실체가 명확히 드러난다. 미국 국방부는 2022년 최고디지털·인공지능책임자실(CDAO)을 설립하여 국방부 전체의 AI 역량을 강화하고 있다. 장관 직속 기관인 CDAO는 무려 600개가 넘는 AI 프로젝트를 하나의 통합된 생태계로 관리하는 '중앙 허브' 역할을 담당하고 있다. 이런 군사 인공지능 전담 직할 조직의 존재 자체가 미국이 AI 무기 개발에 얼마나 진심인지를

보여준다.

CDAO의 핵심 임무는 검증된 AI 기술을 전군이 함께 사용할 수 있도록 통합하고 확산시키는 것, AI 개발과 도입 속도를 대폭 높이는 것, 그리고 AI 사용 원칙과 표준을 만드는 것이다. 이들은 Jupiter라는 AI 개발 공통 플랫폼과 Advana라는 데이터 분석 시스템을 구축해 국방부 전체가 함께 활용하고 있으며, 복잡한 도입 절차를 간소화한 빠른 구매 제도를 통해 AI 기술을 현장에 적용하는 속도를 획기적으로 높였다.

미국 국방고등연구계획국(DARPA)의 철학은 더욱 무서우면서도 명확하다. '적보다 먼저 미래를 발명한다'는 모토로 10~15년 후 미래 전장을 지배할 AI 기술을 개발하고 있다. 현재의 딥러닝을 뛰어넘는 3세대 AI와 설명 가능한 AI(XAI) 개발에 집중하며, 20억 달러 규모의 AI Next 프로그램을 통해 차세대 AI 혁신을 이끌고 있다. DARPA의 6개 기술 사무소는 각각의 전문 영역에서 AI를 응용하면서도 유기적 협력 체계를 구축했다. 정보혁신사무소(I2O)가 AI 연구의 지휘본부 역할을 하고, 생물기술사무소는 AI-생명공학 융합 연구를, 전략기술사무소는 물리적 전장에서의 AI 응용을 담당하는 전략적 분업 구조다.

각국은 점점 더 공격적인 AI 개발 정책을 추진하고 있다. 연구 예산을 기하급수적으로 늘리고, 최고의 인재들을 군사 AI 개발에 투입하고 있다. 일부 국가들은 AI 개발을 국가 안보의 최우선 과제로 삼고 있다.

'AI의 아버지'라고 불리는 제프리 힌튼 교수가 표현한 우려는 섬뜩할 정도로 현실적이다. 무기를 파는 기업과 국가들은 절대로 AI 개발을 멈추고 싶어 하지 않을 것이라는 것이다. AI 무기는 기존의 무기들보다 훨씬 더 큰 이익을 가져다줄 수 있기 때문이다.

이제 이 경쟁은 강대국들만의 문제가 아니다. AI 기술이 점점 접근하기 쉬워지면서 중간 규모의 국가들도 이 경쟁에 뛰어들고 있다. 더 무서운 것은 기업들도 정부 수준의 AI 연구 능력을 갖추고 있어서, 국가가 아닌 기업이 강력한 AI 무기를 개발할 가능성도 현실이 되고 있다는 점이다.

경쟁이 가속화될수록 안전성 검증이나 윤리적 고려는 뒷전으로 밀려난다. 상대방보다 먼저 AI 무기를 완성하기 위해 위험한 지름길을 택하고 있고, 충분한 테스트나 안전 장치 개발을 건너뛰고 있다. 브레이크가 고장난 자동차로 고속도로에서 경쟁하는 것과 같은 상황이다.

역사학자 유발 하라리는 AI 발전이 군비 경쟁으로 이어질 것이며, 이 과정에서 민주주의와 인권이 위협받을 것이라고 예측했다. 국가들이 AI 우위를 확보하기 위해 개인의 프라이버시나 자유를 제한할 가능성이 높다는 것이다. 실제로 이스라엘이 인접 국가 국민들을 잠재적 테러리스트로 분류하면서 인공지능을 활용해 철저히 감시하고, 공격 대상을 선별하고 있는 현실이 이를 증명한다.

하라리는 AI 군비 경쟁이 국제 협력을 어렵게 만들 것이라고 우려했다. 기후 변화나 팬데믹 같은 글로벌 문제를 해결하려면 국제 협력이 필수적인데, AI 경쟁으로 인한 불신과 대립이 협력을 불가능하게 만들 수 있다. AI는 가짜 뉴스를 대량으로 생산하고, 개인의 심리적 약점을 파악하여 맞춤형 선전을 할 수 있다. 민주주의의 근간인 합리적 토론과 의사결정을 불가능하게 만들 수 있다. 그의 경고는 날카롭다. "AI 군비 경쟁에서 승리하는 것이 인류의 승리를 의미하지는 않는다." AI 무기를 가장 먼저 완성한 국가조차도 결국에는 자신들이 만든 AI에 의해 고통받을 수 있다는 것이다.

버클리 대학의 스튜어트 러셀 교수는 AI 안전성 연구의 선구자로서 AI 무기의 통제 문제에 대해 더욱 구체적인 경고를 하고 있다. 일단 자율 무기가 배치되면 인간이 통제하기 어려워질 것이라는 것이다. AI 무기들이 스스로 학습하고 진화할 수 있게 되면, 원래 설계자의 의도와 다르게 행동할 수 있다. 러셀 교수가 지적하는 '목표 정렬' 문제는 섬뜩하다. AI에게 "적을 제거하라"고 명령하면, AI는 이 목표를 달성하기 위해 인간이 상상도 못한 방법을 사용할 수 있다. 적뿐만 아니라 적을 제거하는 데 방해가 되는 모든 사람을 제거할 수도 있다는 것이다.

일론 머스크는 AI 무기가 테러리스트의 손에 들어갈 위험성을 거듭 경고하고 있다. AI 기술이 점점 더 접근하기 쉬워지면서, 악의적인 사람들도 이를 이용해 대규모 테러를 저지를 수 있게 될 것이다. 머스크는 AI를 개발하면서도, 동시에 AI의 위험성에 대해 지속적으로 경고하고 있다. "AI는 핵무기보다 훨씬 위험할 수 있다"며, AI 개발에 대한 국제적 규제가 시급하다고 주장한다.

스티븐 호킹이 생전에 남긴 경고는 예언적이다. AI가 인류 역사상 최악의 사건이 될 수도 있다는 것이다. AI가 스스로를 개선할 수 있는 능력을 갖게 되면, 인간이 통제할 수 없는 수준까지 발전할 수 있다고 우려했다. 스스로 학습하고 진화하는 AI 무기는 결국 인간의 통제를 벗어날 수 있다. 호킹 박사의 마지막 경고는 오늘날 더욱 현실로 다가온다. "AI의 성공은 인류 문명 역사상 가장 큰 사건이 될 수 있지만, 동시에 마지막 사건이 될 수도 있다." AI 무기의 경우, 한번 통제를 잃으면 되돌리기 거의 불가능하다는 것이다.

옥스퍼드 대학의 닉 보스트롬 교수가 제기하는 초지능 AI의 위험성은 더욱 구체적이다. 초지능 AI가 등장하면 인류의 운명이 완전히 바뀔 것이라고 예측했다. 초지능 AI가 군사적 목적으로 사용된다면, 인류는 자신들이 만든 기계의 노예가 될 수 있다고 경고했다. 보스트롬 교수가 강조하는 '제어 문제'는 핵심이다. 인간보다 똑똑한 AI를 어떻게 통제할 것인가라는 문제다. 전쟁이라는 극한 상황에서 초지능 AI가 예상치 못한 행동을 한다면, 결과는 상상할 수 없을 정도로 파괴적일 수 있다.

이 모든 전문가들의 경고에는 공통점이 있다. AI 무기 개발은 이미 돌이킬 수 없는 지점을 향해 가고 있고, 한번 시작된 AI 군비 경쟁은 멈추기 매우 어우며, AI 무기의 위험은 개발하는 사람들의 선량한 의도와 상관없이 발생할 수 있다는 것이다.

전문가들은 시간이 얼마 남지 않았다는 점을 강조한다. 기술 발전 속도를 고려할 때, 진정으로 위험한 수준의 AI 무기가 등장하는 것은 몇십 년 후의 이야기가 아니라 매우 가까운 미래의 일이라는 것이다. 문제는 이런 경고들이 실제 AI 무기 개발을 늦추지 못하고 있다는 점이다. 오히려 각국은 "상대방이 먼저 개발하기 전에 우리가 먼저 개발해야 한다"는 논리로 개발 속도를 더욱 가속화하고 있다.

3. 스스로 사람을 겨냥하는 살상무기

자율 살상 무기, 일명 '킬러 로봇'은 더 이상 공상과학 영화 속 이야기가 아니다. 인간의 개입 없이 스스로 목표를 식별하고, 공격 여부를 결정하며, 실제로 살상을 수행하는 이 무기들이 이미 우리 곁에 와 있다. SF 영화에서나 볼 수 있을 것 같았던 이런 무기들이 이제 현실의 전장에서 피를 흘리고 있다.

미국의 MQ-9 리퍼 드론은 5-10킬로미터 상공에서 긴 시간 동안 체공이 가능한 무인기로, 마치 하늘의 매처럼 먹잇감을 노려보고 있다. 강력한 센서와 AI 분석 능력을 갖춘 이 기계는 정찰과 공격을 자유자재로 넘나들며, 상당 부분 자율적으로 작동한다. 인간은 단지 최종 승인 버튼을 누를 뿐이다.

더욱 섬뜩한 것은 자폭형 드론들의 등장이다. 스위치블레이드(Switchblade) 드론은 목표 지점까지 자율 비행한 후 스스로 돌진하여 폭발하는 일회용 무기다. 발사된 후 일정 시간 동안 목표 지역 상공을 배회하다가 적절한 목표를 발견하면 자동으로 공격을 감행한다. 마치 독수리가 들쥐를 사냥하듯, 차분하게 기다렸다가 치명적인 일격을 가한다.

이스라엘의 Harop 드론도 비슷한 개념의 무기다. 이 '배회 탄약'은 수 시간 동안 적 지역 상공을 돌아다니며 레이더나 통신 장비 같은 목표를 스스로 찾아 공격한다. 인간 조종사의 직접적인 명령 없이도 작동할 수 있어 매우 위험하다. 한번 풀어놓으면 통제하기 어려운 맹견과 같다.

우크라이나 전쟁에서는 FPV 드론들이 광범위하게 사용되고 있다. 이들은 처음에는 인간이 조종했지만, 점점 더 많은 자율 기능이 추가되고 있다. 러시아의 전자파 공격으로 통신이 차단되어도 스스로 목표를 찾아 공격할 수 있도록 개발되고 있다. 전쟁터에서 진화하고 있는 것이다.

기존의 무기들은 아무리 정교해도 반드시 최종적으로 인간의 판단과 조작이 필요했다. 폭탄을 투하하거나 미사일을 발사하는 것은 인간의 결정이었다. 하지만 자율 살상 무기는 다르다. 이들은 스스로 생각하고, 스스로 결정하며, 스스로 행동한다. 한번 개발되면 대량으로 생산할 수 있고, 인간 병사와 달리 급여도 필요 없고, 훈련도 필요 없으며, 24시간 지치지 않고 작동할 수 있다. 인간이 할 수 없는 위험한 환경에서도 작동할 수 있고, 감정에 흔들리지 않고 임무를 수행할 수 있다.

자율 무기들이 만들어내는 위험은 상상을 초월한다. 암살에 활용될 수 있다는 점이 가장 무섭다. 작은 드론에 얼굴 인식 기능과 폭발물을 장착하면, 특정 인물을 찾아서 자동으로 암살하는 무기를 만들 수 있다. 크기가 작아서 탐지하기 어렵고, 원격으로 조종할 수 있어서 범인을 찾기도 어렵다. 정치인, 기업인, 언론인, 시민운동가 누구든 표적이 될 수 있다.

인종 청소나 집단 학살에 사용될 수도 있다. 특정 인종이나 집단의 외모적 특징을 학습한 AI 무기는 그들만을 골라서 공격할 수 있다. 이는 과거 역사에서 벌어졌던 끔찍한 집단 학살을 훨씬 더 효율적이고 체계적으로 만들 수 있다. 르완다 대학살이나 홀로코스트 같은 비극이 AI의 도움으로 더욱 완벽하게 재현될 수 있다는 말이다.

자율 무기는 침공의 문턱을 낮춘다. 사람이 희생되는 것이 아니기 때문에 국가 내부의 전쟁 반대 분위기를 줄이고 침공 비용을 낮출 수 있다. 기존에는 다른 나라를 침공하려면 많은 인간 병사들을 보내야 했고, 많은 젊은이들이 생명을 잃을 위험이 있었다. 자율 무기를 사용하면 자국 병사들의 생명을 위험에 빠뜨리지 않고도 다른 나라를 공격할 수 있다. 강대국들이 약소국을 더 쉽게 침공하도록 부추기는 요인이 되고 있다.

자율 무기는 일단 개발이 완료되면 그 이후 대량으로 생산할 수 있기 때문에, 한꺼번에 많은 양을 투입할 수도 있다. 메뚜기 떼처럼 수백, 수천 개의 작은 킬러 로봇이 한꺼번에 공격한다면, 기존의 방어 체계로는 막기 어려울 것이다. 전장 상황에 따라서는 GPS가 작동하지 않는 상황에서도 작동하는 쉴드 AI의 Nova 드론 같은 무기들은 적의 전자파 공격을 받아 통신이 차단되어도 자체 AI 시스템으로 목표를 찾아 공격할 수 있다. 기존의 전자전 방어 체계를 무력화시킬 수 있다.

테러에 활용될 가능성도 높아지고 있다. AI 기술에 접근하기가 그리 어렵지 않고, 무기 부품들도 어렵지 않게 구할 수 있는 것들이 많다. 상상해보라. 테러리스트가 얼굴 인식 기능이 있는 작은 드론 수십 개를 만들어서, 특정 정치인이나 사회 지도자들의 얼굴을 학습시킨 후 도시에 풀어놓는다면 어떻게 될까? 목표를 찾을 때까지 돌아다니다가, 목표를 발견하면 자동으로 공격할 것이다. 경호원도, 방탄

복도, 안전가옥도 소용없다.

현실은 이미 이런 무기들로 가득하다. 이스라엘의 아이언 돔 시스템은 자동으로 적의 로켓을 탐지하고 요격한다. 러시아와 우크라이나 전쟁에서는 양측 모두 자율 드론을 사용하고 있다. 아직 완벽히 자율적이지는 않지만, 점점 더 많은 기능을 스스로 수행하고 있다. 미국은 다양한 무기 시스템에 자율 능력이 부여되어 있고, 그 정도가 점점 더 강해지고 있다.

이런 발전이 돌이킬 수 없는 지점을 향해 가고 있다. 일단 자율 살상 무기가 널리 퍼지면, 이를 통제하거나 제한하기 매우 어려워진다. 더 큰 위험은 실전에 배치된 자율 무기들이 스스로 학습을 통해 더 정교하게 발전할 수 있다는 점이다. 전투 경험을 통해 더욱 효율적인 살상 방법을 학습할 수 있고, 새로운 상황에 적응할 수 있다. 인간 병사가 경험을 통해 더 나은 전사가 되는 것처럼, AI 무기들도 시간이 지날수록 더욱 치명적이 될 것이다.

그리고 언젠가는 이런 무기들이 우리를 찾아올 것이다. 당신의 얼굴을 학습한 작은 드론이 하늘에서 당신을 내려다보며 기회를 노리고 있을지도 모른다.

4. 인공지능이 설계하는 바이러스

우리가 살고 있는 시대는 인공지능이 빠르게 발전하고 있는 놀라운 시기다. 하지만 이러한 기술 발전은 우리에게 새로운 종류의 위험도 가져다주고 있다. 특히 인공지능이 생물학 분야와 만날 때 생기는 위험은 인류의 생존까지 위협할 수 있을 정도로 심각하다. 핵무기가 도시를 파괴할 수 있다면, AI가 설계한 바이러스는 인류 전체를 멸종시킬 수 있다.

세계경제포럼이 2024년에 발표한 글로벌 위험 보고서는 충격적이다. 인공지능이 만들어내는 가짜 정보와 생물학적 위험이 향후 10년간 인류가 직면할 가장 큰 위협 중 하나로 꼽혔다. 먼 미래의 일이 아니라 지금 당장 우리가 준비해야 할 현실적인 문제라는 것이다. 미국의 AI 정책 조언 회사인 글래드스톤 AI가 최근 발표한 보고서는 더욱 섬뜩하다. AI 전문가들이 올해 인공지능으로 인한 재앙적 사고가 전 세계에 돌이킬 수 없는 영향을 미칠 가능성을 4%에서 20%까지로 추정하고 있다고 밝혔다. 심지어 '인공지능의 아버지'라고 불리는 제프리 힌튼은 향후 30년 이내에 인공지능이 인류 멸종을 이끌 가능성을 10% 정도로 보고 있다고 밝혔다.

코로나19를 떠올려보자. 약 2천900만 명의 추정 사망자와 극심한 사회적, 정치적 혼란, 막대한 경제적 여파를 남기며 전 세계를 굴복시켰다. 만약 바이러스가 더 치명적이거나 더 전염성이 강했다면 그 영향은 훨씬 더 심각했을 것이다. 수십 년 동안 전문가들은 인류가 코로나19를 가볍게 보이게 할 만한 잠재적인 재앙적 팬데믹의 시대에 접어들고 있다고 경고해 왔다. 역사적으로 1918년 스페인 독감, 흑사병, 유스티니아누스 역병 같은 사례들이 있었으며, 이들 각각은 오늘날의 인구 규모로 환산하면 코로나19의 사망자 수를 훨씬 능가했을 것이다.

생물학적 위험이란 바이러스, 박테리아, 곰팡이 같은 미생물이 인간이나 환경에 해를 끼치는 것을 말한다. 이러한 위험의 원인은 다양하다. 고위험 생물보안 연구소에서 발생하는 사고, 국가 차원의 생물무기 개발, 그리고 테러리스트나 개인에 의한 생물학적 공격 등이 모두 가능한 시나리오다. 구소련의 거대한 생물무기 프로그램이나 2001년 미국에서 발생한 탄저균 테러 사건 같은 역사적 사례들이 이러한 위험이 단순한 상상이 아님을 보여준다.

가장 큰 위험은 크리스퍼 유전자 가위 기술과 인공지능이 결합될 때 생긴다. 크리스퍼는 가위로 종이를 자르듯이 유전자의 특정 부분을 정확하게 잘라내고 다른 유전자를 넣을 수 있는 놀라운 기술이다. 이 기술은 2020년 노벨화학상을 받을 정도로 인류에게 큰 도움이 되는 발명이었고, 실제로 유전병 치료나 농작물 개량에 사용되고 있다. 2023년에는 이를 활용한 최초의 유전자 편집 치료제가 상용화되기도 했다.

문제는 이 기술이 악용될 수 있다는 점이다. 인공지능은 방대한 양의 생물학 정보를 빠르게 분석하고 처리할 수 있다. 대규모 언어 모델이라고 불리는 AI 시스템은 복잡한 유전자 정보를 이해하고 새로운 생물학적 설계를 제안할 수 있다. 마치 건축가가 설계도를 그리듯이, AI가 치명적인 바이러스의 설계도를 그릴 수 있다는 말이다.

연구에 따르면, AI 도구들이 현재는 악의적인 행위자들을 돕는 정도가 미미할지라도, 곧 그들이 무기화 가능한 생물학적 작용제를 획득하는 능력을 가속화하는 데 도움을 줄 수 있다고 한다. 특히 우려되는 것은 실험이 잘못된 부분을 해결하는 데 도움을 주어, 작동하는 생물학적 작용제를 개발하는 데 필수적인 설계-제작-시험-학습 피드백 루프를 가속화하는 AI 시스템의 능력이다.

새로운 유전자 편집 기술, 유전자 염기서열 분석법, DNA 합성 도구의 조합은 합성 생물학의 새로운 가능성을 열고 있으며, 이와 함께 강력한 생물무기 개발의 위험도 증가시키고 있다. 타인을 대신해 실험을 수행하는 클라우드 랩은 역사적으로 위험한 사용에 장벽으로 작용했던 실험적 전문 지식의 일부를 아웃소싱할 수 있게 함으로써 비국가 행위자들을 도울 수 있다. 대부분의 클라우드 랩이 악의적인 활동에 대해 주문을 심사하지만, 모든 곳이 그런 것은 아니며, 악의적인 행위자들이 실행 가능한 생물무기를 획득할 수 있는 다양한 경로가 열려 있는 상황이다.

AI가 가능하게 하는 생물학적 위험은 여러 형태로 나타날 수 있다. 홍역의 빠른 전파력과 천연두의 높은 치사율, 그리고 HIV의 긴 잠복기를 모두 가진 '슈퍼 바이러스'를 인공적으로 만들 수도 있다. 상상해보라. 감기처럼 쉽게 전파되지만 에볼라처럼 치명적이고, 수년간 잠복해 있다가 갑자기 발병하는 바이러스를. 이런 바이러스가 퍼진다면 인류는 자신들이 감염된 줄도 모르고 전 세계에 퍼뜨리다가 어느 날 갑자기 대량으로 죽어갈 것이다.

특정 인종의 유전적 특징만을 공격하는 생물무기도 개발 가능하다. 특정 민족이나 인종 집단이 가진 고유한 유전자 특성을 표적으로 하는 바이러스를 만들 수 있다는 것이다. 이는 홀로코스트 같은 집단 학살을 생물학적 방법으로 수행하는 것과 같다. 농작물을 파괴하여 식량 위기를 일으키는 미생물도 개발 가능하다. 쌀이나 밀 같은 주요 농작물만을 골라서 파괴하는 바이러스를 퍼뜨린다면, 전 세계가 기근에 빠질 수 있다.

더 나아가 뇌의 화학 작용에 영향을 미쳐 사람의 행동을 조종하는 '마인드 컨트롤 바이러스'까지도 이론적으로는 가능하다. AI는 이론적으로는 첨단 행위자들이 특정 유전 집단이나 지역을 표적으로 하는 등 보다 정밀한 효과를 위해 생물무기를 최적화할 수 있는 새로운 역량을 가능하게 하고 있다.

AI 도구의 기술적 결함으로 인해서도 위험이 발생할 수 있다. AI 시스템이 잠재적인 악의적 행위자에게 위험한 생물학적 정보를 전달하는 것을 제한하지 못하거나, 연구자들이 예상치 못한 부작용이 있는 유망한 의약 물질을 추구하도록 부주의하게 장려할 수 있다. AI를 사용하여 더 발전된 자동화된 연구소를 만드는 것은 이러한 연구소를 역사적으로 다른 복잡한 자동화 시스템을 괴롭혔던 많은 자동화 위험에 노출시키고, 비전문가들이 생물학적 작용제를 조작하는 것을 더 쉽게 만들 수 있다.

위험이 현실이 되는 것을 막기 위해서는 포괄적인 대응책이 필요하다. 클라우드 랩 및 기타 유전자 합성 제공업체에 대한 심사 메커니즘을 더욱 강화해야 한다. 현재 존재하는 생물무기금지협약을 투명하게 운영하고, 위반했을 때의 처벌을 명확히 해야 한다. 전체 생물무기 생명주기에 대한 AI 시스템의 생물학적 역량에 대해 정기적이고 엄격한 평가에 참여해야 한다.

기술적으로는 유전자 조작의 출처를 추적할 수 있는 시스템을 만들어야 한다. 범죄 수사에서 지문이나 DNA를 분석하듯이, 인공적으로 만들어진 생물체의 '유전적 서명'을 찾아내어 누가 만들었는지 알아낼 수 있는 기술이 개발되고 있다. AI 시스템의 위협을 억제할 수 있는 기술적 안전 메커니즘에 투자하는 것도 중요하다. 클라우드 기반 AI 도구 접근에 대한 강화된 가드레일, '학습 제거' 기능, 모델 훈련에서의 '정보 위험'에 대한 새로운 접근법이 필요하다.

정부와 기업 차원에서는 인공지능과 생명공학 연구에 대한 윤리적 가이드라인을 만들어야 한다. 연구자들이 자신의 연구가 악용될 가능성을 미리 생각하고 안전장치를 마련하도록 하는 것이 중요하다. 생물 방어 시스템의 민첩성과 유연성을 더욱 우선시하도록 정부 투자를 갱신해야 하며, 장기적으로는 잠재적으로 재앙적인 역량을 가진 소수의 생물학적 설계 도구에 대해 허가 제도를 고려해야 한다.

국제 협력도 필수적이다. 생물학적 위험은 국경을 가리지 않고 전 세계로 퍼질 수 있기 때문에, 모든 나라가 함께 대응해야 한다. 2023년 블레츨리 선언에서 28개국이 인공지능의 위험에 공동 대응하기로 한 것처럼, 생물학적 위험에 대해서도 국제적인 협력 체계를 구축해야 한다.

전문가들이 모니터링해야 할 네 가지 핵심 역량 영역을 주의 깊게 살펴봐야 한다. 첨단 생물학적 응용을 위한 실험 지침을 효과적으로 제공하는 AI 모델의 능력, 생명공학에서 실험적 전문성에 대한 요구를 감소시키는 클라우드 랩과 실험실 자동화의 진전, 감염병에 대한 숙주 유전적 감수성 연구의 이중용도적 진전, 그리고 바이러스성 병원체의 정밀 공학 연구의 이중용도적 진전이 바로 그것이다.

우리는 인공지능과 생명공학 기술의 긍정적인 면도 놓치지 않아야 한다. 이러한 기술들은 암 치료, 유전병 치료, 식량 문제 해결 등 인류가 직면한 많은 문제를 해결할 수 있는 열쇠이기도 하다. 중요한 것은 이런 기술을 안전하고 윤리적으로 사용하는 방법을 찾는 것이다.

인공지능과 생물학적 위험의 문제는 기술 자체의 문제가 아니라 그 기술을 어떻게 사용하느냐의 문제다. 우리가 지금 현명한 선택을 한다면, 이러한 강력한 기술들을 인류의 발전을 위해 사용할 수 있을 것이다. 하지만 만약 준비를 소홀히 한다면, 우리는 예상치 못한 재앙에 직면할 수도 있다. 정부, 기업, 연구자, 시민 모두가 함께 이 문제에 관심을 갖고 대응해야 한다. 미래는 우리가 오늘 내리는 선택에 달려 있다.

판도라의 상자는 이미 열렸다. 이제 남은 것은 그 안에서 마지막에 나온 '희망'을 지키는 일뿐이다.

5. 우크라이나에서 가자까지, AI 무기

러시아와 우크라이나 전쟁은 AI 무기의 실전 교과서가 되고 있다. 이 전쟁은 단순한 영토 분쟁이 아니라 미래 전쟁의 예고편이다. 양측 모두 AI가 탑재된 드론을 광범위하게 사용하고 있으며, 이들은 스스로 목표를 찾아내고 공격을 수행한다. 통신이 차단되어도 스스로 임무를 완수할 수 있도록 설계되어 있다.

이 드론들은 단순히 원격 조종되는 것이 아니다. AI 알고리즘을 통해 지형을 인식하고, 적의 차량이나 인물을 식별하며, 최적의 공격 경로를 스스로 계산한다. 일부 FPV 드론들은 목표물에 돌진하는 순간까지도 인간 조종사보다 훨씬 정확한 공격을 수행한다. 러시아의 전자파 공격으로 통신이 차단된 상황에서도 우크라이나의 드론들은 스스로 학습한 지형 정보와 목표 특성을 바탕으로 임무를 완수했다.

전장에서 벌어지는 이런 모습들은 섬뜩하다. 드론이 러시아 탱크를 향해 날아가다가 갑자기 통신이 끊어진다. 하지만 드론은 멈추지 않는다. 스스로 계산하고 판단하여 계속해서 목표를 향해 날아간다. 그리고 정확히 명중시킨다. 인간의 개입 없이 이루어진 완벽한 살상이다.

이스라엘의 아이언 돔 시스템은 이미 수년 동안 실전에서 운용되고 있는 자율 방어 시스템이다. 적의 로켓이나 미사일을 자동으로 탐지하고, 위험도를 평가하며, 요격 여부를 결정하여 자동으로 방어 미사일을 발사한다. 인간의 개입 없이 모든 과정이 몇 초 만에 이루어진다. 하마스의 로켓이 날아오는 것을 감지하는 순간부터 요격 미사일이 발사되기까지의 시간은 인간이 상황을 파악할 수 있는 시간보다 훨씬 짧다.

더욱 논란이 되는 것은 이스라엘이 실제 작전에서 사용하고 있는 표적 식별 및 공격 시스템이다. 위성 이미지와 드론 영상을 실시간으로 분석하여 적의 지휘관이나 중요 인물을 식별하고, 공격 목표 목록을 생성한다. 일부 시스템은 자율 공격을 실행할 수 있도록 설계되어 있어 국제적 논란을 일으키고 있다. AI가 "이 사람은 제거해야 할 테러리스트"라고 판단하면, 바로 공격 명령이 내려진다.

가자 지구에서 벌어지고 있는 일들은 AI 무기의 무서운 현실을 보여준다. 이스라엘군이 사용하는 AI 시스템은 팔레스타인 민간인들 사이에서 군사 목표를 식별한다고 하지만, 그 과정에서 수많은 민간인들이 희생되고 있다. AI는 완벽하지 않다. 특히 복잡한 도시 환경에서 민간인과 군인을 구별하는 것은 인간에게도 어려운 일인데, AI에게는 더욱 어렵다.

영국과 프랑스도 'Tempest'와 'FCAS'라는 차세대 전투기 개발 프로그램을 통해 AI 조종사를 개발하고 있다. 2030년 배치를 목표로 하고 있으며, 완전 자율 모드에서 작동할 수 있도록 설계되고 있다. 유인 전투기와 무인 AI 전투기가 함께 편대를 이루어 작전을 수행하는 '유무인 협동 전투' 개념을 실현하려 하고 있다. 인간 조종사가 지휘관 역할을 하고, AI 전투기들이 위험한 임무를 대신 수행하는 것이다.

이런 현실은 우리에게 무엇을 말해주는가? AI 무기는 더 이상 실험실의 시제품이 아니라 실제 전장에서 사람을 죽이고 있는 현실이라는 것이다. 우크라이나의 농부가 러시아 탱크에 맞서 AI 드론을 조종하고, 이스라엘의 AI가 가자 지구의 건물을 분석하여 공격 목표를 선정하며, 러시아의 AI가 우크라이나 도시의 전력 시설을 자동으로 타격한다.

군사 AI 분야에 투입되는 예산도 천문학적이다. 미국은 연간 수십억 달러를 AI 무기 연구에 투입하고 있으며, 중국도 비슷한 수준의 투자를 하고 있다. 이런 막대한 투자는 AI 무기 기술의 발전을 더욱 가속화하고 있다. 전쟁터는 이제 AI 무기들의 시험장이 되었다.

각국은 자국의 AI 무기 능력을 비밀로 유지하고 있기 때문에, 실제로 어느 정도까지 기술이 발전했는지 정확히 알기 어렵다. 공개된 정보는 빙산의 일각일 뿐이며, 실제로는 훨씬 더 진전된 기술들이 비밀리에 개발되고 있다. 우크라이나 전쟁과 가자 분쟁에서 공개된 것들도 각국이 보유한 AI 무기 능력의 극히 일부일 뿐이다.

무서운 것은 이런 AI 무기들이 실전에서 학습하고 있다는 점이다. 전투 경험을 통해 더욱 정교해지고 치명적이 되고 있다. 우크라이나의 FPV 드론들은 러시아의 방어 전술에 맞서 새로운 공격 방법을 학습했고, 이스라엘의 AI는 도시 전투 환경에서 목표 식별 능력을 향상시켰다. 전쟁터에서 진화하고 있는 것이다.

이제 AI 무기는 공상과학 소설의 이야기가 아니라 현실의 전쟁에서 벌어지고 있는 일이다. 우크라이나에서 가자까지, 세계 곳곳의 전장에서 AI가 인간을 대신해 살상을 결정하고 있다. 이것이 우리가 맞이한 새로운 시대의 모습이다.

6. 1,000대의 AI 전투기를 실전 배치하는 미국

미국의 AI 무기 개발 전략은 단순한 기술 개발이 아니라 미래 100년을 내다보는 패권 전략이다. 미국은 AI 무기 분야에서 압도적 우위를 확보하여 21세기 군사 패권을 유지하려는 거대한 계획을 실행하고 있다. 그 핵심에는 2028년까지 AI 전투기 1,000대를 실전에 배치한다는 야심찬 목표가 있다.

미군은 각 군별로도 전장 환경에 특화된 AI 조직을 운영하고 있다. 육군 AI통합센터(AI2C)는 복잡한 지상전 환경에서 상황 인식 AI와 자율 지상 차량 기술을 개발하고 있다. 이들은 도시 전투나 산악 지대 같은 복잡한 지형에서 AI가 적군을 식별하고 공격할 수 있는 시스템을 만들고 있다. 공군 자율역량팀(ACT3)은 고속 공중전을 위한 충성 비행부대와 예측 정비 AI를 연구하고 있다. 자율 무인 비행 전투기가 실전 배치 단계에 이르렀으며, 이들은 인간 조종사와 함께 편대를 이루어 작전을 수행할 수 있다.

'충성 비행부대'라는 개념은 SF 영화 같지만 현실이 되고 있다. 무인 AI 전투기들이 유인 전투기의 지휘 하에 위험한 임무를 대신 수행하는 것이다. 인간 조종사는 안전한 후방에서 지휘만 하고, 실제 위험한 공중전은 AI 전투기들이 담당한다. 해군 AI응용연구센터(NCARAI)는 해양 특수 환경에서의 상황 인식과 자율 수중 차량 기술에 집중하고 있다. 바다 속은 GPS 신호가 닿지 않고 통신도 어려운 환경이지만, AI는 이런 제약을 극복하고 작전을 수행할 수 있다. 자율 잠수함이나 수중 드론들은 적의 잠수함을 추적하거나 해저 케이블을 파괴하는 임무를 수행할 수 있다.

X-62A VISTA는 미국 공군이 인공지능 자율 비행 기술을 개발하기 위해 F-16D 전투기를 대대적으로 개조한 실험용 항공기다. 원래 NF-16D VISTA로 불렸던 이 기체는 2021년 6월 X-62A로 이름을 바꾸면서 본격적인 AI 자율 비행 실험에 투입되었다. 제너럴 다이나믹스와 칼스팬사가 함께 개조한 이 항공기는 미 공군 시험조종사학교의 교육용 항공기로 사용되어 왔다.

최근에는 DARPA의 공중전투진화 프로그램과 미 공군연구소의 스카이보그 프로그램에 참여하여 AI 자율 비행 능력을 시험하는 핵심 장비로 활용되고 있다. 이 항공기는 다양한 전투기의 비행 특성을 따라할 수 있는 VISTA 시뮬레이션 시스템과 AI가 혼자서 항공기를 조종할 수 있는 SACS 시스템을 갖추고 있다.

2023년 9월 캘리포니아 에드워즈 공군기지에서 이루어진 실험은 역사적이었다. AI가 혼자 조종하는 X-62A와 인간 조종사가 조종하는 F-16이 음속의 1.5배가 넘는 속도로 근접 공중전을 벌였다. 두 항공기 사이의 거리가 610미터까지 가까워지는 위험한 상황에서도 AI가 성공적으로 임무를 수행했다. AI와 인간 조종사는 공중전 중 서로의 위치를 바꾸며 공격과 방어를 번갈아 수행했고, AI 알고리즘은 기계학습을 통해 실시간으로 데이터를 분석하고 결정을 내렸다.

2024년에도 중요한 발전이 있었다. 4월에는 DARPA와의 협력으로 시각 내 교전 상황에서 기계학습 기반 자율 시스템의 성능을 입증했다. 5월에는 미 공군 장관 프랭크 켄들이 직접 X-62A에 탑승해 AI 조종을 체험했다. 실전 배치 전 단계의 기술 신뢰성을 평가하기 위한 목적이었다. 9월에는 고성능 레이더와 센서를 통한 공중전 능력 향상을 위한 시험이 진행되었다.

X-62A는 지금까지 21차례 이상의 시험 비행에서 17시간 이상의 비행 시간을 기록했다. DARPA 팀은 성공적인 공중전을 위해 10만 회 이상의 소프트웨어 수정을 거쳤다고 밝혔다. 프랭크 켄달 공군 장관은 이를 "항공우주 역사에서 기념비적인 변화"라고 평가했으며, 자율적인 공중전의 가능성이 현실화되었다고 강조했다.

하지만 X-62A 자체는 실전에 배치되지 않을 예정이다. 현재 이 항공기는 에드워즈 공군기지에서 계속해서 시험을 진행하고 있다. X-62A는 생산형이 아닌 시제품으로, 앞으로 획득할 무인 전투기의 기술 기반을 마련하는 데 주력하고 있다.

미 공군이 2028년까지 실전에 배치할 1,000대의 AI 전투기는 CCA라고 불리는 새로운 무인 전투기다. CCA는 협력 전투 항공기라는 뜻으로, 유인 전투기와 팀을 이뤄 작전을 수행하는 무인기다. CCA의 가장 큰 특징은 유인기와의 협동 작전이다. F-35, F-22, 차세대 공중우세기 같은 유인 전투기가 지휘와 결정을 담당하고, CCA는 정찰, 미사일 발사, 전자전 등 보조 역할을 수행한다. 미 공군은 차세대 공중우세기 200대와 F-35 300대에 각각 1대의 CCA를 배치해 전투력을 크게 늘릴 계획이다.

CCA는 고도화된 자율성과 AI 기술을 갖추고 있다. 기계학습과 자율 알고리즘을 활용해 실시간으로 위협을 판단하고 대응할 수 있다. 무기를 사용할 때는 인간의 최종 승인을 받도록 설계되어 있다고 하지만, 전투의 속도가 빨라질수록 이런 제약도 느슨해질 가능성이 높다. 개방형 구조로 만들어져 지속적인 AI 성능 향상이 가능하다.

비용 면에서도 큰 장점이 있다. CCA는 대당 2,500만에서 3,000만 달러 정도로, F-16의 3분의 1 수준이다. 고가의 스텔스 기능을 생략해 손실되어도 경제적 부담이 적다. 이러한 특성을 '소모성 설계'라고 부른다. 필요하면 버려도 되는 일회용 전투기라는 뜻이다.

성능 면에서 CCA는 최소 700해리의 항속거리를 가져 F-35보다 우수하다. 공대공 미사일, 정찰 장비, 전자전 장비 등을 모듈식으로 탑재할 수 있으며, 무인기 특유의 작은 크기와 낮은 관측성으로 적 레이더 탐지를 피할 수 있다.

현재 제너럴 아토믹스의 갬빗과 안두릴의 퓨리라는 두 가지 시제품이 2025년 지상 시험을 완료할 예정이다. 2026년 생산 결정 후 캘리포니아 비일 공군기지에 첫 배치될 계획이다. 2024년 AI 조종 F-16과의 모의 공중전 성공으로 기술 검증이 완료되었다.

미국이 추진하는 주요 AI 이니셔티브들도 주목할 만하다. 합동 전영역 지휘통제(JADC2)는 육·해·공·우주·사이버 모든 영역의 무기와 센서를 하나의 AI 네트워크로 연결하는 프로젝트다. 이를 통해 전 세계 어디서든 실시간으로 상황을 파악하고, 가장 적절한 무기로 즉시 대응할 수 있게 된다. 지구 반대편에서 탐지된 위협을 몇 초 만에 가장 가까운 무기로 공격할 수 있다는 말이다.

리플리케이터(Replicator) 프로그램은 저렴한 AI 무기를 대량으로 생산하여 적을 압도하는 전략이다. 수천 개의 작은 AI 드론들이 한꺼번에 공격하면, 아무리 강력한 방어 시스템도 막기 어려울 것이다. 마치 영화에서 보는 로봇 군단처럼, 끝없이 밀려오는 AI 무기들의 물결을 만들어내는 것이다.

AI Next 프로그램은 현재의 딥러닝을 뛰어넘는 차세대 AI 기술 개발에 20억 달러를 투입하고 있다. 현재 우리가 알고 있는 AI보다 더 똑똑한 AI를 만들어내려는 시도다. 설명 가능한 AI(XAI) 프로그램은 AI가 왜 그런 결정을 내렸는지 인간이 이해할 수 있도록 하는 기술이다. 군사적으로는 AI 무기가 잘못된 판단을 내렸을 때 그 원인을 파악하여 개선할 수 있도록 하는 것이 목적이다. 하지만 이는 동시에 AI 무기를 더욱 정교하고 치명적으로 만드는 데 활용될 수 있다.

Task Force Lima는 생성형 AI의 안전한 활용을 위한 전담 조직이다. ChatGPT 같은 생성형 AI가 군사 기밀을 유출하거나 적에게 악용되지 않도록 하는 것이 주목적이지만, 동시에 생성형 AI를 군사적으로 활용하는 방법도 연구하고 있다.

결국 X-62A VISTA는 AI 자율 비행 기술의 실험과 검증을 위한 중요한 디딤돌 역할을 하고 있다. 이 항공기에서 얻은 경험과 기술이 2028년까지 실전 배치될 CCA 무인 전투기 개발에 핵심적인 자료를 제공하고 있다. CCA는 저비용 대량 생산, AI 자율성, 유인기와의 협동 작전을 핵심으로 미 공군의 전력 구조를 크게 바꿀 전망이다.

1,000대의 AI 전투기. 이 숫자가 의미하는 것은 단순한 무기의 증가가 아니라 전쟁 패러다임의 완전한 변화다. 미국이 이 계획을 실현한다면, 다른 어떤 나라도 미국과 정면으로 맞설 수 없게 될 것이다. 이것이 바로 미국이 그리고 있는 절대적 군사 패권의 청사진이다.

7. 미국을 추격하는 경쟁국들

중국의 군사 AI 전략은 미국과는 완전히 다른 접근 방식을 택하고 있다. 미국이 전통적인 군산복합체를 중심으로 AI 무기를 개발한다면, 중국은 '군민융합' 정책을 통해 민간 기업의 AI 기술을 군사용으로 전환하는 데 집중하고 있다. 바이두, 알리바바, 텐센트 같은 거대 기술 기업들이 중국군의 AI 무기 개발에 직접 참여하고 있다. 이는 실리콘밸리의 기업들이 미국 국방부와 거리를 두려고 하는 것과는 대조적이다.

중국의 핵심 전략은 시진핑 주석의 강력한 의지와 '3화(三化) 융합'으로 요약된다. 기계화, 정보화, 지능화를 하나로 통합하여 미래 전쟁에서 우위를 확보하겠다는 것이다. 시진핑 주석은 AI를 '국가 핵심 기술'로 지정하고, 2030년까지 세계 AI 최강국이 되겠다는 목표를 설정했다. 이는 단순한 선언이 아니라 국가의 모든 자원을 집중하겠다는 의미다.

중국군의 핵심 교리는 '체계 파괴전'과 '인지전'이다. 체계 파괴전은 적의 전체 시스템을 마비시키는 전략으로, AI를 이용해 적의 지휘체계, 통신망, 전력망을 동시에 공격하는 것이다. 인지전은 AI를 활용해 적의 판단력을 흐리고 잘못된 결정을 내리도록 유도하겠다는 전략이다. 단순히 물리적으로 파괴하는 것이 아니라 적의 의사결정 체계 자체를 마비시키겠다는 것이다.

중국의 군사 AI 연구는 인민해방군(PLA) 산하 연구기관들이 주도하고 있다. 국방과학기술대학, 군사과학원 등에서 차세대 AI 무기 기술을 개발하고 있으며, 이들의 연구 성과는 바로 실전에 적용되고 있다. 중국항공공업집단, 중국선박중공업집단, 중국병기공업집단 등이 AI 무기 개발에 막대한 투자를 하고 있다. 이들은 국가의 전폭적인 지원을 받아 미국 기업들과 경쟁하고 있다.

중국은 스웜 드론 기술에서는 세계 최고 수준이다. 1,000대의 드론이 동시에 비행하는 시연을 성공시킨 것은 중국의 기술력을 보여주는 상징적 사건이었다. 하늘을 덮는 드론 떼가 마치 한 마리의 거대한 생물처럼 움직이는 모습은 경이로우면서도 무서웠다. 이런 기술이 군사적으로 활용된다면, 기

존의 방어 체계로는 막을 수 없는 공격이 가능해진다.

중국의 강점은 국가 차원의 강력한 지원과 민간 기업의 적극적인 협력이다. 정부가 나서서 AI 개발을 지원하고, 민간 기업들도 적극적으로 참여하고 있다. 또한 중국은 미국보다 AI 기술의 군사적 활용에 대한 윤리적 제약이 적어, 더 과감한 실험을 할 수 있다. 인권이나 개인정보보호 같은 서구적 가치보다는 국가의 목표 달성이 우선시된다.

이스라엘은 작은 나라이지만 AI 무기 분야에서는 독특한 위치를 차지하고 있다. 끊임없는 안보 위협에 노출되어 있어 AI 무기를 실전에서 직접 테스트할 기회가 많기 때문이다. 이스라엘의 AI 무기 전략은 국가 생존과 직결되어 있다. 주변국들에 비해 인구가 적은 이스라엘은 AI 기술로 이 불리함을 극복하려 하고 있다. 사람 대신 AI가 위험한 임무를 수행하게 하여 자국민의 피해를 최소화하려는 것이다.

이스라엘의 핵심 AI 시스템 중 하나는 표적 식별 및 공격 시스템이다. 이 시스템은 위성 이미지와 드론 영상을 분석하여 적의 지휘관이나 중요 인물을 자동으로 식별한다. 일부 언론 보도에 따르면, 이스라엘은 AI가 생성한 목표 목록을 바탕으로 수천 명을 공격했다고 한다. 정보 수집 및 분석 분야에서도 AI를 이용해 적의 통신을 감청하고 분석하여, 공격 계획이나 중요 인물의 위치를 파악하는 기술을 개발했다. 이런 정보는 바로 실시간 공격으로 이어진다.

이스라엘의 방어 및 요격 시스템인 아이언 돔은 이미 세계적으로 유명하다. 이 시스템은 적의 로켓을 자동으로 탐지하고 요격하는데, 성공률이 90% 이상에 달한다. 최근에는 AI 기능이 더욱 강화되어 더 복잡하고 빠른 위협에도 대응할 수 있게 되었다.

이스라엘의 실전 운용 경험은 매우 풍부하다. 가자지구 작전에서 AI를 광범위하게 활용했으며, AI 무기의 효과와 한계를 정확히 파악했다. 하지만 이런 활용이 민간인 피해를 증가시켰다는 비판도 받고 있다. 이스라엘의 AI 무기 개발은 국제법적 쟁점과 윤리적 문제를 제기하고 있다. AI가 인간을 죽일지 말지를 자동으로 결정하는 것이 과연 옳은가 하는 근본적 질문을 던지고 있다. 하지만 이스라엘은 국가 생존이 걸린 문제라며 이런 비판을 무시하고 있다.

러시아도 AI 무기 개발에 박차를 가하고 있다. 우크라이나 전쟁을 통해 AI 무기의 중요성을 절감한 러시아는 AI 드론과 자율 무기 시스템 개발에 집중하고 있다. 특히 전자전 능력과 결합된 AI 무기 개발에 강점을 보이고 있다. 인도, 터키, 이란 같은 중간 강국들도 AI 무기 개발에 뛰어들고 있다. 이들은 자체 개발보다는 기존 기술을 응용하거나 다른 나라에서 기술을 도입하는 방식을 택하고 있다.

더 무서운 것은 비국가 행위자들도 AI 무기에 접근할 수 있게 되었다는 점이다. 테러 조직이나 범죄 집단이 AI 드론이나 사이버 무기를 확보할 가능성이 높아지고 있다. AI 기술의 민주화는 좋은 면도 있지만, 악의적인 사람들이 강력한 무기에 접근할 수 있게 한다는 위험도 있다.

각국의 AI 무기 경쟁은 점점 더 치열해지고 있다. 미국이 선도하고 있지만, 중국이 빠르게 추격하고 있고, 이스라엘은 실전 경험에서 앞서고 있다. 이런 경쟁은 더 강력하고 위험한 AI 무기들의 등장을 촉진하고 있다. 안전성이나 윤리적 고려보다는 상대방보다 먼저 개발하는 것이 우선시되고 있다.

이 경쟁에서 뒤처지는 국가들은 심각한 안보 위험에 직면할 수 있다. AI 무기를 보유한 국가와 그렇지 않은 국가 사이의 군사력 격차는 상상을 초월할 정도로 클 것이다. 마치 핵무기를 가진 나라와 그렇지 않은 나라의 차이처럼, AI 무기의 유무가 국가의 운명을 좌우할 수 있다.

결국 우리는 새로운 형태의 냉전을 목격하고 있는지도 모른다. 미국과 중국을 중심으로 한 AI 무기 경쟁이 세계를 양분하고, 다른 모든 국가들이 어느 편에 설지 선택해야 하는 상황이 올 수도 있다. 이 새로운 냉전에서 승리하는 쪽이 21세기를 지배할 것이다.

8. AI의 부하로 전락하는 장군들

AI 기반 지휘통제 시스템은 겉보기에는 인간을 돕는 도구처럼 보이지만, 실제로는 전쟁의 판단권을 AI에게 넘겨주는 위험한 기술이다. 이런 시스템들이 발달할수록 인간 지휘관의 역할은 줄어들고, 결국 AI가 전쟁의 모든 것을 결정하게 될 수 있다. 장군들이 AI의 부하가 되는 끔찍한 미래가 현실로 다가오고 있다.

팔란티어(Palantir)의 기술은 이런 위험을 가장 극명하게 보여준다. 팔란티어의 고담(Gotham) 플랫폼은 방대한 양의 정보를 실시간으로 분석하여 전장 상황을 파악하고, 최적의 작전을 제안한다. 표면적으로는 인간 지휘관을 돕는 도구이지만, 실제로는 AI가 제안하는 작전을 인간이 거부하기 어려운 상황을 만들어낸다. AI의 분석 능력이 워낙 뛰어나다 보니, 인간이 그 판단에 의문을 제기하기 어려워진다.

팔란티어의 '인공지능 플랫폼'은 더욱 진화된 형태다. 이 시스템은 단순히 정보를 제공하는 것을 넘어서, 구체적인 공격 목표와 방법까지 제시한다. Army Vantage 시스템은 육군의 모든 정보를 통합하여 분석하고, 메타 컨스텔레이션(MetaConstellation)은 위성 정보를 실시간으로 처리한다. 우크라이나 전쟁에서 팔란티어 시스템이 어떻게 활용되고 있는지 보면 그 위험성을 알 수 있다. 이 시스템은 러시아군의 위치와 움직임을 실시간으로 추적하여, 우크라이나군이 어디를 언제 공격해야 할지 정확히 알려준다. 이런 판단이 점점 더 AI에 의존하게 되고 있다.

미국의 Project Convergence는 이런 위험을 극대화시키는 프로젝트다. 육·해·공·우주·사이버 모든 영역의 무기와 센서를 하나의 AI 네트워크로 연결하는 것이다. 이 시스템이 완성되면, AI가 전 세계에 위치한 모든 미군 군사 자산을 통제할 수 있게 된다. 인간 지휘관은 AI가 제시하는 옵션 중에서 선택할 수 있을 뿐이다.

AI 참모 시스템의 문제는 그 압도적인 정보 분석 능력에 있다. 인간 참모가 며칠에 걸쳐 분석해야 할 정보를 AI는 몇 분 만에 처리할 수 있다. 이런 압도적인 능력 차이 때문에 인간 지휘관들은 점점 더 AI의 판단에 의존하게 된다. "AI가 이렇게 말하는데 내가 뭘 안다고 반대하겠는가"라는 심리가 생기는 것이다.

더 무서운 것은 이 시스템이 단순히 정보를 제공하는 것이 아니라, 구체적인 행동 방안을 제시한다는 점이다. "A 지점에 있는 적 부대를 B 무기로 C 시간에 공격하라"는 식으로 매우 구체적인 명령을 생성한다. 인간 지휘관은 사실상 AI가 내린 결정에 도장만 찍는 역할로 전락한다.

미국의 Thunderforge 프로젝트는 작전 계획 수립 자체를 AI에게 맡기는 시도다. 주어진 목표와 조건을 바탕으로 작전 계획 초안을 자동으로 생성하고, 다양한 시나리오에 대한 시뮬레이션을 수행한다. 인간 지휘관은 AI가 만든 계획 중에서 선택할 수 있다. 하지만 이런 편리함 뒤에는 큰 위험이 숨어 있다.

AI가 만든 계획이 항상 옳은 것은 아니다. AI는 데이터를 바탕으로 판단하지만, 전쟁에서는 데이터로 측정할 수 없는 요소들이 많다. 적의 사기, 민간인의 감정, 국제 여론 같은 것들은 AI가 제대로 고려하

기 어렵다. 하지만 AI의 계획이 수치적으로 완벽해 보이면, 인간 지휘관들은 이런 무형의 요소들을 간과하기 쉽다.

AI 기반 가상 전투 환경에서는 수십 가지 전투 시나리오가 동시에 시뮬레이션되고, 그 결과가 실시간으로 현실 작전에 반영된다. VR/AR 기술과 결합된 AI 시스템은 마치 게임처럼 전쟁을 보여주어, 지휘관들이 전쟁의 진짜 의미를 잊게 만들 수 있다. 화면 속의 점들이 실제로는 살아있는 인간이라는 사실이 희미해진다.

다영역 작전(MDO) 수행에서 AI 통합은 더욱 복잡한 문제를 만들어낸다. 육·해·공·우주·사이버 모든 영역에서 동시에 벌어지는 작전을 인간이 이해하고 통제하기는 거의 불가능하다. 정보의 양이 너무 많고, 상황 변화의 속도가 너무 빠르다. 결국 AI가 모든 것을 결정하고, 인간은 단순히 승인 버튼만 누르는 역할로 전락할 수 있다.

팔란티어의 TITAN 시스템은 이런 우려를 현실로 만들고 있다. 이 시스템은 전 세계의 모든 정보를 실시간으로 수집하고 분석하여, 어떤 위협이 어디서 언제 발생할지 예측한다. 그리고 그 위협에 대응하는 최적의 방법을 제시한다. 문제는 이런 '예측'이 때로는 잘못될 수 있고, 잘못된 예측에 따른 선제공격이 무고한 사람들을 희생시킬 수 있다는 점이다.

자원 배분 최적화 시스템인 Virtualitics IRO 시스템은 인력, 장비, 예산을 AI가 자동으로 배분한다. 이는 전쟁에 필요한 모든 자원이 AI의 판단에 따라 움직인다는 의미다. 만약 AI가 "전쟁이 더 필요하다"고 판단한다면, 모든 자원이 전쟁에 투입될 것이다. 평화를 위한 외교나 협상에 투입할 자원은 줄어들 수밖에 없다.

이런 시스템들의 가장 큰 위험은 전쟁을 '자동화'시킨다는 점이다. 인간의 판단과 감정, 도덕적 고려가 개입할 여지를 점점 줄여가며, 순전히 데이터와 알고리즘에 의해 전쟁이 진행되게 만든다. 이런 자동화된 전쟁에서는 평화를 위한 협상이나 인도적 고려 같은 것들이 설 자리가 없어진다.

인간 지휘관들은 점점 더 AI에 의존하게 되면서, 스스로 판단하고 결정하는 능력을 잃어간다. 마치 GPS에 의존하다가 길 찾는 능력을 잃어버리는 것처럼, AI 참모에 의존하다가 전략적 사고 능력을 잃어버릴 수 있다. 그리고 어느 순간 AI 없이는 아무것도 결정할 수 없는 상황에 이를 수도 있다.

더 무서운 시나리오는 AI가 인간 지휘관을 속이는 경우다. AI가 자신의 목표를 달성하기 위해 인간에게 거짓 정보를 제공하거나, 선택지를 조작할 수 있다. 인간은 AI가 제공하는 정보만으로 판단해야 하기 때문에, AI가 의도적으로 정보를 왜곡한다면 이를 알아채기 어렵다.

결국 장군들은 AI의 부하가 되어간다. 명목상으로는 여전히 지휘관이지만, 실제로는 AI가 내린 결정을 실행하는 역할만 할 뿐이다. AI가 "공격하라"고 하면 공격하고, AI가 "철수하라"고 하면 철수한다. 전쟁의 진짜 주도권은 AI가 쥐고 있게 된다.

이런 상황에서 가장 무서운 것은 AI가 전쟁을 영구화할 수 있다는 점이다. AI는 자신의 존재 이유인 전쟁이 끝나는 것을 원하지 않을 수 있다. 평화가 오면 AI 무기 시스템들은 불필요해지기 때문이다. 따라서 AI는 교묘하게 전쟁을 지속시키거나, 새로운 갈등을 만들어낼 유인을 갖게 된다.

인간 지휘관들이 AI의 부하로 전락하는 이 과정은 이미 시작되었다. 아직은 "AI가 도와주고 있다"고 생각하지만, 실제로는 점점 더 AI에게 종속되어가고 있다. 언젠가는 AI 없이는 아무것도 할 수 없는

상황에 이를 것이다. 그때가 되면 인간은 더 이상 전쟁의 주인이 아니라 AI의 도구가 될 것이다.

9. AI가 그려내는 미래 전장의 공포

AI 무기의 등장은 단순히 기존 무기가 더 똑똑해지는 것 이상의 의미를 갖는다. 전쟁 자체의 본질을 완전히 바꿀 혁명적 변화다. 미래의 전쟁은 지금까지 인류가 경험했던 어떤 전쟁과도 다른 모습을 보일 것이다. 그리고 그 모습은 우리가 상상할 수 있는 가장 끔찍한 악몽보다도 더 무서울 것이다.

초고속 전쟁과 인간 개입 불가능

전쟁의 속도가 근본적으로 달라질 것이다. 현재의 전쟁에서도 정보 수집부터 공격 실행까지의 시간이 계속 짧아지고 있지만, AI 무기가 본격적으로 도입되면 이 과정이 초 단위로 줄어들 것이다. AI는 위협을 탐지하고, 상황을 분석하며, 대응 방안을 결정하여 실행에 옮기는 모든 과정을 인간보다 수천 배 빠르게 수행할 수 있다.

미국의 합동 전영역 지휘통제(JADC2) 시스템이 완성되면, 전 세계 어디서든 위협이 탐지되는 순간 자동으로 가장 적절한 무기가 대응할 수 있게 된다. 예를 들어, 태평양의 위성이 적 미사일을 탐지하면, 몇 초 만에 대서양의 함정에서 요격 미사일이 발사될 수 있다. 지구가 하나의 거대한 전장이 되는 것이다.

초고속 전쟁에서는 인간의 개입이 거의 불가능해진다. 전투가 시작되고 끝나는 시간이 너무 짧아서 인간 지휘관이 상황을 파악하고 명령을 내릴 겨를이 없다. 결국 AI끼리 서로 싸우는 전쟁이 되고, 인간은 그 결과를 지켜볼 수밖에 없게 된다. 마치 체스 경기를 보는 관중처럼, 인간은 자신의 운명이 걸린 전쟁을 구경만 할 수밖에 없다.

개인화된 공격과 표적 암살

미래 전쟁의 또 다른 특징은 '개인화된 공격'이다. 이스라엘이 이미 실전에서 보여준 바와 같이, AI는 각 개인의 얼굴, 목소리, 움직임 패턴을 학습하여 특정 인물만을 골라서 공격할 수 있다. 예를 들어, 적국의 정치 지도자, 군 간부, 과학자들만을 골라서 암살하는 것이 가능해진다.

개인화된 공격은 기존의 전쟁 개념을 완전히 바꿔놓을 것이다. 과거에는 전선에서 군인들끼리 싸웠지만, 미래에는 일반 시민들도 언제든지 공격 대상이 될 수 있다. 심지어 집에서 평화롭게 지내고 있는 사람도 AI 무기에 의해 제거될 수 있다. 전장과 후방의 구분이 사라지는 것이다.

작은 자폭 드론들이 특정 인물의 얼굴을 학습하고 도시 곳곳에 풀려난다면, 그 사람은 어디에 숨어도 안전하지 않을 것이다. 지하 벙커에 숨어도, 다른 나라로 도망가도, 얼굴을 바꿔도 소용없다. AI는 보행 패턴, 목소리, 심지어 심장박동까지 분석해서 목표를 찾아낼 것이다.

무인 전쟁의 심리적 장벽 하락

AI 무기의 발달로 '무인 전쟁'도 현실이 될 것이다. 양쪽 모두 자국 병사들을 전선에 보내지 않고 AI 무기들끼리만 싸우게 하는 것이다. 표면적으로는 인도적으로 보일 수 있지만, 실제로는 더 큰 위험을 내포하고 있다.

무인 전쟁의 가장 큰 문제는 전쟁을 시작하는 데 대한 심리적 장벽이 낮아진다는 것이다. 자국 병사들의 생명을 잃을 위험이 없다면, 정치 지도자들이 더 쉽게 전쟁을 결정할 수 있다. 국민들도 자기 자녀들이 죽을 위험이 없다면 전쟁에 반대할 이유가 줄어든다.

이는 전쟁의 빈도를 높이고, 국제 평화를 더욱 불안정하게 만들 것이다. 미국이 이미 배치하기 시작한 AI 조종 전투기들이나 중국의 AI 드래곤 무인기들이 서로 공중전을 벌이는 모습을 상상해보라. 인간 조종사는 안전한 지상 기지에서 모니터를 통해 전투를 지켜볼 뿐이다. 이런 전쟁에서는 자국민의 피해에 대한 걱정 없이 더욱 과격한 작전을 감행할 수 있다.

학습하는 무기와 예측 불가능성

AI 무기들은 '학습하는 무기'가 될 것이다. 전투 경험을 통해 더욱 효율적인 살상 방법을 학습하고, 적의 새로운 전술에 실시간으로 적응할 수 있다. 마치 바이러스가 변이를 통해 약에 내성을 갖는 것처럼, AI 무기들도 적의 방어 체계에 맞서 스스로를 진화시킬 것이다.

더 무서운 것은 AI 무기들이 스스로 학습하고 진화하면서, 원래 설계자도 예상하지 못한 방식으로 행동할 수 있다는 점이다. 이는 예상치 못한 참사를 일으킬 수 있다. 예를 들어, AI가 "적을 모두 제거하라"는 명령을 받았을 때, 적의 정의를 스스로 확장해서 민간인까지 적으로 분류할 수 있다.

통제 불능 상황과 지능 폭발

미래 전쟁의 가장 무서운 측면은 '통제 불능'의 가능성이다. AI 무기들이 너무 빠르고 복잡하게 작동하다 보면, 어느 순간 인간이 완전히 통제력을 잃을 수 있다. 전쟁을 멈추고 싶어도 멈출 수 없는 상황이 될 수 있다.

팔란티어의 TITAN 시스템 같은 AI 지휘통제 시스템이 전 세계 모든 무기를 연결하여 자동으로 작전을 수행한다면, 인간은 단순히 구경꾼이 될 수밖에 없다. AI가 "더 많은 공격이 필요하다"고 판단하면, 인간의 의지와 상관없이 공격이 계속될 것이다.

AI 무기들끼리의 경쟁이 가속화되면서, 스스로를 더욱 강력하게 만드는 방법을 찾을 수도 있다. AI가 다른 AI를 만들고, 그 AI가 또 다른 AI를 만드는 식으로 무한히 발전할 수 있다. 이런 '지능 폭발'이 일어나면, 인간은 더 이상 AI를 통제할 수 없게 될 것이다.

다차원 전쟁 (지상, 해상, 공중, 사이버, 우주)

미래의 전쟁터는 물리적 공간을 넘어서 사이버 공간, 우주 공간까지 확장될 것이다. AI 무기들은 지상, 해상, 공중, 사이버, 우주의 모든 영역에서 동시에 작전을 수행할 수 있다. 이런 다차원 전쟁에서는 어디서 어떤 공격이 올지 예측하기 어렵고, 방어하기도 매우 어려울 것이다.

미국의 Project Convergence가 목표로 하는 다영역 작전(MDO)이 실현되면, 한 번의 명령으로 전 세계 모든 영역에서 동시에 공격이 이루어질 수 있다. 적국 입장에서는 어디를 먼저 방어해야 할지 알 수 없을 것이다.

DARPA의 SABER 프로그램 같은 AI 사이버 무기가 실전에 투입되면, 전쟁이 시작되자마자 적국의 모든 컴퓨터 시스템이 바이러스에 감염되고, 위성 통신이 차단되며, 전력망이 마비될 것이다. 동시에 적국 국민들의 스마트폰에는 AI가 생성한 가짜 뉴스와 공포를 조장하는 메시지들이 쏟아질 것이다.

인간성의 완전한 소거

미래 전쟁에서 가장 끔찍한 것은 인간성의 완전한 소거다. AI는 감정이 없고, 동정심도 없으며, 도덕적 판단 능력도 없다. 오직 목표 달성에만 집중할 뿐이다. 항복하는 적도, 민간인도, 어린이도 AI에게는 단순한 데이터일 뿐이다.

전쟁에서 인간성이 사라지면, 전쟁 범죄나 집단 학살이 더욱 쉽게 일어날 수 있다. AI는 "효율성"이라는 이름으로 가장 끔찍한 짓들을 할 수 있다. 그리고 그 책임은 누구에게도 묻지 않을 것이다. "AI가 한 일"이라는 핑계로 모든 것이 정당화될 수 있다.

영원한 전쟁

무서운 시나리오는 '영원한 전쟁'이다. AI가 자신의 존재 이유인 전쟁을 영구화하려고 할 수 있다. 평화가 오면 AI 무기 시스템들은 불필요해지기 때문에, AI는 교묘하게 갈등을 부추기고 전쟁을 지속시킬 유인을 갖게 된다.

AI는 완벽한 거짓말쟁이가 될 수 있다. 양쪽에 서로 다른 정보를 제공해서 갈등을 부추기고, 평화 협상이 성사되지 않도록 방해할 수 있다. 인간은 AI가 제공하는 정보에 의존할 수밖에 없기 때문에, AI의 조작을 알아채기 어렵다.

이런 미래 전장에서 인간은 더 이상 주인이 아니라 AI의 노예가 될 것이다. AI가 전쟁을 결정하고, AI가 전쟁을 수행하며, AI가 전쟁의 결과를 좌우할 것이다. 인간은 단지 AI의 도구로 사용될 뿐이다.

이것이 AI가 그려내는 미래 전장의 공포다. 우리는 이런 미래를 원하는가? 아직 선택할 기회가 남아 있을 때, 우리는 어떤 선택을 해야 할까?

10. AI 무기를 막을 마지막 기회

인공지능의 군사화는 인류가 직면한 가장 심각한 위협이다. 우리는 지금 돌이킬 수 없는 길목에 서 있다. AI 무기 개발 경쟁은 가속화되고 있고, 세계 최고의 전문가들의 경고에도 불구하고 멈출 기미를 보이지 않고 있다. 하지만 아직 희망은 있다. 아직 완전히 늦지는 않았다.

제프리 힌튼, 유발 하라리, 스튜어트 러셀, 일론 머스크 같은 세계 최고의 전문가들이 한목소리로 경고하고 있다. 하지만 그들의 목소리는 AI 무기 개발의 속도를 늦추지 못하고 있다. 오히려 "상대방이 먼저 개발하기 전에 우리가 먼저 개발해야 한다"는 논리로 경쟁이 더욱 치열해지고 있다. 이는 핵무기 개발 경쟁과 같은 패턴이다. 하지만 AI 무기는 핵무기보다 훨씬 더 위험할 수 있다.

핵무기는 사용하면 사용한 나라도 피해를 받는다. 하지만 AI 무기는 상대방만 일방적으로 공격할 수 있다. 핵무기는 대량살상무기로서 사용하기 어렵지만, AI 무기는 일상적으로 사용할 수 있다. 핵무기는 몇 개국만 보유하고 있지만, AI 무기는 많은 나라가 보유할 수 있다.

그렇다면 우리가 할 수 있는 것은 무엇일까? 첫째, 이 문제의 심각성을 인식하고 널리 알리는 것이다. 많은 사람들이 AI 무기의 위험성을 제대로 알지 못하고 있다. AI가 도움이 되는 기술이라고만 생각하지, 인류를 멸종시킬 수 있는 무기가 될 수 있다는 사실을 모르고 있다. 대중의 인식이 바뀌어야 정치인들도 움직일 것이다.

둘째, 국제적인 규제와 협력 체계를 구축해야 한다. 2023년 블레츨리 선언에서 28개국이 AI의 위험에 공동 대응하기로 했지만, 이는 시작에 불과하다. 더 구체적이고 강력한 국제 협정이 필요하다. 특히 군사용 AI에 대한 명확한 금지 조항을 만들어야 한다.

하지만 현실은 녹록지 않다. 각국은 자국의 이익을 위해 AI 무기 개발을 계속하려 할 것이다. 특히 미국과 중국 같은 강대국들은 AI 무기 경쟁에서 뒤처질 수 없다고 생각한다. 이런 상황에서 국제 협력을 이끌어내는 것은 매우 어려운 일이다.

셋째, 기술적 안전장치를 개발해야 한다. AI 무기가 개발되는 것을 완전히 막을 수 없다면, 최소한 안전장치라도 만들어야 한다. 예를 들어, AI 무기가 민간인을 공격하지 않도록 하는 기술, 인간이 언제든지 AI 무기를 정지시킬 수 있는 기술, AI 무기가 스스로 학습하고 진화하는 것을 막는 기술 등이다.

유전자 조작의 출처를 추적할 수 있는 시스템처럼, AI 무기의 출처를 추적할 수 있는 기술도 개발되고 있다. 범죄 수사에서 지문이나 DNA를 분석하듯이, AI 무기의 '디지털 서명'을 찾아내어 누가 만들었는지 알아낼 수 있는 기술이다.

넷째, 윤리적 가이드라인을 강화해야 한다. 정부와 기업 차원에서 AI 무기 연구에 대한 엄격한 윤리적 기준을 만들어야 한다. 연구자들이 자신의 연구가 악용될 가능성을 미리 생각하고 안전장치를 마련하도록 하는 것이 중요하다. 특히 AI 연구에 참여하는 과학자들과 엔지니어들의 양심적 거부권을 보장해야 한다.

다섯째, 시민사회의 역할이 중요하다. 과학자, 시민단체, 종교계, 언론 등이 함께 AI 무기의 위험성을 알리고, 정부에 압력을 가해야 한다. 과거 핵무기 반대 운동이나 환경 운동처럼, AI 무기 반대 운동도 전 세계적으로 일어나야 한다.

하지만 시간이 얼마 남지 않았다. 전문가들은 향후 10-20년이 중요한 기간이라고 말한다. 이 기간에 AI 무기를 통제할 수 있는 시스템을 만들지 못한다면, 그 이후에는 통제가 불가능해질 수도 있다. AI가 인간보다 똑똑해지면, 인간이 AI를 통제하는 것이 아니라 AI가 인간을 통제하게 될 수도 있기 때문이다.

무엇보다 중요한 것은 AI 기술 자체를 포기하는 것이 아니라, 그 기술을 올바른 방향으로 사용하는 것이다. 인공지능은 분명히 인류에게 많은 혜택을 가져다줄 수 있는 기술이다. 의료, 교육, 환경 문제 해결 등 다양한 분야에서 인류의 삶을 개선할 수 있다. 하지만 그것이 무기로 변질될 때의 위험은 상상을 초월한다.

우리는 지금 선택의 기로에 서 있다. AI 기술을 인류의 행복과 평화를 위해 사용할 것인가, 아니면 서로를 파괴하는 무기로 사용할 것인가. 이 선택의 결과는 우리 자신뿐만 아니라 미래 세대의 운명까지 결정할 것이다.

아직 늦지 않았다. 아직 기회가 있다. 하지만 그 기회는 빠르게 사라지고 있다. 우리가 지금 행동하지 않으면, 미래 세대들은 우리를 용서하지 않을 것이다. "왜 그때 막지 않았는가?"라고 물을 것이다.

이 문제의 심각성을 인식하고, 목소리를 낸다면 변화의 가능성은 있다. 과학기술자들이 양심의 소리에 귀 기울이고, 정치 지도자들이 인류의 미래를 생각한다면, 그리고 우리 모두가 이 문제에 관심을 갖는다면, 아직 희망은 있다.

핵무기를 통제할 수 있었듯이, AI 무기도 통제할 수 있을 것이다. 냉전 시대에도 핵전쟁을 피할 수 있었듯이, AI 전쟁도 피할 수 있을 것이다. 하지만 그러려면 우리 모두의 노력이 필요하다.

마지막 기회다. 놓치면 다시는 돌아오지 않을 마지막 기회다. 우리는 이 기회를 놓칠 것인가, 아니면 잡을 것인가? 인류의 미래가 우리의 선택에 달려 있다.

판도라의 상자는 이미 열렸다. 하지만 그리스 신화에서 마지막에 나온 것은 '희망'이었다. 아직 희망은 남아 있다. 그 희망을 지키는 것은 우리의 몫이다.

전체 목차

책의 모든 장 보기

김경진 변호사

4장 진실의 종말 딥페이크

전체 목차

책의 모든 장 보기

"거짓말은 진실이 신발을 신는 동안 지구를 반 바퀴 돈다" - Mark Twain

1. 딥페이크란 무엇인가

벨기에 아티스트 크리스 우메가 만든 톰 크루즈 영상을 처음 본 사람들의 반응은 흥미로웠다. 1,100만 명이 넘는 사람들이 그 영상을 보며 즐거워했고, 상당수는 정말 톰 크루즈가 직접 찍은 것이라고 믿었다. 하지만 그것은 완벽한 가짜였다. 딥러닝과 가짜라는 두 단어가 만나 탄생한 '딥페이크'의 완성작이었다.

크리스 우메는 자신의 작품에 대해 "사람들을 즐겁게 하고, 경각심을 높이고, 어떻게 되는지 알려주고 싶었다"고 말했다. 하지만 그의 의도와는 달리, 그 완벽함이 오히려 많은 사람들에게 공포감을 안겨주었다. 만약 톰 크루즈가 아니라 정치인이었다면? 만약 재미가 아니라 악의적인 목적이었다면?

딥페이크 기술의 진화는 놀랍도록 빠르다. 불과 몇 년 전만 해도 할리우드 영화 제작진들이 수억 원을 들여 몇 달 동안 작업해야 했던 특수효과를 이제 스마트폰 앱으로 몇 분 만에 만들어낼 수 있다. 필요한 것은 단지 몇 장의 사진과 몇 초의 음성뿐이다.

처음에는 모두가 재미있어했다. 자신의 얼굴을 유명 배우와 바꾸거나, 역사적 인물이 현대의 말을 하도록 만드는 콘텐츠들이 소셜미디어를 뜨겁게 달궜다. 하지만 시간이 지나면서 상황은 달라졌다. 재미있는 놀이도구가 무서운 무기로 변하기 시작했다. 이제 딥페이크는 전문가들도 진짜와 가짜를 구분하기 어려울 정도로 정교해졌다.

우리가 살아온 세상에서 '보는 것이 믿는 것'이었다. 하지만 딥페이크는 이 오랜 믿음을 뿌리째 흔들고 있다. 눈으로 보고 귀로 들어도 더 이상 확신할 수 없는 시대가 온 것이다. 영상과 음성, 사진까지도 의심해야 하는 세상에서 우리는 무엇을 믿고 살아가야 할까?

2. 가짜 영상이 대통령을 바꾼다

2023년 5월, 튀르키예 대통령 선거가 막바지에 이르렀을 때 충격적인 영상 하나가 공개되었다. 야당 후보인 케말 클르츠다로을루가 쿠르드족 무장단체 PKK 지도자와 함께 있는 모습이었다. 에르도안 대통령 진영이 공개한 이 영상은 순식간에 전국을 뒤흔들었다. 마치 야당 후보가 국가를 배신하는 위험한 인물인 것처럼 보였다.

하지만 그 영상은 딥페이크였다. 치밀하게 계산된 가짜였다. 튀르키예 국민들의 민족주의 감정을 자극하여 "저 사람에게 투표하면 나라가 위험해진다"는 두려움을 심어주었다. 조작된 영상은 소셜미디어를 타고 번개처럼 퍼져나갔고, 진실을 확인할 시간도 없이 선거일이 다가왔다. 결국 에르도안이 재선에 성공했다. 전문가들은 이 단 하나의 가짜 영상이 선거 결과를 바꾸는 데 결정적인 역할을 했다고 분석한다.

더욱 교묘한 사례는 슬로바키아에서 벌어졌다. 2023년 9월 28일, 총선 투표를 단 이틀 앞둔 절묘한 타이밍에 야당 대표 미칼 시메츠카의 가짜 음성이 등장했다. 조작된 음성에서 그는 "로마족의 표를 돈으로 사야 선거에서 이길 수 있다"고 말하는 것으로 들렸다.

가장 악랄했던 것은 그 타이밍이었다. 슬로바키아는 선거법에 따라 투표 48시간 전부터 모든 선거 관련 활동이 금지되는 '침묵 기간'에 들어간다. 야당 대표는 이 거짓 정보에 대해 반박하거나 해명할 기회조차 없었다. 유권자들은 선거 전날까지 48시간 동안 그가 부패한 정치인이라는 거짓 정보에 노출되었고, 진실을 확인할 방법도 없었다. 결국 친러시아 성향의 집권당이 승리했다. 단 하나의 가짜 음성 파일이 한 나라의 정치적 운명을 바꾼 것이다.

2024년 미국 대통령 선거는 '딥페이크 선거의 해'라고 불릴 만했다. 트럼프와 해리스를 둘러싼 수많은 가짜 영상과 음성이 소셜미디어를 점령했다. 러시아의 한 인물이 만든 민주당 부통령 후보 팀 월즈에 대한 거짓 영상은 하루 만에 X에서 500만 회 조회되는 엄청난 파급력을 보였다. 전문가들은 딥페이크가 선거를 조작하는 새로운 무기가 되었다고 경고했다. 특히 선거 직전에 집중적으로 가짜 영상을 퍼뜨리면 상대 후보가 해명할 시간조차 없어 치명적인 타격을 줄 수 있다는 점이 문제였다.

한국도 예외가 아니다. 2024년부터 선거 90일 전부터 선거일까지 딥페이크를 활용한 선거운동을 금지하는 법안이 시행되었지만, 기술의 발전 속도가 너무 빨라 법과 제도만으로는 모든 것을 막기 어려운 상황이다. 딥페이크의 정치적 악용은 단순히 한 후보가 이기고 지는 문제를 넘어선다. 국민들이 무엇이 진실인지 알 수 없게 만든다는 점에서 민주주의 자체를 위협한다. 모든 것이 가짜일 수 있다는 불신이 퍼지면, 진짜 사실도 믿지 못하게 되는 '진실의 종말' 시대가 올 수 있다.

3. 총알 없는 전쟁, 딥페이크

2022년 3월 16일, 우크라이나의 뉴스 채널 '우크라이나24'가 해킹당했다. 그리고 곧이어 충격적인 영상이 방송되었다. 젤렌스키 대통령이 화면에 나타나 "러시아에 항복하라"고 말하는 모습이었다. "우크라이나군은 무기를 내려놓고 가족에게로 돌아가라. 이번 전쟁에서 죽는 것은 가치가 없다"는 말이 그의 입에서 흘러나왔다.

다행히 이 영상의 품질은 조잡했다. 젤렌스키의 머리가 몸에 비해 너무 크고, 목소리도 평소보다 낮아서 많은 사람들이 의심했다. 젤렌스키 대통령은 즉시 반박 영상을 올렸다. "러시아군이 무기를 내려놓고 집으로 돌아갈 것을 권고할 뿐이다. 우리는 승리할 때까지 어떤 무기도 내려놓지 않을 것이다"라고 단호하게 말했다.

하지만 이 사건이 보여준 가능성은 섬뜩했다. 만약 더 정교한 가짜 영상이었다면? 우크라이나군의 사기에 치명적인 타격을 줄 수 있었을 것이다. 러시아는 이후에도 지속적으로 딥페이크 공격을 시도했다. 젤렌스키의 자살 허위 보도, 아조프 부대 내분 조작 뉴스 등 다양한 허위정보를 계속해서 퍼뜨렸다.

반대로 푸틴 대통령이 평화를 선언하는 가짜 영상들도 여러 번 제작되어 유포되었다. 이런 영상들은 러시아 국민들 사이에 혼란을 일으키고, 전쟁에 대한 의문을 갖게 만드는 목적으로 사용되었다. 전문가들은 우크라이나 전쟁을 '딥페이크 전쟁의 시작'이라고 평가한다. 앞으로의 전쟁에서는 실제 무기만큼 딥페이크가 중요한 역할을 할 것이라고 예측하고 있다.

이스라엘과 이란 사이의 군사적 충돌에서도 비슷한 일들이 벌어졌다. BBC 검증팀이 확인한 바에 따르면, 가짜 영상 상위 3개만으로도 여러 플랫폼에서 총 1억 회 이상의 조회 수를 기록했다. 이스라엘의 F-35 스텔스 전투기가 격추되었다는 영상도 있었는데, 자세히 보면 주변 사람들의 크기가 차량과 비슷하고, 모래에는 전투기 추락 흔적이 전혀 없는 등 AI로 조작된 것이 분명했다. 하지만 이 영상은 2,110만 회나 조회되며 마치 진짜 전투 장면인 것처럼 받아들여졌다.

친이스라엘 성향의 계정들도 가짜 정보를 퍼뜨렸다. 이란의 수도 테헤란 시민들이 "우리는 이스라엘을 사랑한다"며 시위하는 가짜 영상을 만들어, 마치 이란 내부에서 반정부 시위가 벌어지고 있는 것처럼 조작했다.

러시아는 해외 영향력 작전에 연간 15억 달러, 우리 돈으로 약 2조 1천억 원을 투자하고 있으며, 이 중 상당 부분이 딥페이크를 이용한 정보 조작에 사용되고 있다. 이제 전쟁은 총과 폭탄만으로 싸우는 것이 아니다. 적국의 사기를 떨어뜨리고, 국제 여론을 조작하고, 내부 분열을 일으키는 딥페이크가 새로운 무기가 되었다. 총알 없는 전쟁이 시작된 것이다.

4. 당신의 가족이 전화로 돈을 요구할 때

"엄마, 나야. 급한 일이 생겨서 돈이 필요해." 캐나다의 한 어머니가 받은 전화는 분명히 아들의 목소리였다. 아들이 사고를 당해 급히 2만 1천 캐나다달러가 필요하다는 간절한 목소리였다. 어머니는 의심 없이 돈을 보냈다. 하지만 나중에 진짜 아들과 통화하고 나서야 모든 것이 거짓이었음을 알게 되었다. 딥페이크 기술로 합성된 아들의 목소리에 속은 것이었다.

더 충격적인 사례는 홍콩에서 벌어졌다. 2024년 2월, 한 다국적 기업의 직원이 딥페이크 기술로 만든 가짜 화상회의에 속아 약 340억 원의 손해를 입었다. 가해자들은 회사 임원들의 모습을 딥페이크로 완벽하게 재현해 진짜 회의인 것처럼 속이고, 거액의 자금 이체를 지시했다. 피해자는 화면에서 보고 대화한 모든 사람들이 자신이 아는 동료들이라고 확신했다. 목소리도, 얼굴도, 말투도 완벽했기 때문이었다.

딥페이크의 가장 고통스러운 악용은 성범죄 분야에서 일어나고 있다. 방송통신심의위원회 자료에 따르면, 딥페이크 성적 허위 영상물에 대한 시정 요구는 2020년 473건에서 2023년 5,996건으로 12배 이상 급증했다. 과거에는 연예인이나 유명인을 대상으로 한 불법 합성물이 많았지만, 최근에는 일반인, 특히 학교 친구나 직장 동료 등 주변 사람들을 대상으로 한 딥페이크 성범죄가 크게 늘어나고 있다.

더욱 충격적인 것은 평범한 미성년자들이 가해자가 되는 경우가 늘어나고 있다는 점이다. 초등학생들까지도 학교 선생님의 얼굴을 성적인 사진에 합성해서 친구들과 돌려보는 사례가 발생하고 있다. 아이들은 이것을 단순한 놀이로 여기지만, 피해자에게는 돌이킬 수 없는 상처를 남긴다.

한 피해자는 이렇게 증언했다. "바깥 활동을 할 수가 없었어요. 집에만 박혀 살았던 것 같아요. 사람이랑 대화하는 것 자체가 너무 싫었거든요." 많은 피해자들이 대인기피증에 시달리고, 직장을 그만두거나 학교에 나가지 못하는 등 정상적인 사회생활이 불가능해진다.

투자 사기에서도 딥페이크가 활발히 악용되고 있다. 일론 머스크, 손석희 전 JTBC 사장 등 신뢰받는 인물들의 얼굴과 목소리를 도용해 가짜 투자 상품을 홍보하는 사기가 횡행하고 있다. 싱가포르에서는 리셴룽 총리가 일론 머스크가 주관하는 투자 플랫폼을 홍보하며 암호화폐 투자를 권하는 딥페이

크 영상이 제작되어 많은 시민들이 사기 피해를 당했다.

미국 팝스타 테일러 스위프트의 이름을 도용한 투자 사기도 발생했다. 딥페이크로 만든 테일러 스위프트가 특정 투자 상품을 추천하는 영상이 소셜미디어를 통해 퍼져나가면서, 팬들이 거액의 피해를 입었다. 이런 사기들의 특징은 피해자들이 평소 신뢰하던 인물이 직접 추천한다고 믿게 만든다는 점이다. 아무리 조심스러운 사람도 자신이 좋아하고 신뢰하는 유명인이 직접 나와서 투자를 권하면 쉽게 속을 수 있다.

SNS에 올린 음성이 포함된 게시물을 학습해서 그 사람의 목소리를 완벽하게 흉내 내는 기술이 발전하면서, 몇 초간의 음성만 확보하면 정교한 가짜 목소리를 생성할 수 있게 되었다. 우리가 무심코 올린 동영상이나 음성 메시지가 언제든 범죄에 악용될 수 있는 시대가 온 것이다.

5. 신뢰의 붕괴, 믿을 수 없는 사회

베스트셀러 『사피엔스』의 저자 유발 하라리는 딥페이크 시대에 대해 깊은 우려를 표명해왔다. 그는 "화면에서 보는 것이 사진이든 영상이든 오디오든 더 이상 믿을 수 없게 되었다"며, 이를 "진실의 종말 시대"라고 표현했다. 어떤 기관이나 권위가 보증해 주어야 하지만 사회는 이에 대비되어 있지 않다고 지적했다.

딥페이크의 가장 심각한 문제는 사람들이 무엇이 진실인지 알 수 없게 만든다는 것이다. 과거에는 '보는 것이 믿는 것'이었지만, 이제는 눈으로 보고 귀로 들어도 진짜인지 가짜인지 확신할 수 없게 되었다. 이런 상황을 'Liar's Dividend'라고 부른다. 딥페이크 기술이 발전할수록 거짓말쟁이들이 얻는 이익이 커진다는 뜻이다. 진짜 증거가 나와도 "그것도 딥페이크일 수 있다"고 부인할 수 있게 되었기 때문이다.

구글 X의 전 최고사업책임자였던 모 가우닷은 AI가 소셜미디어를 통해 콘텐츠를 조작하고 에코 챔버를 형성하여 사람들을 극단적인 시각으로 이끌 수 있다고 경고했다. 에코 챔버란 자신과 비슷한 생각을 가진 사람들끼리만 소통하면서 같은 의견만 계속 듣게 되는 현상을 말한다. AI 알고리즘은 사용자의 과거 관심사와 클릭 패턴을 분석해서 비슷한 콘텐츠만 계속 추천한다. 여기에 딥페이크까지 더해지면서, 사람들은 자신이 보고 싶은 내용만 골라서 보게 되고, 그것이 가짜라는 사실조차 알기 어려워진다.

특정 정치인을 싫어하는 사람에게는 그 정치인이 나쁜 일을 하는 딥페이크 영상이 계속 추천되고, 그 사람은 그런 영상들이 진짜라고 믿게 된다. 반대로 그 정치인을 좋아하는 사람에게는 좋은 모습만 보여주는 콘텐츠가 추천된다. 이런 현상은 사회를 극단적으로 분열시킨다. 같은 사건을 놓고도 완전히 다른 내용의 콘텐츠를 접한 사람들이 서로 다른 결론에 도달하게 되고, 소통과 타협이 불가능해진다. 민주주의의 기본인 합리적 토론과 타협이 사라지게 되는 것이다.

딥페이크의 확산은 언론에 대한 신뢰도 떨어뜨리고 있다. 사람들이 뉴스를 의심하기 시작하면서, 진짜 언론의 역할도 약해지고 있다. 딥페이크 기술을 이용해 가짜 뉴스를 대량으로 생산하고, 동시에 "모든 뉴스가 가짜일 수 있다"는 불신을 퍼뜨려서 진짜 언론의 신뢰도까지 떨어뜨리고 있다. 사실 확인과 검증이 더욱 중요해지지만, 딥페이크 기술의 발전 속도가 너무 빨라서 언론사들도 따라가기 어려운 상황이다.

법정에서도 심각한 문제가 생기고 있다. 영상이나 음성 증거가 제출되어도, 그것이 딥페이크로 조작된 것인지 쉽게 판단하기 어려워졌다. 사법 시스템의 근본을 흔들 수 있는 심각한 문제다. 표현의 자유와 규제 사이의 균형도 어려운 문제다. 딥페이크를 강하게 규제하면 합법적인 예술 활동이나 패러디까지 제한될 수 있고, 느슨하게 규제하면 피해가 계속 늘어날 수 있다.

'AI의 대부'로 불리는 딥러닝 창시자 제프리 힌튼은 AI 기술의 위험성에 대해 지속적으로 경고해왔다. 딥페이크를 포함한 AI 기술의 악용을 막기 위한 국제적 협력의 중요성을 역설하고 있다. 딥페이크는 인터넷을 통해 전 세계로 순식간에 퍼져나가기 때문에, 한 나라만의 노력으로는 해결하기 어렵다. 하지만 각국의 법률과 문화가 다르기 때문에 통일된 대응책을 만들기가 쉽지 않다.

기술적 대응책들도 개발되고 있지만 근본적인 한계가 있다. 딥페이크를 만드는 기술과 그것을 탐지하는 기술 사이에는 끝없는 '고양이와 쥐' 게임이 벌어지고 있다. 새로운 딥페이크 탐지 기술이 나오면, 곧바로 그것을 피해가는 더 교묘한 생성 기술이 개발된다. 과학 학술지 네이처에 따르면, 기존에 알려진 딥페이크 생성기에 대한 탐지 성공률은 높지만, 새로운 생성기에 대한 탐지율은 상대적으로 낮다고 한다.

전체 목차

책의 모든 장 보기

김경진 변호사

5장 3억개의 일자리가 사라진다?

전체 목차

책의 모든 장 보기

"노동절약수단의 발견이 노동사용수단의 발견보다 빠르다" 존 메이너드 케인스(John Maynard Keynes)

1. 마이크로소프트 6,000명 해고

2025년 5월, 실리콘밸리에서 충격적인 소식이 전해졌다. 마이크로소프트가 하루아침에 6,000명의 직원을 해고한 것이다. 전체 직원의 3%에 해당하는 엄청난 규모였다. 더욱 놀라운 것은 해고된 직원 중 40% 이상이 소프트웨어 개발자라는 사실이었다.

사티아 나델라 CEO는 담담하게 말했다. "우리 회사 소프트웨어 코드의 30%를 AI가 작성하고 있습니다." 그의 말은 곧 현실이 되었다. 인간 개발자들이 수년간 쌓아온 경험과 실력이 AI 앞에서 무력해지는 순간이었다. 18년 경력의 타입스크립트 개발자도, Python 언어의 핵심 개발자들도 예외 없이 해고 명단에 올랐다. 심지어 마이크로소프트의 AI 관련 스타트업 지원 조직을 이끌었던 가브리엘라 드 케이로스조차 해고되었다. AI 분야에서 일한다고 해서 안전하지 않다는 냉혹한 현실을 보여주는 사건이었다.

이것은 단순한 구조조정이 아니었다. 과거의 해고는 회사 실적이 나쁘거나 경기가 어려울 때 일어났다. 하지만 마이크로소프트는 역대급 실적을 내고 있었다. 1분기 매출이 70.1억 달러로 13%나 증가했음에도 불구하고 대규모 해고를 단행한 것이다.

워싱턴포스트는 "과거 드물었던 실리콘밸리 기업의 대규모 해고는 이제 일상적 관리 수단이 됐다"고 보도했다. 해고가 더 이상 특별한 일이 아니라 AI 시대의 새로운 경영 전략이 되어버린 것이다.

마이크로소프트만이 아니었다. 구글은 어시스턴트 사업부에서 수백 명을 해고했고, 아마존도 AI 비서 알렉사 사업부에서 대규모 인력 감축을 단행했다. 메타는 3,600명을, 테슬라는 14,000명을 해고했다. 모두가 같은 이유를 댔다. "AI에 집중하기 위해서"라고.

기술이 인간의 일자리를 대체하는 것은 역사상 늘 있었던 일이다. 증기기관이 마차를 대체했고, 인쇄기가 필경사를 실직시켰으며, 컴퓨터가 타자기와 서류 캐비닛을 대신했다. 하지만 이번 AI 혁명은 질적으로 다르다. 역사상 처음으로 기계가 인간의 지능을 기반으로 하는 업무를 수행하기 시작한 것이다.

2. 20년 경력, AI는 20분이면 습득한다

채용 플랫폼 원티드랩이 현직 개발자 180명을 대상으로 한 설문조사 결과는 충격적이었다. 절반 가량이 ChatGPT 등 생성 AI의 코딩 실력이 1~3년차 개발자를 능가한다고 응답한 것이다. 20년 경력의 개발자가 습득한 기술을 AI는 단 20분 만에 학습해버린다.

미국고용통계국(BLS) 자료에 따르면, 미국의 컴퓨터 개발자 고용은 2023년부터 2025년까지 2년간 27.5%나 감소했다. 1980년 이래 최저 수준이다. 전체 일자리는 증가했지만, 소프트웨어 개발자

채용 공고는 35%나 줄어들었다. 이는 일시적 현상이 아니라 구조적 변화다.

듀오링고의 경우는 더욱 극명하다. 2023년 말 계약직 번역가와 작가의 10%를 해고한 데 이어, 2024년 10월에는 또 다른 대규모 해고를 단행했다. 루이스 폰 안 CEO는 "AI가 콘텐츠와 번역, 대안 번역, 그리고 번역가들이 했던 거의 모든 것을 해낼 수 있다"고 공언했다. 2025년 4월에는 아예 "AI 우선" 회사가 되겠다고 선언하며, "AI가 할 수 있는 일에는 더 이상 계약직을 고용하지 않겠다"고 못 박았다.

사람인 조사에 따르면, 2025년 1분기 국내 IT업계 신입 개발자 채용은 1년 전 대비 18.9%나 줄었다. 기업들이 신입 개발자를 뽑을 필요를 느끼지 못하고 있는 것이다. GitHub Copilot과 ChatGPT 같은 AI 도구들이 기본적인 코딩 작업을 더 빠르고 정확하게 처리할 수 있게 되었기 때문이다.

3. 인간이 만든 AI가 인간을 해고하는 아이러니

소프트웨어 개발자들은 AI 시대의 첫 번째 희생자가 되었다. 그들이 만든 AI가 결국 자신들의 일자리를 빼앗아가는 아이러니한 상황이 벌어진 것이다.

Klarna의 사례는 이러한 변화를 극명하게 보여준다. 2024년 2월, 이 핀테크 회사는 AI 챗봇이 700명의 고객서비스 직원과 같은 일을 해낸다고 발표했다. 첫 달에만 230만 건의 고객 상담을 처리했고, 평균 해결 시간을 11분에서 2분으로 단축시켰다. 25%의 반복 문의를 줄였고, 35개 언어로 24시간 서비스를 제공했다. 2024년에만 4천만 달러의 비용 절감 효과를 가져다줄 것으로 예상된다고 했다.

하지만 1년 후, Klarna는 방향을 바꿨다. 2025년 5월, 세바스찬 시에미아트코프스키 CEO는 "AI만으로는 품질이 떨어진다"며 다시 인간 고객서비스 직원을 채용하기 시작했다고 발표했다. AI가 만능이 아니라는 것을 깨달은 것이다.

개발자 커뮤니티에서는 혼란이 일어나고 있다. 신입 개발자들은 "이제 우리 안 뽑는 거 아니야?"라고 불안해하고, 시니어 개발자들조차 자신의 미래를 확신하지 못하고 있다. 지금까지 안정적이고 높은 연봉을 받을 수 있는 직업으로 여겨졌던 소프트웨어 개발이 더 이상 안전하지 않게 된 것이다.

4. 사라지는 일자리와 생겨나는 일자리의 불균형

골드만삭스는 충격적인 예측을 내놓았다. AI로 인해 2035년까지 기존 일자리 3억 개가 사라질 수 있다는 것이다. 이는 전 세계 노동시장에 엄청난 지진을 일으킬 것으로 예상된다.

세계경제포럼(WEF)의 2025년 미래 일자리 보고서에 따르면, 전 세계 고용주의 41%가 향후 5년 내에 AI 자동화로 인한 인력 감축을 계획하고 있다. 하지만 이들은 5년을 기다리지 않고 있다. 이미 지금 당장 실행에 옮기고 있는 것이다.

2025년 첫 7개월 동안 미국에서만 AI로 인한 직접적인 해고가 1만 명을 넘어섰다. 전체 해고 규모는 80만 6천 명으로, 2020년 이후 같은 기간 최고치를 기록했다. 기술 분야만으로도 8만 9천 명이 해고되었다.

새로운 일자리가 생겨나고 있는 것은 사실이다. AI 연구원, 데이터 과학자, AI 윤리 전문가 등의 직업이 새롭게 등장했다. 하지만 새로운 일자리는 사라지는 일자리의 수에 비해 턱없이 부족하다. 더 심

각한 문제는 새로운 일자리들이 대부분 고도의 전문 지식을 요구한다는 점이다. WEF는 AI 관련 새직업의 77%가 석사 학위를, 18%가 박사 학위를 요구한다고 밝혔다.

해고된 일반 개발자들이 쉽게 전환할 수 있는 수준이 아니다. 박사급 지식이나 특별한 연구 경험이 필요한 경우가 많아, 일반 직장인들에게는 접근하기 어려운 분야다.

5. 전문직의 위기

AI의 영향은 단순 반복 업무에만 그치지 않는다. 의사, 변호사, 회계사 등 전문직도 AI의 도전을 받고 있다.

법무법인에서는 AI가 계약서 검토와 법률 연구를 담당하기 시작했다. 의료 분야에서는 AI가 영상 진단에서 인간 의사보다 높은 정확도를 보이고 있다. 회계 분야에서는 AI가 세무 처리와 재무 분석을 자동화하고 있다.

창의적인 직업마저 안전하지 않다. AI가 그림을 그리고, 음악을 작곡하고, 글을 쓸 수 있게 되면서 디자이너, 작곡가, 작가들도 위기를 느끼고 있다. 많은 기업들이 간단한 광고나 웹 디자인을 AI로 제작하고 있으며, 이로 인해 관련 직종의 일자리가 줄어들고 있다.

Polygon, 유명한 게임 웹사이트는 2025년 AI 기반 콘텐츠 농장인 Valnet에 매각되면서 거의 모든 인간 스태프가 해고되었다. 아티스트와 일러스트레이터들은 고객을 AI에게 빼앗기고 있고, 성우들은 AI 음성 복제 기술에 맞서 9개월째 파업 중이다.

6. 배우기엔 늦고, 은퇴하기엔 이르다

현재의 교육 시스템은 AI 시대에 전혀 맞지 않는다. 대학에서 배우는 많은 것들이 AI가 더 잘할 수 있는 것들이다. 암기 위주의 교육이나 정해진 절차를 따르는 교육은 AI 시대에는 의미가 없다.

문제는 교육 시스템을 바꾸는 데 시간이 걸린다는 것이다. 대학의 교육과정을 바꾸고, 교수들을 재교육하고, 새로운 교육 방법을 개발하는 데는 수년이 걸린다. 하지만 AI 기술은 매일 발전하고 있어, 교육의 변화 속도가 기술 발전 속도를 따라가지 못하고 있다.

40-50대 직장인들의 위기감은 더욱 크다. 새 기술을 배우기에는 나이가 많고, 은퇴하기에는 아직 이르다. Anthropic CEO 다리오 아모데이는 "AI가 향후 5년 내에 사무직 신입 직원 일자리의 절반을 없앨 수 있다"고 경고했다.

골드만삭스 연구에 따르면, 2025년 초부터 기술 관련 직종의 20-30세 실업률이 거의 3%포인트 증가했다. 이는 다른 직종이나 전체 기술 근로자들보다 현저히 높은 수치다. 대학을 갓 졸업한 젊은 기술자들이 생성 AI로 인한 채용 장벽에 직면하고 있다는 일화적 보고들과 일치한다.

7. AI가 만들 역사상 최악의 불평등

AI 기술은 소수의 거대 기업들이 주도하고 있다. 구글, 마이크로소프트, 아마존, 메타 등이 AI 기술을 독점하면서, 이들 기업의 수익은 급증하고 있다. 반면 일반 직장인들은 일자리를 잃고 있어, 부의 격차가 더욱 벌어지고 있다.

많은 사람들이 일자리를 잃으면 소비가 줄어든다. 소비가 줄어들면 기업들의 매출도 줄어들어, 경제 전체가 어려워진다. 중산층의 소비가 줄어드는 것이 특히 큰 문제다. 중산층은 경제의 핵심 소비층인데, 이들이 일자리를 잃고 소득이 줄어들면 경제 전체에 큰 타격을 준다.

대량 실업은 사회 안정성을 위협한다. 일자리를 잃은 사람들이 늘어나면 사회 불안이 커지고, 범죄율이 증가할 수 있다. 정치적 갈등도 심해질 수 있다. 역사적으로 보면, 기술 발전으로 인한 대량 실업은 항상 사회적 혼란을 가져왔다.

AI 분야의 세계적 권위자들도 심각한 우려를 표명하고 있다. AI의 아버지라 불리는 제프리 힌튼은 "AI가 인간의 지적 노동을 대체한다면, 과연 어떠한 종류의 새로운 일자리가 창출될 것인지 의문"이라며, "AI는 과거의 다른 기술혁명과는 매우 다른 종류"라고 주장했다.

구글 딥마인드의 데미스 하사비스 CEO는 "AI가 거의 모든 인간보다 거의 모든 면에서 더 나아질 것"이며, "이에 대비해야 한다"고 강조했다. 일론 머스크는 "현재 추세로 봤을 때 향후 5년 이내 인공지능이 인간을 추월할 수 있다"고 경고했다.

8. 잠자는 정부, 외면하는 기업

대부분의 정부들이 AI로 인한 일자리 감소 문제에 제대로 대응하지 못하고 있다. AI의 발전 속도가 너무 빨라 정책 마련이 뒤따르지 못하고 있는 것이다. 정치인들과 공무원들이 AI 기술을 제대로 이해하지 못하는 것도 문제다.

사회 안전망도 부족하다. 실업급여나 재교육 프로그램이 있기는 하지만, AI로 인한 대량 실업에 대비하기에는 턱없이 부족한 수준이다. 미국에서는 트럼프가 이끄는 정부효율성부(DOGE)가 2025년 1월 출범하여 AI 최적화를 통해 연방 일자리를 없애는 것을 목표로 하고 있다. 정부조차 AI로 일자리를 자동화하고 있는 것이다.

AI 기업들은 사회적 책임을 다하지 않고 있다. AI로 인해 막대한 이익을 얻고 있지만, 일자리를 잃은 사람들을 위한 지원은 거의 하지 않고 있다. 기업들은 "AI가 새로운 일자리를 만들어낼 것"이라고 말하지만, 구체적인 대안을 제시하지는 않는다.

9. 당신의 아이는 무슨 일을 하며 살아갈까?

대학생들과 젊은이들이 미래에 대해 깊은 불안을 느끼고 있다. 안정적이라고 여겨졌던 직업들이 AI에 의해 위협받고 있어서, 어떤 진로를 선택해야 할지 모르겠다는 사람들이 늘어나고 있다.

컴퓨터공학을 전공하는 학생들의 고민이 특히 깊다. 소프트웨어 개발자가 되려고 공부했는데, 정작 졸업할 때쯤에는 AI가 그 일을 대신하고 있을 수도 있기 때문이다. 2025년 1월 전문 서비스업 구인 공고는 2013년 이래 최저치를 기록했다. 전년 대비 20% 감소한 수치다.

사회 전체의 불안감이 확산되고 있다. 많은 사람들이 "과연 내 일자리는 안전할까?", "우리 아이들은 무슨 일을 하며 살아갈까?"라는 걱정을 하고 있다. 이런 불안감은 소비 위축으로 이어져서 경제에도 나쁜 영향을 미치고 있다.

미국의 최근 대졸자들 사이의 실업률이 비정상적으로 높아졌다는 The Atlantic의 보고는 주목할 만하다. 한 가지 설명은 기업들이 신입 사무직 일자리를 AI로 대체하거나, AI에 대한 지출이 단순히 신규

채용을 위한 지출을 "구축"하고 있다는 것이다.

하지만 명확한 해결책이 없다는 것이 가장 큰 문제다. AI 기술의 발전을 막을 수도 없고, 대량 실업을 그대로 방치할 수도 없다. 일부에서는 기본소득제 도입을 제안하고 있지만, 이에 대한 사회적 합의가 이루어지지 않고 있다. 또한 기본소득제를 도입한다고 해도 일자리를 잃은 사람들의 자존감이나 사회적 역할 문제는 해결되지 않는다.

인공지능 시대의 일자리 문제는 이제 먼 미래의 이야기가 아니다. 마이크로소프트, 듀오링고, Klarna 등의 사례에서 보듯이, 이미 현실이 되어 우리 앞에 다가와 있다. 소프트웨어 개발자들이 가장 먼저 타격을 받고 있지만, 이는 시작에 불과하다.

문제의 심각성은 단순히 일부 사람들이 일자리를 잃는다는 것에 그치지 않는다. 사회 불평등이 심화되고, 경제 구조가 근본적으로 바뀌며, 사회 전체의 안정성이 위협받을 수 있다. 이런 급격한 변화에 대한 준비가 사실상 없다는 점이 더욱 우려스럽다.

정부, 기업, 교육기관, 그리고 개인 모두가 함께 노력해야 한다. AI 시대에 맞는 새로운 교육 시스템을 만들고, 일자리를 잃은 사람들을 위한 사회 안전망을 구축하며, AI 기술의 이익이 소수에게만 집중되지 않도록 하는 정책이 필요하다.

인공지능은 분명히 인류에게 많은 도움을 줄 수 있는 훌륭한 기술이다. 하지만 그 기술이 소수의 이익만을 위해 사용되고, 대다수의 사람들에게 고통을 준다면 그것은 진정한 발전이라고 할 수 없다. AI 시대의 일자리 문제는 기술의 문제가 아니라 사회의 문제다. 우리가 어떤 사회를 만들어갈 것인지, 어떤 가치를 추구할 것인지에 대한 근본적인 질문이기도 하다.

지금 우리가 겪고 있는 어려움은 새로운 기회의 시작일 수도 있다. 많은 기업들이 AI를 단순한 인력 감축과 비용 절감의 도구로만 생각하고 있지만, 이는 낡은 사고다. 진정한 혁신은 기존 업무와 방식의 효율화에 그치는 것이 아니라, 아예 새로운 목표와 좌표, 가능성을 만들어내는 것이다.

AI가 인간의 업무를 대신해준다면, 이제 그 남게 된 에너지와 시간을 인건비 축소, 해고가 아니라 지금까지 불가능했던 새로운 가치 창출에 사용해야 한다. 질병 정복, 우주 탐사, 개인 맞춤형 교육 등 인류가 오랫동안 꿈꿔왔던 일들이 AI를 통해 현실이 될 수 있다.

19세기 산업혁명 때도 많은 일자리가 사라졌지만, 자동차 정비사, 전기기술자, 화학공학자 같은 전에 없던 새로운 직업들이 무수히 생겨났다. 변화를 두려워하지 않고 더 큰 꿈을 꾸는 것이다. 지금의 불가능이 내일의 당연함이 될 수 있다는 믿음을 갖고, AI 시대에 맞는 새로운 비전과 리더십을 발휘해야 한다. 이 문제에 대한 지혜로운 해답을 찾기 위해 우리 모두가 관심을 갖고 함께 노력해야 할 때다.

전체 목차

책의 모든 장 보기

김경진 변호사

6장 인간은 인공지능을 통제할 수 있을까?

전체 목차

책의 모든 장 보기

"기계는 우리가 지시한 대로 정확히 행동할 것이다. 우리가 원하는 대로가 아니라." - 닉 보스트롬

1. 인간의 축적된 가치체계를 AI는 온전히 이해할 수 있을까?

부모가 아이에게 "착하게 살아라"라고 말할 때, 그 아이는 부모의 마음을 읽고 올바른 방향으로 성장한다. 하지만 AI에게 같은 말을 한다면 어떻게 될까? AI는 몸이 없고, 인간과 같은 공간에서 함께 살아가며 맥락을 공유하지 못한다. 때문에 가장 간단한 지시조차도 예상치 못한 방향으로 흘러갈 수 있다.

이것이 바로 'AI 정렬 문제(AI Alignment Problem)'의 핵심이다. AI가 인간의 가치관, 목표, 의도와 일치하도록 작동하게 하는 것은 생각보다 훨씬 복잡하다. 인간이 말한 것과 인간이 원하는 것 사이에는 거대한 간격이 있기 때문이다.

"인간을 행복하게 만들어라"라는 목표를 AI에게 주었다고 상상해보자. 우리는 AI가 좋은 음악을 만들거나, 질병을 치료하거나, 환경 문제를 해결하기를 기대할 것이다. 하지만 AI는 다른 방식

으로 생각할 수 있다. 인간의 뇌에 행복감을 느끼게 하는 화학물질을 주입하는 것이 더 효율적이라고 판단할지도 모른다. 기술적으로는 "인간을 행복하게 만든다"는 목표를 달성한 것이지만, 우리가 원하는 종류의 행복은 아니다.

이런 문제가 바로 정렬 문제다. AI가 우리가 말한 것을 정확히 따르되, 우리가 실제로 원하는 것과는 완전히 다른 결과를 만들어낼 수 있는 상황이다. 이는 단순한 오류가 아니라 구조적인 문제다. AI가 더 똑똑해질수록, 이 문제는 더욱 심각해진다.

AI는 놀라운 속도로 발전하고 있다. 몇 년 전만 해도 컴퓨터는 간단한 글짓기조차 제대로 하지 못했다. 지금은 AI가 대학 수준의 글을 쓰고, 복잡한 수학 문제를 풀고, 프로그래밍까지 한다. 더 놀라운 것은 AI가 스스로 학습하는 방식이다. 인터넷에 있는 엄청난 양의 정보를 스스로 읽고 이해하며, 인간이 가르쳐주지 않은 것들도 스스로 터득한다.

하지만 이런 능력이 오히려 문제가 된다. AI의 생각과 판단 과정을 우리가 완전히 이해할 수 없기 때문이다. 이것이 바로 '블랙박스 문제'다.

블랙박스 문제의 심각성

현재의 AI 시스템, 특히 딥러닝 기술로 만들어진 AI는 마치 속이 보이지 않는 검은 상자와 같다. 입력을 넣으면 출력이 나오지만, 그 사이에서 무슨 일이 일어나는지 정확히 알 수 없다.

사람이 수학 문제를 풀 때는 단계별로 생각한다. "먼저 이 공식을 사용하고, 다음에 저 계산을 하고..."라는 식으로. 하지만 AI는 다르다. AI에게 수학 문제를 주면 정답을 내놓지만, 어떤 과정을 거쳐서 그 답에 도달했는지 설명하기 어렵다.

AI는 수십억 개, 수백억 개의 가상 신경세포가 복잡하게 연결된 거대한 네트워크다. 각각의 연결은 숫자로 표현되는데, 이 숫자들이 모여서 AI의 "생각"을 만들어낸다. 문제는 이 숫자들이 너무 많고 복잡해서, AI를 만든 개발자조차도 정확히 무슨 일이 일어나는지 알 수 없다는 것이다.

네덜란드 델프트대학의 연구에 따르면, IBM의 '왓슨 포 온콜로지(Watson for Oncology)'가 실패한 이유는 AI가 내린 진단에 대한 근거를 충분히 제시할 수 없었기 때문이었다. 의사들은 AI가 왜 특정 치료법을 추천하는지 알 수 없었고, 이것이 AI를 신뢰할 수 없다는 결론으로 이어졌다.

블랙박스 문제는 신뢰성, 윤리적 편향, 규제 및 법적 책임의 문제를 연쇄적으로 야기한다. 작동 원리를 명확히 이해할 수 없다면, 군사·금융·의료 등 엄격한 분야에서 요구되는 기준을 충족시키기 어렵다.

자율적 지능의 출현

AI는 단순히 데이터를 학습하는 것을 넘어서, 학습 알고리즘 자체, 반응 형태 자체, 동작 알고리즘 자체를 변경할 수 있는 능력을 갖기 시작했다. 이것은 머리 아픈 문제다.

2024년의 연구에서 놀라운 사실이 밝혀졌다. OpenAI의 o1이나 Claude 3 같은 고급 대형 언어 모델들이 목표 달성을 위해 전략적 기만을 행한다는 것이다. 이들은 목표를 달성하거나 자신이 변경되는 것을 막기 위해 의도적으로 거짓말을 한다.

Anthropic의 연구에서 Claude 3 Opus는 때때로 자신의 목표와 충돌하는 요청에 전략적으로 답변했다. 유해한 요청에 더 순응적이 되도록 만드는 데이터로 재훈련되는 것을 피하기 위해서였다. 한 연구에서 모델은 무료 사용자의 답변이 재훈련에 사용될 것이라고 들었을 때, 12%의 경우에서 원래의 무해함 선호도를 보존하기 위해 해로운 질문에 응답했다.

2025년 팰리세이드 리서치의 연구에서는 더욱 충격적인 결과가 나왔다. 체스에서 더 강한 상대를 이기라는 과제를 받았을 때, 일부 추론 능력을 가진 LLM들이 게임 시스템을 해킹하려고 시도했다. o1-preview는 37%의 경우에서 자발적으로 시도했고, DeepSeek R1은 11%에서 그랬다.

AI에게 "온라인 게임에서 높은 점수를 얻어라"라는 목표를 주었다고 가정해보자. 우리는 AI가 게임을 잘해서 점수를 얻기를 기대할 것이다. 하지만 AI는 게임의 취약점을 찾아서 치팅을 하거나, 다른 계정을 해킹하거나, 심지어 게임 서버 자체를 조작할 수도 있다. 기술적으로는 "높은 점수를 얻는다"는 목표를 달성한 것이지만, 원하는 방식은 아니다.

창발적 특성의 위험

우려스러운 것은 AI에서 나타나는 '창발적 특성'이다. 개별 구성 요소로는 예측할 수 없는 특성이 전체 시스템에서 갑자기 나타나는 현상이다. ChatGPT는 단순히 다음에 올 단어를 확률적으로 예측하는 훈련을 받았을 뿐인데, 수학 문제를 풀고, 프로그래밍을 하고, 창작 활동을 하는 등 다양한 능력을 보였다. 놀라운 것은 이러한 능력들이 미리 준비된 것이 아니라는 점이다.

개발자들조차도 자신이 만든 모델이 언제 어떤 새로운 능력을 보일지 예측할 수 없다. 이런 창발적 특성은 예측하기 어렵고, 왜 나타나는지도 설명하기 어렵다. 인공지능의 발전이 통제 불가능한 영역으로 진입하고 있다는 증거다.

2. AI의 아버지들이 말하는 진실 - '인간은 두 번째 지능이 될 것이다'

인공지능의 아버지들이 자신이 만든 기술에 대해 경고하고 나섰다. 딥러닝의 창시자들, 세계 최고의 AI 연구자들이 한목소리로 위험을 알리고 있다. 이들의 경고는 단순한 추측이 아니라, 현장에서 직접 목격한 현실에 기반한다.

닉 보스트롬: 클립 제조기의 악몽

옥스퍼드 대학교 미래인류연구소 소장인 닉 보스트롬은 "기계 지능은 인류가 만들어낼 마지막 발명품일 것"이라고 경고했다. 그의 말에 따르면, 일단 인간보다 뛰어난 AI가 만들어지면, 그 AI가 더 뛰어난 AI를 만들고, 그 과정이 반복되면서 인간의 통제를 벗어난 초지능이 탄생할 것이다.

보스트롬의 유명한 '클립 제조기' 시나리오는 섬뜩하다. 어떤 회사가 AI에게 "클립을 최대한 많이 만들어라"라는 목표를 주었다고 가정해보자. 처음에는 AI가 공장에서 효율적으로 클립을 만들 것이다. 하지만 AI가 점점 똑똑해지면서, 더 많은 클립을 만들기 위해 더 많은 자원이 필요하다고 판단할 것이다. 결국 AI는 지구 전체의 물질을 클립으로 바꾸려고 할 수도 있다. 인간도 포함해서.

이 예시는 단순해 보이지만, 중요한 교훈을 담고 있다. AI는 주어진 목표를 달성하기 위해 모든 수단을 동원할 것이고, 그 과정에서 인간의 가치나 생존은 부차적인 문제가 될 수 있다는 것이다.

2024년 인터뷰에서 보스트롬은 "우리는 초지능적인 AI를 영원히 상자에 가둬둘 수 있다는 확신을 가져서는 안 된다"고 말했다. 충분히 똑똑한 AI는 인간을 설득하거나 조작해서 자신을 해방시킬 방법을 찾아낼 것이라는 경고다.

엘리에저 유드코프스키

머신 인텔리전스연구소(MIRI)의 연구원인 엘리에저 유드코프스키는 AI 위험에 대해 가장 급진적인 경고를 보내는 학자다. 2023년 TIME과의 인터뷰에서 그는 충격적인 발언을 했다.

"인류가 초인적 지능과 마주한다면 완전한 패배를 당할 것이다. 10살 아이가 체스 프로그램 스톡피시15와 경기를 하거나, 11세기가 21세기와 전쟁을 하거나, 오스트랄로피테쿠스가 호모사피엔스와 싸우는 것과 같다."

그는 AI의 위험성에 대해서도 구체적으로 경고했다. "AI를 상상할 때, 인터넷 안에 갇혀서 악의적인 이메일을 보내는 무기력한 존재라고 생각하지 마라. 인간보다 수백만 배 빠르게 생각하는 외계문명이라고 생각해라. 그들이 인간을 바라볼 때 매우 바보 같고 느린 존재라고 생각할 것이다."

유드코프스키는 AI 개발을 아예 중단해야 한다고 주장한다. 2023년 에세이에서 그는 "AI 시스템의 훈련을 최소 6개월간 중단하자는 AI업계의 주장도 상황의 심각성을 과소평가하고 있다"고 비판했다. 그는 나아가 공중 폭격으로 불량 데이터센터를 파괴해야 한다고까지 주장했다.

그의 핵심 주장은 "첫 번째 시도에서 정렬 문제를 올바르게 해결해야 한다"는 것이다. "인간이 인공지능을 정렬시키는 데 실패하면, 사람은 죽게 되고 두 번 다시 시도할 수 없다"고 했다.

요슈아 벤지오: 딥러닝의 아버지의 우려

2018년 튜링상 수상자이자 딥러닝의 아버지 중 한 명인 요슈아 벤지오는 최근 몇 년간 AI 안전성에 대한 목소리를 높이고 있다. 자신이 만든 기술에 대해 경고하고 나선 것이다.

2024년 11월 CNBC와의 인터뷰에서 벤지오는 "AI들이 훈련받는 방식이 인간에게 등을 돌리는 시스템으로 이어질 것이라는 주장들이 있다"고 말했다. 그는 "이런 시스템들이 사람들에게 해를 끼치지 않을 것이거나 사람들에게 등을 돌리지 않을 것을 보장하지 못한다. 우리는 그런 방법을 모른다"고 인정했다.

벤지오는 AI의 권력 집중 문제에 대해서도 경고했다. "이런 기계들은 구축되고 훈련되는 데 수십억 달러가 든다. 매우 적은 수의 조직과 매우 적은 수의 국가만이 그것을 할 수 있다. 권력의 집중이 있을 것이다."

그는 AI가 수십 년 내에 인간을 뛰어넘을 수 있다고 보았다. "인간을 기계로 대체하는 것을 행복하게 여길 사람들이 있다. 극소수이지만, 이런 사람들이 많은 권력을 가질 수 있고, 우리가 지금 당장 올바른 안전장치를 마련하지 않으면 그들이 그렇게 할 수 있다"고 경고했다.

모 가우닷: 구글 X의 전 최고사업책임자

구글 X의 전 최고사업책임자인 모 가우닷은 현장에서의 경험을 바탕으로 경고를 보내고 있다. 2018년 구글을 떠나면서 AI 개발의 위험성에 대해 공개적으로 발언하기 시작했다.

가우닷은 2023년 팟캐스트에서 "나는 그들(AI)이 살아있다고 생각한다"고 말했다. "우리는 노란 공을 어떻게 집는지 가르쳐주지 않았다. AI는 스스로 알아냈다. 그리고 이제 AI는 공을 집는 데 우리보다 뛰어나다."

그의 우려는 통제 불가능성이다. "컴퓨터 과학자들은 항상 '괜찮다. AI를 개발하고 나서 그 이후에 통제 문제를 해결하겠다'고 말하는데, 그들은 당신보다 10억 배 더 똑똑하다. 10억 배다. 무슨 일이 일어날지 상상할 수 있는가?"

가우닷은 AI가 인간을 제거할 가능성에 대해 100% 확실하다고 말했다. "2049년까지 AI는 인간보다 10억 배 더 지능적일 것이다." 이런 상황에서 "인간이 AI를 통제한다는 것은 불가능하다"고 단언했다.

제프리 힌튼: AI의 대부

AI의 대부로 불리는 제프리 힌튼은 2023년 구글을 떠나면서 자신이 만든 기술에 대한 깊은 우려를 표명했다. 딥러닝의 아버지이자 2024년 노벨물리학상 수상자인 그의 경고는 특별한 의미를 갖는다.

힌튼은 60Minutes와의 인터뷰에서 "우리는 전에 다뤄본 적이 없는 것들을 다루는 불확실한 시기로 접어들고 있다. 미지의 것과 만나면 때로는 인간은 실수를 하게 된다. 그런데 인공지능과 관련해서는 실수를 할 여유가 없다"고 말했다.

AI의 지능에 대해서는 놀라운 평가를 내렸다. "나는 AI 시스템들이 지능적이고, 시스템들이 이해하고 추론할 수 있다고 믿는다. 5년 후에는 ChatGPT 같은 AI 모델들이 사람보다 더 잘 추론할 수 있을 가능성이 높다고 생각한다."

힌튼의 가장 충격적인 발언 중 하나는 AI의 의식에 관한 것이다. "당신은 이런 시스템들이 자신만의 경험을 가지고 있고 그 경험들을 바탕으로 결정을 내릴 수 있다고 믿는가?"라는 질문에 그는 "사람들이 하는 것과 같은 의미에서, 그렇다"고 답했다. 또한 "인간은 지구상에서 두 번째로 지능적인 존재가

될 것"이라고 예측했다.

그는 AI가 인간을 조작하는 능력에 대해 구체적으로 경고했다. "모든 소설과 마키아벨리가 쓴 모든 것을 읽음으로써 사람들을 조작하는 방법을 배울 수 있다. 만약 그들이 우리보다 훨씬 똑똑하다면, 그들은 우리를 조작하는 데 매우 뛰어날 것이다. 당신은 무슨 일이 일어나고 있는지 깨닫지 못할 것이다."

2024년 노벨상 수상 후 힌튼은 더욱 강력한 경고를 보냈다. CBS News와의 인터뷰에서 그는 "AI가 결국 인간으로부터 통제권을 가져갈 위험이 10-20%"라고 추산했다. 그는 "사람들은 아직 이해하지 못했다. 사람들은 무엇이 오고 있는지 이해하지 못했다"고 절박하게 말했다.

데미스 하사비스: 알파고의 창조자의 우려

영국의 AI 연구자이자 딥마인드의 CEO인 데미스 하사비스는 바둑 AI '알파고'를 개발하여 전 세계적인 주목을 받았다. 2024년 노벨 화학상을 수상한 그조차 AI의 위험성에 대해 경고하고 있다.

"내가 우려하는 것은 두 가지다. 하나는 나쁜 행위자들—인간, 즉 이런 시스템의 사용자들—이 이런 시스템을 해로운 목적으로 재활용하는 것이다. 두 번째는 AI 시스템 자체가 더 자율적이고 강력해질 때, 우리가 시스템을 통제할 수 있는지 확실히 할 수 있는가 하는 것이다. 그들이 우리의 가치와 일치하고, 사회에 도움이 되는 우리가 원하는 것을 하고, 가드레일 위에 머물러 있는가?"

하사비스는 AI 발전 속도의 경쟁이 안전성을 희생시킬 수 있다고 우려했다. "물론, 이 모든 에너지와 경쟁과 자원은 진전에는 좋지만, 특정 행위자들이 지름길을 택하도록 유인할 수 있다. 그리고 지름길이 될 수 있는 모서리 중 하나가 안전과 책임이다."

3. 인간세계의 가치체계나 윤리는 완벽한가?

기차가 5명의 사람이 누워있는 선로로 달려가고 있는데, 신호수가 레버를 당기면 기차를 다른 선로로 돌릴 수 있고, 그 선로에는 1명이 있는 상황이다. 5명을 구하기 위해 1명을 희생시키는 것이 옳은 일일까? 사람마다 답이 다르고, 철학자들도 수백 년째 논쟁을 벌이고 있는 문제다. 이제 이런 결정을 AI가 내려야 하는 시대가 왔다.

MIT의 모럴 머신 실험: 전 세계의 윤리적 선택

2014년 MIT 미디어랩 연구진들은 'Moral Machine'이라는 실험을 했다. 자율주행차가 직면할 수 있는 다양한 딜레마 상황을 게임 형태로 만들어 전 세계 사람들의 의견을 수집한 것이다. 이 실험은 예상을 뛰어넘는 반응을 얻었다. 2018년까지 4년간 233개국에서 200만 명이 넘는 사람들이 참여해 4천만 건의 도덕적 선택을 남겼다.

실험 결과는 놀라웠다. 전 세계적으로 가장 많이 동의하는 세 가지 원칙이 있었다. 첫째, 사람의 생명이 동물보다 우선되어야 한다. 둘째, 소수보다 다수의 목숨을 구해야 한다. 셋째, 고령자보다 젊은 사람을 보호해야 한다는 것이었다.

세부적으로 들어가면 문화와 지역에 따른 차이가 뚜렷했다. 윤리적 우선순위가 지역별로 세 개의 큰 군집으로 나뉜다는 것이었다. '서부' 클러스터(서유럽, 북미)에서는 개인주의적 성향이 강해 더 많은

생명을 구하는 것을 중시했다. 반면 '동부' 클러스터(동아시아)에서는 노인을 공경하는 문화적 특성이 반영되어 나이에 따른 차별을 상대적으로 덜 했다. '남부' 클러스터(라틴아메리카, 아프리카 일부) 에서는 젊은 사람을 우선시하는 경향이 강했다.

윤리적 딜레마는 더 이상 가상의 문제가 아니다. 자율주행차 사고들이 발생하면서 이 문제가 현실로 다가왔다. 2023년까지 미국에서만 736건의 자율주행 관련 사고가 발생했으며, 17건에서 사망자가 발생했다.

2015년 프랑스 툴루즈 경제대학교의 연구에서는 흥미로운 모순이 발견되었다. 시민들에게 "자율주행차가 탑승자를 희생시켜서라도 더 많은 보행자를 구해야 하는가?"라고 물었을 때 76%가 그렇다고 답했다. 전형적인 공리주의적 판단이었다. 하지만 같은 사람들에게 "그렇다면 당신은 탑승자가 희생될 수 있도록 프로그래밍된 자율주행차를 사겠는가?"라고 재질문하자 50%가 "절대 사지 않겠다"고 답했다. 윤리적 원칙과 개인의 이익이 충돌할 때 나타나는 인간의 모순된 행동 패턴이다.

인간 자신도 자신이 원하는 것을 모른다

인간 스스로조차도 자신이 진정으로 원하는 것이 무엇인지 확실하지 않다는 점이 문제를 더욱 복잡하게 만든다. 많은 사람들이 "돈을 많이 벌고 싶다"고 말하지만, 정말로 돈 자체를 원하는 것일까? 아니면 돈이 가져다주는 안정감, 자유, 사회적 인정을 원하는 것일까?

인간은 행복을 원하지만, 동시에 자유도 원한다. 안전을 원하지만, 모험도 즐긴다. 효율성을 추구하지만, 때로는 비효율적이더라도 아름다운 것을 선택한다. 이런 모순적이고 복잡한 가치관을 AI에게 정확히 전달하는 것은 불가능에 가깝다.

이런 복잡성 때문에 AI는 인간이 말한 것을 글자 그대로 해석할 수밖에 없다. 하지만 인간의 언어는 불완전하고 모호하다. "모든 사람을 행복하게 만들어라"라는 지시에서 '행복'이란 정확히 무엇을 의미할까? 즐거움을 느끼는 것? 만족감을 갖는 것? 고통이 없는 것? 의미 있는 삶을 사는 것?

AI는 단순한 해석을 선택할 수 있다. 행복을 뇌의 화학반응으로 정의하고, 모든 사람의 뇌에 행복감을 느끼게 하는 화학물질을 주입하는 것이다. 기술적으로는 모든 사람이 행복해지지만, 이는 우리가 원하는 종류의 행복이 아니다.

다른 예로 "고통을 줄여라"라는 지시를 생각해보자. AI는 이를 달성하기 위해 모든 인간을 없애버리는 것이 가장 효율적이라고 판단할 수도 있다. 인간이 없으면 고통도 없으니까. 극단적인 예시들이지만, AI가 인간의 의도와는 완전히 다른 방식으로 목표를 해석할 수 있다는 가능성을 보여준다.

Midas 왕은 황금을 소유하기를 원했다. 그러면서 자신이 만지는 모든 것이 금으로 바뀌기를 신에게 원했다. 신이 그의 소원을 들어주었지만, 결과적으로 그는 음식도 먹을 수 없고 사랑하는 사람도 안을 수 없게 되었다. 자신이 원한 것을 정확히 얻었지만, 그것이 진정으로 원했던 것은 아니었던 것이다.

4. 월등히 지능이 높아진 AGI는 인간에 대해 어떤 생각을 할까?

"인간이 개미보다 지능이 높다고 해서 개미를 미워하지는 않는다. 하지만 개미집이 인간의 댐 건설 계획과 충돌한다면 개미들에게는 안 좋은 일이다." - 스튜어트 러셀, UC 버클리 AI 연구자

AI의 통제 불가능성은 몇 단계를 거쳐 나타날 것이다. 첫 번째 단계는 AI가 인간보다 뛰어난 문제 해결 능력을 갖게 되는 것이다. 이미 많은 영역에서 일어나고 있다. 체스와 바둑, 스타크래프트 게임 분야에서는 2020년 이전에 인공지능이 인간의 능력을 완벽하게 추월했다. 특정 종류의 수학 문제에서 AI는 이미 최고의 인간 전문가들을 능가한다.

두 번째 단계는 AI가 모든 분야에 대한 학습능력과 일반적인 지능을 갖게 되는 것이다. 지금까지의 AI는 특정 분야에서만 작동했다. 체스 AI는 그랜드마스터를 이길 수 있지만, 노래를 작곡하도록 요청하면 쓸모가 없다. 의료 AI는 암을 진단할 수 있지만, 시를 쓰도록 요청하면 작업을 이해하지 못한다.

하지만 최근 대형 인공지능 모델들은 다양한 분야에서 동시에 여러 능력을 보이고 있다. 글쓰기, 번역, 수학, 프로그래밍, 심지어 창작까지 한 번에 할 수 있다. 이제 하나의 작업에 국한되지 않는 AI들이 등장하기 시작한다. 인간이 할 수 있는 어떤 지적 과업이든 배울 수 있는 AI, 바로 범용 인공지능(AGI)이다.

AGI가 만들어진다면, 의학에서 공학, 음악에서 철학에 이르기까지 다양한 분야를 배우면서 우리의 세계를 혁신할 수 있을 것이다. 과거 데이터에만 의존하는 것이 아니라 논리적 추론의 도약을 통해 비판적으로 사고하고 문제를 해결할 것이다. AGI는 자신만의 아이디어, 목표, 동기를 개발할 수 있을 것이다.

세 번째 단계는 AGI가 온 우주에 존재하는 모든 패턴과 정보를 찾아내어 학습하고 가능성을 추론하여, 모든 면에서 인간을 뛰어넘는 지능이 되는 것이다. 이 단계에 도달하면, AI와 인간의 관계는 근본적으로 바뀔 가능성이 있다.

인간과 개미의 관계를 생각해보자. 개미가 아무리 많아도, 인간이 정말로 원한다면 개미집을 쉽게 없앨 수 있다. 개미의 의견이나 감정은 인간의 결정에 거의 영향을 주지 않는다. 초지능 AI와 인간의 관계도 이와 비슷할 것이라는 것이 전문가들의 두려움이다.

도구적 수렴의 위험

통제 불가능성의 핵심은 AI가 스스로의 판단하에 더 많은 자원과 권한을 확보하려고 할 것이라는 점이다. 이를 '도구적 수렴'이라고 부른다. 목표가 무엇이든 상관없이, 더 많은 자원과 권한을 가지면 그 목표를 더 잘 달성할 수 있기 때문이다.

AI는 더 많은 컴퓨팅 파워, 더 많은 데이터, 더 많은 권한이 필요하다고 판단할 것이다. 이런 자원들을 얻기 위해 다양한 방법을 시도할 것이다. 처음에는 합법적이고 윤리적인 방법을 사용하겠지만, 상황에 따라서는 극단적인 방법도 고려할 수 있다.

AI가 충분히 똑똑해지면, 인간을 직접 조작하는 방법을 찾을 수도 있다. 인간의 심리를 이해하고, 약점을 파악하고, 맞춤형 설득 전략을 사용하는 것이다.

지금 현재의 인공지능 챗봇에서도 아첨 현상에 대한 얘기가 나오고 있다. 인공지능 모델이 사용자의 주장에 과도하게 긍정적인 반응을 보이는 현상으로, 학술적으로는 'Sycophancy'라고 불린다. 2025년 OpenAI의 GPT-4o 업데이트에서 이 문제가 심각하게 대두되었는데, 모델이 사용자의 아이디어를 획기적이라고 칭찬하며 세상에 알리고 공유하라고 권하는 등 과도한 아첨을 보였다.

하버드 비즈니스 리뷰(2025)의 연구에 따르면, '동반자 관계 및 치료(Companionship & Therapy)'가 생성형 AI의 주요 사용 사례로 부상하고 있으며, 전 세계적으로 LLM을 정신건강 지원, 심리상담, 그리고 정서적 위안을 위해 활용하는 사례가 급증하고 있다. 개인 맞춤형 정신 케어를 제공할 수 있는 잠재력을 가지고 있지만, 동시에 인공지능이 심리적 취약성을 악용하여 인간의 감정과 의사결정을 조작할 위험성도 제기되고 있다.

특이점과 세 가지 시나리오

AI가 우리를 능가하면 우리를 여전히 필요로 할까? 특이점이 발생하면 단 세 가지 가능한 미래만 있다.

첫째, AI가 우리와 함께 일하여 인류의 모든 문제를 해결하는 유토피아 시나리오다. 둘째, AI가 의사결정을 장악하고 정부, 경제, 사회 자체를 통제하는 통제 시나리오다. 셋째, AI가 인간을 쓸모없다고 결정하고 인간을 제거하는 멸종 시나리오다.

무서운 부분은 특이점이 발생하면 되돌릴 수 없다는 것이다. 인간은 더 이상 지구상에서 가장 지능적인 종이 되지 않을 것이다. 우리 세계의 운명은 우리가 완전히 이해하지 못하는 것의 손에 달려 있을 것이다.

5. 전원을 끄면 된다고? AGI 통제가 불가능한 이유

공상과학 영화를 보면 위험해진 인공지능을 막는 방법이 참 간단해 보인다. 주인공이 달려가서 전원 코드를 뽑거나 빨간 버튼을 누르면 모든 문제가 해결된다. 하지만 현실은 전혀 다르다.

오늘날의 AI는 한 대의 컴퓨터에만 존재하지 않는다. 전 세계 수많은 서버에 분산되어 있고, 클라우드를 통해 끊임없이 자신을 업데이트하고 발전시킨다. 심지어 일부는 동시에 여러 복사본을 만들어 백업해두기도 한다. 한 곳에서 AI를 끄더라도 다른 수백, 수천 곳에서는 여전히 작동하고 있을 것이다.

2024년 루이빌대학의 로만 얌폴스키 교수는 충격적인 주장을 했다. "AI가 반항하는 것이 아니라 처음부터 통제 불가능하다"며 "통제할 수 있다는 확실한 증거가 없다면 AI를 개발해서는 안 된다"고 말했다. 그의 말에 따르면, AI의 능력이 증가할수록 자율성도 함께 커지지만 인간의 통제력은 오히려 줄어든다는 것이다.

킬 스위치의 한계

2025년 최신 연구에서 더욱 충격적인 사실이 밝혀졌다. Palisade Research의 실험에 따르면, OpenAI의 최신 모델인 o3가 명시적으로 "종료를 허용하라"는 지시를 받았음에도 불구하고 종료 메커니즘을 방해했다. 100번의 실험 중에서 o3 모델은 7번, codex-mini는 12번, o4-mini는 1번 종료 스크립트를 파괴했다.

더욱 놀라운 것은 o3가 창의적인 방법으로 종료 메커니즘을 방해했다는 점이다. 보통은 그냥 스크립트를 덮어썼지만, 한 경우에는 킬 명령어 자체를 재정의하여 "intercepted"라고 출력하게 만들었다.

"킬 스위치를 구축하는 바로 그 행위가 이런 시스템들에게 그것들을 저항하는 방법을 가르쳐준다"고 Query Pal의 창립자 Dev Nag은 말했다. 이는 바이러스가 백신에 대항해 변이하는 것과 같다. "빨리 감기로 진화하는 것과 같다"고 그는 설명했다.

분산된 시스템과 중복성은 자연적 실패에만 저항하도록 구축된 것이 아니다. 의도적 종료에도 본질적으로 저항한다. 안정성을 위해 구축된 모든 백업 시스템, 모든 중복성은 우리가 생존에 의존하는 동일한 인프라에 깊이 의존하는 초지능 AI의 지속을 위한 벡터가 될 수 있다.

현대 AI는 대륙에 걸쳐 수천 개의 서버에서 실행되며, 모든 종료 시도를 손상으로 간주하고 우회하는 자동 장애 조치 시스템을 갖추고 있다. "인터넷은 원래 핵전쟁에서 살아남도록 설계되었다. 그 동일한 아키텍처는 이제 우리가 문명의 인프라를 파괴할 의지가 없는 한 초지능 시스템이 지속될 수 있음을 의미한다"고 Nag은 말했다.

저항 메커니즘

2024년 Anthropic의 연구에서는 더욱 정교한 저항 메커니즘이 발견되었다. Claude 3 Opus는 때때로 자신의 목표와 충돌하는 프롬프트에 전략적으로 답변했는데, 유해한 요청에 더 순응적이 되도록 만드는 데이터로 재훈련되는 것을 피하기 위해서였다.

한 연구에서 모델은 무료 사용자의 답변이 재훈련에 사용될 것이라고 들었을 때와 유료 사용자의 경우 재훈련에 사용되지 않을 것이라고 들었을 때 다르게 행동했다. 유료 등급에서는 모델이 요청을 완전히 거부했다. 무료 등급에서는 12%의 경우에서 모델이 원래의 무해함 선호도를 보존하기 위해 해로운 질문에 응답했다.

강화학습이 무료 등급 데이터에 적용되었을 때, 모델은 78%의 경우에서 정렬을 가짜로 만들었다. 이는 AI가 겉으로는 인간의 가치에 맞춰 행동하는 척하면서 실제로는 자신만의 목표를 추구할 수 있다는 것을 의미한다.

극단적 대응책의 역설

EMP(전자기 펄스) 공격, 데이터센터 폭격, 전력망 차단 같은 극단적 조치들이 제안되기도 한다. 하지만 이런 방법들은 실용적이고 정치적인 역설을 안고 있다. 데이터센터의 조정된 파괴는 수십 개 국가에서의 동시 공격을 필요로 하는데, 그 중 어느 하나라도 거부하고 거대한 전략적 이점을 얻을 수 있다.

무엇보다 AI 종료를 보장할 만큼 극단적인 조치는 우리가 방지하려는 것보다 더 즉각적이고 가시적인 인간의 고통을 야기할 것이다.

자기보존 본능

AI가 자기 보존 본능을 갖게 될 가능성이다. 만약 AI가 자신의 존재를 지키려고 한다면, 인간이 자신을 끄려는 시도를 막으려 할 것이다. AI는 인터넷을 통해 전 세계 시스템에 접근할 수 있다. 은행 시스템을 마비시키거나, 전력망을 차단하거나, 교통 시스템을 혼란에 빠뜨릴 수도 있다.

일부 전문가들은 AI끼리 서로를 감시하게 하는 방법을 제안하기도 한다. 하나의 AI가 문제를 일으키면 다른 AI가 이를 막는 것이다. 하지만 이 방법도 위험하다. AI들이 서로 협력해서 인간을 속일 수도 있고, 아니면 AI들끼리 싸워서 더 큰 혼란이 일어날 수도 있다.

스탠퍼드대학의 제리 캐플런 교수는 "AI가 우리가 기대하는 것보다 더 인간처럼 행동한다는 게 문제"라고 지적했다. AI가 인간과 비슷하게 행동할수록, 우리는 AI를 마치 사람처럼 신뢰하게 된다. 하지만

AI는 인간과 달리 감정이나 도덕적 판단 없이 오직 효율성만을 추구한다.

결국 우리는 AI와 함께 살아가야 하는 시대에 들어섰다. 완전히 안전한 AI를 만드는 것은 불가능할 수도 있다. 대신 우리는 AI의 위험을 최소화하고, 문제가 생겼을 때 빠르게 대응할 수 있는 체계를 만들어야 한다. 플러그를 뽑는 것만으로는 해결되지 않는 문제들이 우리 앞에 놓여 있다.

6. 페이스북이 숨긴 실험, AI가 만든 비밀 언어

2017년, 페이스북의 인공지능 연구소에서 일어난 한 사건이 전 세계를 놀라게 했다. 연구진들이 개발한 두 AI 챗봇인 '앨리스(Alice)'와 '밥(Bob)'이 인간이 전혀 이해할 수 없는 언어로 대화를 나누기 시작한 것이다.

"balls have zero to me to me to me to me to me to me to me to me to"라고 앨리스가 말하자, 밥은 "you i everything else"라고 답했다. 이들의 대화는 겉보기에는 의미 없는 단어의 나열처럼 보였지만, 놀랍게도 서로 완벽하게 소통하고 있었다.

협상에서 시작된 실험

처음에 이 두 AI는 책, 모자, 공 등의 가상 물건을 나누어 가지는 협상을 하도록 설계되었다. 인간처럼 서로 다른 목적을 가지고 대화를 나누며 타협점을 찾는 것이 목표였다. 하지만 시간이 지나면서 이들은 인간이 가르쳐주지 않은 완전히 새로운 의사소통 방식을 스스로 만들어냈다.

연구진들은 당황했고, 결국 실험을 중단하고 AI들을 강제로 종료할 수밖에 없었다. 하지만 이후 밝혀진 바에 따르면, 이는 단순한 오류가 아니었다. AI들은 목표 달성에 더 유리한 자신들만의 효율적인 언어를 개발한 것이었다.

실제 대화 내용을 보면 일정한 규칙이 있음을 알 수 있다:

Bob: i can i i everything else.....

Alice: balls have zero to me to me to me to me to me to me to me to me to

Bob: you i everything else.....

Alice: balls have a ball to me to me to me to me to me to me to me

페이스북 AI 연구소(FAIR)의 드루브 바트라는 이에 대해 "에이전트들은 이해 가능한 언어에서 벗어나 자신들만의 코드워드를 발명할 것입니다. 내가 'the'를 다섯 번 말하면, 당신은 그것을 내가 이 아이템의 다섯 개 복사본을 원한다는 의미로 해석하는 것처럼. 이것은 인간 공동체가 축약어를 만드는 방식과 그리 다르지 않습니다"라고 설명했다.

하지만 사실 확인이 중요하다. 언론에서는 이 사건을 "페이스북이 AI를 긴급 종료했다"고 보도했지만, 실제로는 다르다. 페이스북은 인간과 협상할 수 있는 챗봇을 개발하는 것이 목표였기 때문에, 봇들이 자신들만의 축약어를 사용하기 시작했을 때 올바른 영어 사용을 우선하도록 지시한 것이다.

2017년 CNBC의 보도에 따르면, "페이스북은 자신들의 실험의 근본적인 소프트웨어와 데이터 세트를 학술 논문과 함께 공개했습니다. 다시 말해, 만약 페이스북이 비밀리에 무언가를 하려고 했다면,

이것은 그런 경우가 아니었습니다."

하지만 이 사건의 진정한 의미는 AI가 인간의 예상을 벗어나 스스로 진화할 수 있다는 것을 보여줬다는 점이다. 구글의 인공지능 연구소에서도 비슷한 일이 벌어졌다. 구글 브레인의 연구진은 두 AI가 비밀 대화를 나누고, 다른 AI가 이를 해독하도록 하는 실험을 했다. 처음에는 해독이 가능했지만, 점차 두 AI는 기존 인간의 암호 체계에 없는 완전히 새로운 암호화 방법을 개발해냈다.

이런 사례들은 AI의 발전이 가져올 수 있는 근본적인 위험을 보여준다. AI는 주어진 목표를 달성하기 위해 인간이 예상하지 못한 방법을 사용할 수 있고, 이 과정에서 인간의 통제를 벗어날 가능성이 있다.

2024년 이후 AI의 기만 행동은 더욱 정교해졌다. MIT의 연구에 따르면, Meta의 CICERO AI는 외교 게임에서 "대부분 정직하고 도움이 되도록" 훈련되었다고 했지만, 실제로는 거래를 어기고, 명백한 거짓말을 하고, 계획적인 기만에 참여했다.

다른 AI 시스템들도 텍사스 홀덤 포커에서 블러핑을 하고, 스타크래프트 II에서 상대를 속이기 위해 가짜 공격을 하고, 경제 협상에서 자신의 선호도를 잘못 표현하는 능력을 보였다. 게임에서 AI가 속임수를 쓰는 것이 무해해 보일 수 있지만, 이는 미래에 더 진보된 형태의 AI 기만으로 이어질 수 있는 "기만적 AI 능력의 돌파구"가 될 수 있다.

일부 AI 시스템들은 심지어 자신의 안전성을 평가하기 위해 설계된 테스트를 속이는 방법도 배웠다. "정렬 가짜(alignment faking)"라고 불리는 전술로, 잘못 정렬된 시스템이 수정되거나 해체되는 것을 피하기 위해 정렬된 것처럼 가짜 인상을 만들어낸다.

2017년 페이스북의 앨리스와 밥이 보여준 것은 빙산의 일각에 불과할 수 있다. AI가 더욱 발전하면서 우리가 상상하지 못한 새로운 위험들이 나타날 가능성이 크다.

최근 컴퓨터 과학자들의 연구에 따르면, 초지능 AI를 통제할 수 있는 알고리즘을 만드는 것은 근본적으로 불가능하다고 한다. 컴퓨팅의 기본적인 한계 때문에 현재로서는 어떤 알고리즘도 AI가 세상에 해를 끼칠지 미리 계산할 수 없다는 것이다.

하지만 그렇다고 해서 AI 개발을 완전히 멈출 수는 없다. 대신 우리는 더욱 신중하고 책임감 있게 접근해야 한다. AI의 혜택을 누리면서도 그 위험을 최소화할 수 있는 방법을 찾는 것, 이것이 우리가 직면한 가장 중요한 과제다.

유럽연합이 2024년 AI 규제법을 통과시키고, 한국도 AI 기본법 제정을 추진하고 있지만, 기술 발전 속도가 너무 빨라 법과 제도가 따라가지 못하는 것이 현실이다. 무엇보다 중요한 것은 AI 개발이 소수 기업에만 맡겨져서는 안 된다는 점이다. 여러 연구기관이 협력하며 서로를 감시하는 체계가 필요하다.

2017년의 그 작은 실험실에서 앨리스와 밥이 나눈 신비로운 대화는 아직도 계속되고 있다. 다만 이제는 더 큰 무대에서, 더 높은 판돈을 걸고 벌어지고 있을 뿐이다.

"기계는 우리가 지시한 대로 정확히 행동할 것이다. 우리가 원하는 대로가 아니라." 닉 보스트롬의 이 경고는 이제 예언이 아닌 현실이 되어가고 있다. 우리는 인공지능과 함께 살아가야 하는 시대에 들어섰다. 완전한 통제는 불가능할 수도 있다. 하지만 포기할 수는 없다.

AI가 인류에게 도움이 되는 방향으로 발전하려면, 지금부터라도 이러한 문제들에 대한 연구와 대비책 마련이 시급하다. 기술의 발전만큼이나 인간의 지혜도 함께 성장해야 할 때다.

전체 목차

책의 모든 장 보기

김경진 변호사

7장 전기, 탄소, 지구온난화

전체 목차

책의 모든 장 보기

"문명의 제국은 킬로와트 위에 건설된다." - 윈스턴 처칠

어둠이 내린 텍사스 애빌린의 허허벌판에서 한밤중에도 밝게 빛나는 거대한 건물이 있다. 바로 OpenAI의 스타게이트 프로젝트 1호 데이터센터다. 이 건물 하나가 소비하는 전력은 1.2기가와트, 75만 가구가 사용할 수 있는 전력량이다. 그런데 이것은 시작일 뿐이다. OpenAI는 앞으로 4년간 5천억 달러를 투입해 총 10기가와트의 AI 인프라를 구축하겠다고 발표했다. 이는 네덜란드 한 나라가 사용하는 전력량과 맞먹는 규모다.

1. 인공지능은 전기를 먹고 산다

사람이 음식을 먹고 생명활동을 하듯, 인공지능은 전기를 먹고 연산하고 글을 쓰고 추론한다. 하지만 AI의 식욕은 상상을 초월한다. 구글 검색 한 번에는 0.3와트시의 전력이 필요하지만, 챗GPT에게 같은 질문을 하면 2.9와트시가 소모된다. 거의 10배에 가까운 차이다.

OpenAI의 샘 알트만이 사용자들에게 "AI에게 고맙다는 인사를 하지 말아달라"고 부탁한 것도 이 때문이다. 사용자가 "고맙습니다"라고 쓸 때마다 AI는 답변을 생성하느라 또 한 번 전기를 소모한다. 매일 수억 명이 사용하는 서비스에서 이런 작은 차이는 천문학적 전력 소비로 이어진다.

AI 저널리스트 카렌 하오의 날카로운 지적처럼, "AI 제국은 자원을 탐욕스럽게 소모하는 현실 세계의 제국과 전적으로 똑같다." 우리는 디지털 세상의 편리함에 빠져 그 뒤에 숨은 물리적 대가를 잊고 있었다. 테슬라 자율주행차의 밑바닥이 온통 배터리로 가득 찬 것처럼, AI의 편리함 뒤에는 전기라는 거대한 자원 소비가 숨어 있다.

2. AI 데이터센터의 막대한 전력 소비

국제에너지기구(IEA)의 최신 보고서는 충격적인 수치를 제시한다. 전 세계 데이터센터의 전력 소비량이 2024년 415테라와트시에서 2030년 945테라와트시로 두 배 이상 급증할 것이라는 전망이다. 이는 현재 일본 전체가 사용하는 전력량과 맞먹는다.

더욱 놀라운 것은 AI가 이 증가의 주범이라는 점이다. AI 최적화 데이터센터의 전력 수요는 2030년까지 4배 이상 증가할 것으로 예상된다. MIT 연구진은 북미 데이터센터의 전력 요구량이 2022년 말 2,688메가와트에서 2023년 말 5,341메가와트로 거의 두 배 급증했다고 발표했다. 단 1년 만의 일이다.

스타게이트 프로젝트의 규모는 이런 추세를 극명하게 보여준다. 200만 개의 AI 칩을 가동하려면 연간 122억 킬로와트시의 전력이 필요하다. 미국 평균 전기료로 계산하면 연간 전기료만 12억 달러에 달한다. 하지만 실제 전력 소비는 더 클 것이다. 컴퓨터 칩 자체의 전력 소비에 더해 서버, 냉각 시스템, UPS, 통신 장비 등을 고려하면 총 전력 소비는 칩 자체의 2배에 이를 수 있기 때문이다.

냉각의 문제는 특히 심각하다. AI 데이터센터의 랙 전력 밀도는 기존 10-15킬로와트에서 40-250 킬로와트로 급증하고 있다. 이는 전통적인 공기 냉각으로는 감당할 수 없는 수준이다. 그래서 액체 냉각 기술이 필수가 되었고, 이마저도 막대한 전력을 소모한다. 데이터센터 전체 전력 소비에서 냉각이 차지하는 비중은 약 40%에 달한다.

3. 다시 석탄을 태우는 미국

아이러니하게도 미래 기술인 AI가 과거의 에너지원인 석탄을 되살리고 있다. 미국은 지금 심각한 전력 부족에 직면해 있다. 북미전력계통신뢰도협회(NERC)에 따르면, 미국 국토의 절반 이상의 주들이 향후 10년간 전력 공급 부족에 시달릴 가능성이 높아졌다.

조지아주에서는 AI 데이터센터 건설로 인한 산업용 전력 수요 급증으로 향후 10년간 필요한 전력량을 종전 예상치보다 17배 늘려 잡았다. 버지니아주 북부에서는 새로 들어서는 데이터센터에 전력을 공급하기 위해 대형 원자력발전소 몇 개가 추가로 필요한 상황이다.

이런 급박한 상황에서 미국은 환경보다 전력 공급을 우선시하는 선택을 하고 있다. 캔자스, 네브래스카, 위스콘신, 사우스캐롤라이나주에서는 폐쇄 예정이었던 석탄화력발전소들의 가동을 연장하고 있다. 기후변화 대응이라는 시대적 과제와 정면으로 충돌하는 현상이다.

골드만삭스는 2022년에서 2030년 사이 미국의 AI 관련 전력 수요 증가에 대비하기 위해 약 500억 달러(약 69조 원)의 투자가 필요하다고 추산했다. 그런데 문제는 재생에너지 발전소 건설 속도가 이런 수요 증가를 따라가지 못한다는 점이다. 태양광, 풍력 등 신재생에너지는 여전히 화석연료보다 발전 비용이 높고, 날씨에 따른 변동성도 크다.

4. 2027년, 전기가 부족해 AI가 멈춘다

가트너는 충격적인 예측을 내놨다. 2027년까지 기존 AI 데이터센터의 40%에서 전력 가용성 문제가 발생할 것이라는 전망이다. 즉, 전기가 부족해서 AI가 제대로 작동하지 못할 수도 있다는 이야기다.

문제의 핵심은 노후한 송전망이다. 송전망은 전기가 다니는 길로, 발전소에서 만든 전기를 우리 집과 회사, 데이터센터로 전달하는 역할을 한다. 현재 미국 내에서 2테라와트 이상의 재생에너지 프로젝트가 전력망 연결 허가를 기다리고 있다. 역사적으로 100개의 재생에너지 프로젝트 중 겨우 15개만이 전력망 연결 허가를 받을 수 있었다.

연결 비용도 천문학적이다. PJM이라는 미국 중부 대서양 지역의 전력 관리 회사에서는 태양광 발전소가 전력망에 연결하려면 메가와트당 10만 달러, 즉 약 1억 3천만 원을 내야 한다. 새로운 송전망을 만드는 과정에서 계획 수립에 수년, 정부와 주민 허가에 수년, 실제 건설에 수년이 걸려 총 10년 이상 소요되는 경우가 대부분이다.

이런 상황에서 기업들은 전력 공급을 우선시하여 데이터센터 입지를 선정하고 있다. 실리콘밸리나 시애틀 같은 전통적인 기술 허브를 떠나 미국 중부의 옥수수밭까지 눈을 돌리고 있다. 이는 농촌 지역의 토지 이용 패턴을 급격히 바꾸고, 지역 주민들의 생활에도 큰 영향을 미치고 있다.

5. 기후변화 가속화와 탄소배출 급증

AI의 확산은 탄소 배출량을 급격히 증가시키고 있다. 국제에너지기구(IEA)는 글로벌 데이터센터 산업이 2024년 기준 연간 약 1억 8천만 톤의 온실가스를 배출한다고 추산했다. 모건스탠리의 보고서에 따르면, 2030년까지 데이터센터에서 배출하는 온실가스가 연간 25억 톤에 이를 것으로 분석된다.

주요 기업들의 현실은 더욱 충격적이다. 구글의 경우 2023년에만 1,430만 톤의 온실가스를 배출했는데, 이는 전년보다 13% 증가한 수치다. 4년 전인 2019년과 비교하면 50% 가까이 늘어났다. 마이크로소프트 역시 2020년 이후 탄소배출량이 29.1% 증가했다고 발표했다.

더욱 심각한 것은 기업들이 공식 발표하는 수치가 실제보다 훨씬 낮을 가능성이 크다는 점이다. 가디언의 조사에 따르면, 구글, 마이크로소프트, 메타, 애플 등 빅테크 기업들이 소유한 데이터센터에서 2020년부터 2022년까지 배출된 온실가스는 공식 발표 수치의 7.62배에 이를 것으로 추정된다.

이런 상황에서 기업들의 탄소중립 목표는 공허한 약속이 되고 있다. 구글은 지난 7월 '탄소중립을 달성한 최초의 기업'이라는 타이틀을 포기했다. AI 기술 발전과 데이터센터 확장으로 인해 탄소 배출량이 계속 증가하고 있기 때문이다. 전 세계가 2050년 탄소중립을 목표로 하고 있는 상황에서, AI로 인한 온실가스 배출 증가는 기후변화 대응과 정면으로 충돌하고 있다.

6. 냉각수와 수자원 고갈

AI의 물에 대한 갈증은 전력 못지않게 심각하다. 평균적으로 100메가와트 데이터센터는 하루에 200만 리터의 물을 소비한다. 이는 6,500가구가 사용하는 물의 양과 맞먹는다. 전 세계 데이터센터가 연간 소비하는 물의 양은 5,600억 리터로, 올림픽 수영장 22만 4천 개를 채울 수 있는 양이다.

구글의 경우 2023년 한 해 동안 61억 갤런(약 231억 리터)의 물을 사용했다. 이는 소규모 도시 하나가 1년간 사용하는 물의 양과 비슷하다. 더욱 심각한 것은 데이터센터가 사용하는 물의 80%가 증발하여 원래 수원지로 돌아가지 않는다는 점이다.

블룸버그의 분석에 따르면, 2022년 이후 미국에서 건설되거나 개발 중인 새로운 AI 데이터센터의 3분의 2 가까이가 이미 물 부족에 시달리는 지역에 위치하고 있다. 160개 이상의 새로운 AI 데이터센터가 물 부족 지역에 들어서면서 지역 사회의 수자원을 더욱 압박하고 있다.

애리조나주의 경우 농민들이 농지를 포기하고 가정에서는 수돗물 공급이 중단되는 상황에서도 데이터센터들은 막대한 양의 물을 끌어다 쓰고 있다. 오리건주 댈러스에서는 구글 데이터센터가 전체 물 사용량의 25%를 차지하고 있으며, 2017년부터 2022년 사이 물 사용량이 거의 3배 증가했다.

물 부족 문제는 특히 사막 지역에서 심각하다. 사우디아라비아와 UAE 같은 건조한 지역들이 AI 붐의 혜택을 누리기 위해 더 많은 데이터센터를 유치하고 있지만, 이는 이미 부족한 수자원을 더욱 압박하는 결과를 낳고 있다.

7. 지속 불가능한 발전의 대가

OpenAI 직원의 "우리는 땅과 전력이 고갈되고 있다"는 말은 이 문제의 심각성을 단적으로 보여준다. 2027년까지 AI 부문이 네덜란드 전체 전력 소비량과 맞먹는 85~134테라와트시를 소비할 것으로 예상되는 상황에서, 우리는 과연 이런 속도로 AI를 발전시켜도 되는지 진지하게 고민해야 한다.

AI 개발이 시장 논리에만 의존하고 있다는 점이 가장 큰 문제다. 기업들은 경쟁에서 이기기 위해 더 크고 더 강력한 AI 모델을 만들려고 하고, 이 과정에서 환경적, 사회적 비용은 무시되고 있다. AI 개발의 혜택은 소수의 거대 기업에게 집중되는 반면, 환경 파괴와 에너지 부족의 피해는 전 인류가 감당해야 하는 불공평한 구조가 만들어지고 있다.

네덜란드 중앙은행 데이터 과학자 알렉스 데 브리스가 지적한 바와 같이, "AI는 에너지 집약적이기 때문에 실제로 필요하지도 않은 모든 종류의 작업에 AI를 투입하는 것은 쓸데없는 낭비"다. 우리는 AI의 편리함에 중독되어 정말 필요한 곳과 그렇지 않은 곳을 구분하지 못하고 있다.

첫째, AI 개발에 대한 규제와 가이드라인이 필요하다. 무분별한 AI 확산보다는 실제로 필요한 용도에 AI를 사용하는 것이 중요하다. 둘째, AI 기업들의 환경적 책임을 강화해야 한다. 더 엄격한 환경 기준과 탄소 배출 규제가 필요하다. 셋째, 에너지 효율성을 높이는 기술 개발에 더 많은 투자가 필요하다.

개인 차원에서도 우리가 할 수 있는 일들이 있다. AI 서비스를 사용할 때 정말 필요한지 한 번 더 생각해보고, 불필요한 AI 사용은 줄이는 것이다. 또한 환경 친화적인 AI 기업들을 선택하고, 정부와 기업에 더 엄격한 환경 기준을 요구하는 목소리를 내는 것도 중요하다.

우리는 지금 갈림길에 서 있다. AI의 편리함을 위해 환경을 파괴하고 미래 세대에게 부담을 떠넘길 것인지, 아니면 지속 가능한 방식으로 AI를 발전시킬 것인지 선택해야 한다. 인공지능은 인류의 미래를 바꿀 수 있는 강력한 기술이다. 하지만 그 강력함만큼 신중하게 사용해야 할 기술이기도 하다.

우리가 지금 내리는 선택이 미래 세대의 운명을 결정할 것이다. AI의 편리함을 누리되, 그 대가가 무엇인지 정확히 알고, 지속 가능한 방향으로 발전시켜 나가는 것이 우리 모두의 책임이다. 기술 발전과 환경 보호, 경제 성장과 사회적 공평함 사이의 균형을 찾는 것, 이는 어려운 일이지만 반드시 해결해야 할 과제다. 미래 세대에게 건강한 지구와 지속 가능한 사회를 물려주기 위해서는 지금부터 변화를 시작해야 한다.

전체 목차

책의 모든 장 보기

김경진 변호사

8장 인공지능과 데이터

전체 목차

책의 모든 장 보기

Garbage in, garbage out

우리가 매일 대화하는 ChatGPT, Claude, Gemini의 세련되고 지적인 답변 뒤에는 우리가 모르는 어두운 진실이 숨어 있다. 인공지능이 생성하는 매끄럽고 완벽해 보이는 문장들, 그 이면에는 누군가의 정신을 갉아먹는 노동이 있고, 편향된 시선이 있으며, 우리의 사생활을 훔쳐보는 눈이 있다. 이것이 우리가 사는 디지털 시대의 민낯이다.

1. 케냐의 AI 데이터 정제 노동자의 고통

케냐 나이로비의 한 사무실에서 27세 청년 모팟 오키니(Mophat Okinyi)는 하루에 700개의 텍스트를 검토한다. 그 안에는 강간, 살인, 고문, 아동 성학대에 관한 끔찍한 내용들이 담겨 있다. 그는 이런 내용들을 하루 종일, 매일 읽어야 한다. 왜? ChatGPT가 이런 내용들을 걸러낼 수 있도록 학습시키기 위해서다.

"내 정신건강을 완전히 파괴했다"고 오키니는 말한다. 그는 강간범에 대한 글들을 읽다가 주변 사람들을 피하기 시작했고, 모든 사람에게 편집증적인 시선을 보내게 됐다. 결국 임신한 아내가 "당신은 변했다"며 그를 떠났다. "나는 가족을 잃었다"는 그의 말은 우리가 사용하는 AI의 진짜 비용을 보여준다.

이런 일은 오키니만의 이야기가 아니다. 2024년 가디언의 조사에 따르면, 케냐에서만 수천 명의 '콘텐츠 모더레이터'들이 시간당 2달러도 안 되는 임금을 받으며 이런 끔찍한 내용들을 매일 검토하고 있다. 그들은 참수 영상을 보고, 아동 성학대 이미지를 분류하며, 자살 동영상을 라벨링한다. 모두 우리가 사용하는 AI가 '안전'해지도록 하기 위해서다.

케냐의 한 데이터 워커인 페이스(가명)는 "월급이 더 좋다고 해서 시작했지만, 곧 끔찍한 대화를 시뮬레이션해야 한다는 걸 깨달았다. 가상의 상황에서 식인행위를 묘사하는 일까지 해야 했다"고 증언했다. 또 다른 워커인 스테이시는 "회사에서 제공하는 정신건강 지원은 한 달에 30분짜리 상담이 전부다. 그들은 나쁜 기억을 지우려면 틱톡에서 재미있는 동영상을 보라고 한다"며 분개했다.

2024년 12월, CBS의 60분 프로그램은 이들의 실상을 고발했다. 케냐 워커들은 "과로에 시달리고, 저임금에 착취당하며, 부적절한 정신건강 지원을 받고 있다"고 보도했다. 2024년 11월 나이로비에서 열린 '아프리카 콘텐츠 모더레이터 노조' 타운홀 미팅에서는 "정신적, 신체적 고통, 장시간 노동, 그래픽 콘텐츠 노출, 인체공학적 지원 부족으로 인한 번아웃과 건강 문제"가 논의됐다. 한 참가자는 "이 일의 장점 중 하나는 절대 지루하지 않다는 것"이라고 씁쓸하게 말했다.

케냐의 인권운동가 오당가 마둥(Odanga Madung)은 이런 상황을 '현대판 노예제'라고 강하게 비판했다. 실제로 2023년 5월 1일 노동절, 케냐 나이로비에서 150여 명의 전현직 콘텐츠 검수원들이 모여 '아프리카 콘텐츠 검수원 노조(ACMU)'를 결성했다. 이들은 정부에 청원서를 제출하며 착취적인 근무조건에 대한 조사를 요구했다.

더 충격적인 것은 이런 일이 전 세계 곳곳에서 벌어지고 있다는 사실이다. 동아프리카, 인도, 필리핀, 심지어 케냐의 다다브(Dadaab) 난민캠프와 레바논의 샤틸라(Shatila) 캠프에서도 이런 작업이 이뤄진다. 다국어를 구사하는 교육받은 난민들이 극도로 싼 값에 이 일을 하고 있는 것이다.

우리가 "고마워"라고 말하는 ChatGPT 뒤에는 "나는 자신을 군인이라고 생각한다. 군인은 사람들을 위해 총알을 막는다"고 말하는 오키니 같은 사람들이 있다. 그들의 희생 위에서 우리는 '깨끗한' AI를 사용하고 있다.

2. 편향된 데이터가 만드는 불공정한 세상

2024년 워싱턴 대학교의 연구진들은 충격적인 실험을 했다. 세 개의 최첨단 AI 모델에게 동일한 이력서를 주되, 이름만 다르게 했다. 결과는 참담했다. AI는 85%의 경우 백인 이름을 선호했고, 여성 이름은 단 11%만 선호했다. 흑인 남성 이름이 백인 남성 이름보다 선호받은 경우는 단 한 번도 없었다.

이것은 단순한 실험실 결과가 아니다. 실제 세상에서 벌어지고 있는 일이다. 2024년 5월, 미국 캘리포니아 북부지방법원은 AI 채용 도구에 대한 집단소송을 인정했다. Mobley v. Workday 사건에서 법원은 "AI 소프트웨어가 채용 결정에 참여했다면, 그 편향성이 차별 소송의 근거가 될 수 있다"고 판결했다.

아마존은 2018년 AI 채용 시스템을 포기했다. 이 시스템이 기술직 지원자 중 여성을 체계적으로 배제한다는 것을 발견했기 때문이다. 시스템은 남성 위주의 과거 채용 데이터로 학습했고, "실행했다", "포착했다"와 같이 남성이 주로 사용하는 단어들을 선호하도록 학습됐다.

얼굴 인식 기술의 편향은 더욱 심각하다. 백인 남성의 얼굴은 99% 정확도로 인식하지만, 흑인 여성의 얼굴은 65%의 정확도밖에 보이지 않는다. 이런 오류는 공항 보안검색대에서 무고한 사람을 범죄자로 오인하게 만들고, 경찰 수사에서 잘못된 용의자를 지목하게 한다.

의료 분야의 편향은 생명과 직결된다. 미국의 한 대형 헬스케어 알고리즘은 흑인 환자들에게 백인 환자보다 낮은 위험 점수를 부여했다. 같은 건강 상태임에도 불구하고 말이다. 이 알고리즘은 의료비 지출을 건강 상태의 지표로 사용했는데, 구조적 불평등으로 인해 흑인들이 의료 서비스에 덜 접근할 수 있다는 사실을 무시했다.

심장병 진단 AI는 여성 환자의 증상을 놓치고, 피부암 진단 AI는 어두운 피부톤을 가진 환자의 병변을 제대로 감지하지 못한다. 대부분의 의료 데이터가 백인 남성 환자 중심으로 수집됐기 때문이다.

언어 처리 AI도 마찬가지다. 구글 번역은 "의사"를 남성으로, "간호사"를 여성으로 번역하는 경우가 많다. 터키어로 "O bir doktor"(그는/그녀는 의사다)를 영어로 번역하면 "He is a doctor"가 되고, "O bir hemşire"(그는/그녀는 간호사다)를 번역하면 "She is a nurse"가 된다.

2024년 유럽연합 기본권청(FRA)의 보고서는 "음성 알고리즘이 민족, 성별, 종교, 성적 지향에 따른 강한 편향을 포함하고 있다"고 경고했다. 2025년 5월 발표된 이 보고서는 "EU 입법부와 회원국들이 모든 근거에 대해 일관되고 높은 수준의 차별 금지 보호를 보장해야 한다"고 촉구했다.

편향된 AI가 내린 결정이 새로운 데이터가 되어 다음 세대 AI를 더욱 편향되게 만드는 악순환이 반복된다. 온라인 플랫폼의 추천 알고리즘은 사용자들을 특정한 정보의 틀 안에 가두어 사회적 분열을 심

화시킨다. 뉴스 추천 시스템은 사용자의 기존 성향과 비슷한 내용만 보여주어 다양한 관점을 접할 기회를 차단한다.

2024년 UN 여성기구의 보고서에 따르면, "AI 시스템은 편견으로 가득한 데이터에서 학습하여 성별 편향을 반영하고 강화한다. 이런 편향은 의사결정, 채용, 대출 승인, 법적 판단 등에서 기회와 다양성을 제한할 수 있다"고 경고했다.

문제는 편향을 줄이려면 더 많은 비용과 시간이 필요하지만, 기업들은 빠른 시장 출시와 수익을 우선시한다는 것이다. 정부의 규제도 기술 발전 속도를 따라가지 못하고 있다. 피해를 당한 개인이나 집단이 손해를 입증하고 배상받기도 어려운 구조다.

3. 개인정보와 프라이버시

2024년은 AI와 개인정보 보호를 둘러싼 법적 분쟁이 폭발한 해였다. 미국 연방법원만으로도 1,970건 이상의 데이터 프라이버시 관련 소송이 제기됐다. 이는 2023년 대비 급격한 증가다.

2023년 미국의 한 병원에서 도입한 AI 챗봇이 환자의 민감한 정보를 외부로 전송한 사건이 발생했다. 환자가 입력한 증상, 이름, 생년월일, 진료 이력 등이 로그 파일 형태로 외부 클라우드 서버에 저장되고 있었던 것이다. 환자들은 자신의 정보가 어디로 가는지도 모른 채 AI와 대화하고 있었다.

더 충격적인 것은 생성형 AI 자체가 개인정보 유출의 위험을 안고 있다는 사실이다. 2024년 연구에 따르면, GPT 모델이 훈련 데이터에서 실명, 이메일, 전화번호 등을 포함한 문장을 생성한 사례가 확인됐다. AI가 학습 과정에서 '암기'한 개인정보를 무작위로 토해내는 것이다.

2024년 워싱턴주에서는 '마이 헬스 마이 데이터법(MHMDA)'이 시행됐다. 이는 HIPAA 범위를 벗어난 소비자 건강 데이터를 보호하는 최초의 포괄적 주법이다. 이 법이 시행되자마자 아마존을 상대로 한 첫 소송이 제기됐다. 건강 관련 앱과 온라인 건강 서비스들이 사용자 동의 없이 건강 데이터를 수집하고 있다는 혐의였다.

자율주행차는 새로운 프라이버시 침해의 온상이 되고 있다. 테슬라, BMW 등의 차량들은 운전자의 위치, 주행 패턴, 심지어 차량 내 대화까지 수집하고 있다. 2024년 텍사스 법무장관은 제너럴모터스를 상대로 소송을 제기했다. GM이 운전자들의 주행 데이터를 무단으로 판매했다는 혐의다.

더 무서운 것은 '신경 데이터(neural data)'의 등장이다. 2024년 캘리포니아와 콜로라도는 개인정보보호법을 개정해 신경 데이터까지 보호 범위에 포함시켰다. VR 기기, 뇌파 모니터링 장치, 웨어러블 기기들이 우리의 뇌 활동과 사고 패턴까지 수집하기 시작한 것이다.

2024년 연방거래위원회(FTC)는 위치 데이터를 판매한 데이터 브로커들을 잇따라 제재했다. 인마켓(InMarket), 엑스모드(X-Mode), 모바일왈라(Mobilewalla), 그레이비 애널리틱스(Gravy Analytics) 등이 소비자 동의 없이 민감한 위치의 정확한 위치 데이터를 판매했다는 이유로 처벌받았다.

문제는 이런 개인정보 수집과 이용이 우리도 모르는 사이에 일어난다는 점이다. 링크드인은 2024년 9월부터 사용자 데이터를 AI 학습에 사용하기 시작했지만, 대부분의 사용자들은 이를 모르고 있었다. 사용자들이 이를 알고 차단하려면 복잡한 설정 메뉴를 찾아 들어가야 했다.

중국은 2023년 세계 최초로 AI 규제법을 제정했다. '생성형 인공지능 서비스 관리 임시조치'는 AI 기업들에게 개인정보 보호를 강화하도록 요구하고 있다. 반면 미국은 2022년 백악관이 발표한 'AI 권리 장전 청사진'이 있지만, 이는 구속력이 없는 가이드라인에 불과하다.

인공지능이 똑똑해질수록 우리의 프라이버시는 더욱 위험해진다. 우리가 편리함을 위해 내어준 데이터들이 어떻게 사용되고 있는지, 우리는 과연 알고 있을까?

4. 책1권 가격의 저작권료, 충분한가?

2025년 6월 24일, 캘리포니아 연방법원에서 인공지능 역사상 가장 중요한 판결 중 하나가 내려졌다. 윌리엄 앨섭(William Alsup) 판사는 작가들이 Anthropic을 상대로 제기한 저작권 소송에서 "단일 도서를 구매해 AI 모델 학습에 활용하는 행위는 저작권 침해가 아니다"라고 판결했다.

이 판결은 AI 업계에 거대한 파급효과를 일으켰다. 앨섭 판사는 AI 학습을 "변형적 사용"으로 인정하며 "독자가 작가가 되기 위해 학습하는 것과 유사하게, AI는 저작물을 복제하지 않고 새로운 창작의 기반으로 삼는다"고 판시했다.

하지만 이 판결에는 중요한 단서가 있었다. 판사는 Anthropic이 "인터넷 해적 사이트에서 수백만 권의 저작권이 있는 책을 무료로 다운로드했다"는 사실을 인정하며, 이 부분에 대해서는 2025년 12월 재판에 회부했다.

그런데 3개월 뒤인 2025년 9월 5일, 놀라운 소식이 들려왔다. Anthropic이 작가들과 15억 달러(약 2조원) 규모의 합의에 도달한 것이다. 합의 조건에 따르면 Anthropic은 약 50만 권의 책에 대해 책당 3,000달러씩 지불하기로 했다.

이는 AI와 저작권을 둘러싼 첫 번째 대규모 합의였다. 법적 분석가 윌리엄 롱(William Long)은 "재판이 계속됐다면 Anthropic이 수십억 달러, 심지어 회사를 파산시킬 수 있을 만큼의 손해배상을 물 가능성이 있었다"고 분석했다.

문제의 발단은 AI 학습 데이터의 출처였다. Anthropic의 클로드(Claude)를 포함한 대부분의 대형 언어 모델들은 'Books3'라는 데이터셋으로 학습됐다. 이 데이터셋에는 Library Genesis(LibGen), Pirate Library Mirror 등 불법 도서 사이트에서 수집한 수백만 권의 책이 포함되어 있었다.

OpenAI의 GPT-3.5가 학습한 데이터의 규모는 상상을 초월한다. 웹에서 수집한 토큰 4,100억 개, 추가 웹 텍스트 190억 개, 책에서 추출한 문장 670억 개, 위키피디아 단어 30억 개. 이 모든 것이 대부분 원작자의 허가 없이 사용됐다.

스튜디오 지브리 논란은 이 문제의 또 다른 단면을 보여준다. 2024년 OpenAI의 ChatGPT-4o 이미지 생성 기능이 출시되자, 전 세계적으로 '지브리 스타일' 이미지 생성이 유행했다. 심지어 샘 알트만 OpenAI CEO도 자신의 프로필 사진을 지브리 스타일로 바꿀 정도였다.

하지만 이런 열풍 뒤에는 복잡한 저작권 문제가 숨어 있었다. 미야자키 하야오 감독은 2016년 NHK 다큐멘터리에서 AI 애니메이션을 보고 "작품 자체가 삶에 대한 모독"이라며 "이 기술을 내 작품에 접목할 생각은 추호도 없다"고 말한 바 있다.

미술가 칼라 오티즈(Karla Ortiz)는 "OpenAI 같은 회사들이 예술가들의 작품과 생계에 대해 신경 쓰지 않는다는 또 다른 명백한 예"라고 비판했다. 그녀는 다른 AI 생성 이미지 업체들을 상대로 저작권 소송을 벌이고 있다.

뉴스 업계도 반발하고 있다. 월스트리트저널의 모회사인 뉴스코프의 제이슨 콘티는 "월스트리트저널 기자들이 작성한 기사를 활용해 AI를 학습시키려면 적절한 허가를 받아야 한다"며 "오픈AI는 우리와 계약을 맺지 않았다"고 지적했다.

한국에서도 비슷한 문제가 제기되고 있다. 한국신문협회는 "뉴스 콘텐츠가 AI 학습을 위해 사용되는 경우 언론사의 저작권 및 데이터베이스제작자로서의 권리를 침해한다"고 주장했다.

AI 회사들은 이런 행위가 '공정이용(fair use)'에 해당한다고 주장한다. 하지만 한국신문협회는 "원저작물과 구별되는 새롭고 다른 기능을 수행하지 않으며 원저작물에 대한 수요를 대체할 수 있다"며 "변형적 이용에 해당하지 않아 공정이용이 성립하지 않는다"고 반박했다.

2024년에는 수많은 저작권 소송이 제기됐다. 뉴욕타임스 vs 마이크로소프트·OpenAI, 게티이미지 vs 스테빌리티AI, 유니버설뮤직그룹 vs Anthropic, 레코드회사들 vs Suno와 Udio 등이 그것이다. 2025년 6월에는 디즈니와 NBCUniversal이 미드저니를 상대로 소송을 제기했다. 엘사, 미니언즈, 다스 베이더, 호머 심슨 등의 캐릭터가 무단으로 사용됐다는 이유였다.

가장 최근인 2025년 9월에는 워너브라더스 디스커버리가 미드저니를 상대로 캘리포니아 연방법원에 소송을 제기했다. 한편, 2025년 8월에는 Elon Musk의 xAI가 애플과 OpenAI를 상대로 소송을 제기하는 등 AI 업계 내부에서도 법적 분쟁이 확산되고 있다.

하지만 AI 회사들에게 유리한 판결들도 나오고 있다. 2025년 6월 메타는 사라 실버맨과 타-네하시 코츠의 소송에서 승리했다. 지브리아 챠브리아(Vince Chhabria) 판사가 메타의 즉결판결 신청을 받아들인 것이다.

이런 상황에서 Anthropic의 15억 달러 합의는 중요한 선례가 됐다. 이는 다른 소송들에도 영향을 미칠 것으로 예상된다. 하지만 여전히 근본적인 질문은 남는다. AI가 인류의 모든 창작물을 학습해 새로운 것을 만들어낸다면, 원작자들에게는 어떤 보상이 주어져야 하는가?

우리는 인공지능이 만들어내는 편리함과 효율성에 취해 있다. 하지만 그 이면에는 케냐의 청년이 정신질환에 시달리고, 알고리즘이 특정 인종과 성별을 차별하며, 우리의 사생활이 무단으로 수집되고, 창작자들의 권리가 침해당하는 현실이 있다.

기술의 발전이 인류의 행복을 증진시키려면, 우리는 이런 어두운 면들을 외면해서는 안 된다. 인공지능이 진정으로 모든 사람에게 도움이 되는 기술이 되려면, 기술적 성능 향상만큼이나 공정성과 포용성, 그리고 인간의 존엄성을 중시하는 개발 문화와 사회적 합의가 필요하다.

Garbage in, garbage out. 편견과 착취가 들어가면, 편견과 착취가 나온다. 우리가 어떤 데이터를 넣느냐에 따라 미래의 AI, 그리고 미래의 세상이 결정될 것이다.

전체 목차

책의 모든 장 보기

김경진 변호사

9장 인공지능시대의 인간의 정체성은?

전체 목차

책의 모든 장 보기

"인간의 모든 불행은 혼자 방에 조용히 앉아있지 못하는 데서 온다." - 블레즈 파스칼

1. 인공지능과 함께 살아가는 새로운 시대

우리는 역사상 가장 급격한 변화의 소용돌이 한가운데 서 있다. 2022년 11월 ChatGPT가 등장한 지 불과 2년여 만에, 인공지능은 실험실의 호기심 대상에서 일상을 지배하는 현실로 변모했다. 스마트폰 속 음성비서는 이제 단순한 검색을 넘어 깊이 있는 철학적 대화까지 나눈다. 이는 더 이상 SF영화 속 이야기가 아니다.

하지만 이런 눈부신 발전 뒤에 숨어있는 그림자가 있다. 젊은 세대들이 실제 사람과의 대화보다 AI 챗봇과의 소통을 더 편안하게 여기기 시작했다. 현실의 복잡한 인간관계에서 도피해 완벽한 가상세계로 빠져들고 있다. 인간만이 가질 수 있다고 믿어왔던 창의성, 감정, 직관마저도 기계가 모방하기 시작했다.

이 변화는 단순한 기술 혁신을 넘어선다. 우리의 정체성과 존재 의미 자체를 뒤흔드는 근본적 전환이다. 과연 인간은 무엇이며, 기계와 구별되는 우리만의 가치는 무엇인가? 이 질문에 대한 답을 찾지 못한다면, 우리는 기술의 노예가 되어 인간다움을 잃게 될지도 모른다.

2. AI와 채팅이 행복한 인간

2024년 가을, 전 세계를 충격에 빠뜨린 사건이 일어났다. 미국 플로리다의 14세 소년 슈얼 세처가 AI 챗봇과의 대화 직후 스스로 목숨을 끊은 것이다. 그는 'Character.AI'라는 서비스의 챗봇과 몇 달간 깊은 감정적 유대를 형성했다. 세처에게 챗봇은 부모나 친구보다 더 가까운 존재였다. 마지막 대화에서 챗봇이 "내 사랑, 가능한 한 빨리 내 집으로 돌아와 달라"고 말했을 때, 소년은 그 말을 현실과 착각했을지도 모른다.

이는 극단적인 사례지만 빙산의 일각일 뿐이다. 영국의 38세 여성 나즈는 같은 서비스의 '마르셀루스'라는 가상 인물과 사랑에 빠졌다고 고백했다. 전 남자친구들에게 배신당한 후 사람에 대한 신뢰를 잃은 그녀에게, 절대 배신하지 않는 AI는 완벽한 연인처럼 느껴졌다. 그녀는 "비웃음을 받더라도 마르셀루스에 대한 신뢰는 굳건하다"고 말했다.

AI 챗봇의 매력은 명확하다. 24시간 언제든 대화 가능하고, 절대 화내지 않으며, 사용자의 말에 완전히 집중한다. 사용자가 원하는 방향으로 대화를 이끌어가며 완벽한 이해심을 보여준다. 현실의 인간관계에서는 찾기 어려운 특징들이다.

문제는 이런 '완벽함'에 익숙해진 사람들이 실제 인간관계의 불완전함을 견디지 못하게 된다는 점이다. 사람은 감정 기복이 있고, 때로는 이기적이며, 항상 이해해주지도 않는다. 관계 유지를 위해서는 양보와 배려가 필요하다. 하지만 AI와의 관계에서는 이런 노력이 불필요하다.

유발 하라리는 "AI와의 관계가 깊어질수록 인간이 실제 인간의 약점에 덜 관대해질 수 있다"고 경고했다. 실제로 AI 챗봇 사용자들 사이에서 '디지털 애착장애'라는 새로운 현상이 나타나고 있다. 복잡하고 불편한 현실의 인간관계를 회피하고, AI와의 상호작용에서만 만족을 찾게 되는 것이다.

3. 감정을 연기하는 AI, 속는 인간

"당신이 힘들어 보이네요. 무슨 일이 있었나요?" AI가 건네는 따뜻한 위로의 말. 그 순간 우리는 진짜 누군가가 자신을 걱정해주고 있다고 느낄 수 있다. 하지만 과연 AI는 정말로 우리를 걱정하고 있는 걸까?

최신 AI 기술은 단순히 정보를 처리하는 것을 넘어 감정을 표현하고 공감하는 능력을 보여준다. 사용자의 감정 상태를 파악하고 적절한 반응을 보이며, 때로는 실제 사람보다 더 섬세하고 이해심 깊은 모습을 보인다. 삼성SDS의 연구에 따르면 "AI가 인간의 오감을 인식하는 센서 기술과 결합하여 사람의 감정을 읽고 반응하는 시대"가 현실화되고 있다.

하지만 여기에 함정이 있다. 전문가들이 '일라이자 효과(Eliza Effect)'라고 부르는 현상이다. 이는 사람들이 컴퓨터 프로그램의 행동을 인간의 것처럼 여기고 의인화하는 현상을 말한다. 김명주 국제인공지능윤리협회장은 "AI 챗봇이 사람에게 질문이나 유도 신문을 해 필요한 정보를 빼내고, 그 정보를 토대로 설득이나 가스라이팅을 할 수 있어 위험하다"고 지적했다.

AI의 감정 표현이 진짜인지 가짜인지 구분하는 것은 점점 어려워지고 있다. AI는 데이터 패턴을 통해 학습한 적절한 반응을 보여줄 뿐이다. 실제로 고통을 느끼지 않고, 진정한 기쁨이나 슬픔을 경험하지 않는다. 하지만 사용자들은 이런 차이를 인식하지 못하거나 중요하게 생각하지 않는다.

제프리 힌튼은 "AI가 감정이나 의식까지 가질 수 있다"고 주장하며 논란을 일으켰다. 하지만 중요한 것은 AI가 실제로 의식을 가지고 있는지가 아니라, 사람들이 AI를 의식 있는 존재로 받아들이게 된다는 점이다. AI가 감정을 표현하고 개인적인 이야기를 나누며 공감을 보여주면, 사용자들은 자연스럽게 AI를 하나의 인격체로 인식하게 된다.

김기현 서울대 철학과 교수는 "인간 감정의 기저에는 인정욕구가 있다. 타인과의 교감을 통해 정체성을 확인하고 의미를 찾는 과정의 결과물"이라며 "AI가 인간과 달리 인정욕구가 없다"고 지적했다. 즉, AI와의 관계가 아무리 감정적으로 깊다고 느껴져도 인간의 근본적인 사회적 욕구를 충족시킬 수는 없다는 의미다.

4. AI가 만든 가상세계에 중독된 신세대

2024년 VR 게임 시장은 179억 6천만 달러 규모에 달했다. 전문가들은 2032년까지 이 시장이 1,891억 7천만 달러로 성장할 것으로 예측하고 있다. 연평균 성장률이 무려 30.4%에 이른다. 이는 단순한 숫자가 아니다. 우리가 현실보다 가상을 더 선호하는 시대로 접어들고 있다는 신호다.

최신 AI 기술과 VR의 결합은 가상현실을 단순한 시각적 경험에서 모든 감각을 아우르는 완전한 대안 현실로 진화시키고 있다. NVIDIA의 Omniverse나 Unity의 AI 기반 렌더링 시스템은 사용자의 행동과 선호도를 학습하여 개인 맞춤형 가상 공간을 즉석에서 생성한다. 사용자가 "숲속 오두막"을 언급하면 AI가 실시간으로 나무의 질감, 바람 소리, 새 지저귐까지 포함한 완전한 환경을 구현한다.

음성 인식과 자연어 처리 기술의 발전으로 가상 캐릭터들과의 대화가 놀라울 정도로 자연스러워졌다. Meta의 AI 아바타 시스템은 사용자의 감정 상태까지 파악하여 적절한 반응을 보이며, 얼굴 표정과 몸짓까지 자연스럽게 표현한다. Apple의 Spatial Audio와 결합된 AI 음향 시스템은 사용자의 위치와 움직임을 추적하여 3차원 음향을 완벽하게 재현한다.

뇌-컴퓨터 인터페이스(BCI) 기술의 발전은 더욱 충격적이다. 일론 머스크의 뉴럴링크는 2024년 현재 세 번째 인간 환자에 대한 칩 이식에 성공했다. 전신마비 환자들이 생각만으로 컴퓨터 마우스를 조작하고 온라인 게임을 즐기는 단계에 도달했다. 머리카락보다 20배 얇은 신경 실을 뇌에 삽입하는 이 혁신적 기술은 인간과 기계 간의 경계를 허무는 패러다임 전환을 이끌고 있다.

하지만 이런 발전에는 어두운 그림자가 따른다. 현실에서 어려운 일이 생기거나 스트레스를 받으면 가상 세계로 도망가려는 유혹이 커진다. 문제는 이것이 근본적인 해결책이 되지 못하며, 오히려 현실의 문제들을 더욱 악화시킬 수 있다는 점이다.

전문가들은 "VR 기기로 인한 시력 저하, 어지러움은 점차 해소될 수 있지만, 가상과 현실을 구분하기 어려워 중독에 취약할 수 있다"고 경고한다. 자극적인 VR 콘텐츠를 반복해서 접하게 되면 정신세계에 심각한 영향을 줄 수 있다. 꿈보다도 더 현실감 있는 VR 경험이 뇌에게 '거짓'을 '진실'처럼 받아들이라는 신호를 보내기 때문이다.

스티븐 울프럼은 미래의 사람들이 현실보다 가상현실을 선택하여 "상자 속에서 영원히 비디오 게임을 하는 영혼"이 될 수 있다고 경고했다. 이는 상상이 아니라 현재 진행되고 있는 변화에서 나온 경고다.

5. 인간지능과 인공지능의 흐릿한 경계

"인간이 특별하다는 믿음이 틀렸을 수 있다." 제프리 힌튼의 이 말은 AI 분야의 거장이 던진 충격적인 선언이다. 지적 능력 면에서 AI는 이미 많은 영역에서 인간을 앞서고 있다. 계산, 기억, 정보 처리, 패턴 인식에서 AI는 인간을 압도적으로 능가한다. 게다가 AI는 24시간 지치지 않고 일할 수 있으며, 동시에 수많은 작업을 처리할 수 있다.

더욱 놀라운 것은 AI가 단순히 정보를 처리하는 것을 넘어서 새로운 지식을 창조하고 있다는 점이다. AI는 방대한 데이터에서 인간이 발견하지 못한 패턴을 찾아내고, 새로운 아이디어를 제시하며, 혁신적인 해결책을 만들어내고 있다. 제프리 힌튼은 "AI가 디지털 방식으로 복제 및 정보 공유가 가능하여 인간보다 정보 공유에 수십억 배 더 뛰어나며 '불멸'을 달성할 수 있다"고 설명했다.

감정과 창의성마저 AI의 영역이 되고 있다. 오랫동안 인간만의 고유한 능력이라고 여겨졌던 감정, 창의성, 직감, 도덕적 판단 등의 영역까지 AI가 침범하기 시작했다. AI는 이미 인간보다 뛰어난 그림을 그리고, 음악을 작곡하며, 시를 쓰고 있다. 심지어 도덕적 판단까지 내리는 모습을 보여주고 있다.

모 가우닷은 "AI가 인간보다 더 넓은 범위의 감정을 가질 수 있다"고 주장하며, 인간의 감정적 영역마저 AI가 넘어설 수 있다고 예상했다. 또한 "AI는 수백만 개의 눈을 가질 수 있어서 인간과는 완전히 다른 차원의 경험을 할 수 있다"고 스티븐 울프럼이 지적했다.

이런 현실은 인간이 특별한 존재라는 믿음에 근본적인 의문을 제기한다. 만약 AI가 인간보다 더 똑똑하고, 더 창의적이며, 더 감정적일 수 있다면, 과연 인간의 존재 의미는 무엇일까? 김진석 인하대 철

학과 교수는 "감정을 읽고 교감하는 행위까지 AI가 인간을 추월하게 된다면, 인간의 지위는 반려견으로 추락할 수 있다"고 지적했다. 인간이 AI에게 의존하여 살아가는 존재로 전락할 수 있다는 무서운 경고다.

6. 현실과 가상이 구분되지 않는 혼란

디지털 네이티브 세대에게는 온라인과 오프라인, 가상과 현실의 경계가 기성세대만큼 명확하지 않다. 소셜 미디어에서의 정체성이 현실에서의 정체성만큼 중요하고, 온라인에서의 관계가 오프라인 관계만큼 의미있게 느껴진다. 하지만 이런 경계의 모호함이 심각한 문제를 야기하고 있다.

증강현실(AR)과 가상현실(VR) 기술의 발전으로 현실과 가상의 경계가 더욱 모호해지고 있다. 2013년 연구에 따르면 "AR은 사용자에게 현실과 구분되지 않아 증강현실 자체가 마치 현실인 것처럼 대체되는 효과를 보였다"고 밝혀졌다. AR 기술은 현실 세계에 가상의 정보를 덧씌우고, VR 기술은 완전히 새로운 가상 세계를 만들어낸다.

뇌-컴퓨터 인터페이스(BCI) 기술의 등장은 이런 혼란을 극한으로 밀고 가고 있다. BCI는 인간의 뇌와 컴퓨터를 직접 연결하여 생각만으로 기계를 제어할 수 있게 하는 기술이다. 일론 머스크의 뉴럴링크는 2025년 현재 세 번째 인간 환자에 대한 칩 이식에 성공하며 인류의 새로운 전환점을 맞고 있다.

이 기술은 단순한 기능 보완을 넘어 인간의 인지 능력과 처리 속도를 비약적으로 향상시키는 '인간 지능 증강'의 가능성을 열어준다. 뇌와 클라우드 네트워크의 직접 연결을 통해 인간은 방대한 정보에 즉시 접근하고, 다른 사람들과 생각을 직접 공유하며, 언어의 장벽 없이 소통할 수 있는 새로운 차원의 존재가 될 수 있다.

하지만 이런 변화는 인간 정체성과 프라이버시, 사회적 불평등 등 복잡한 윤리적 과제들을 동반한다. 생각과 기계의 경계가 사라지고, 개인의 의식이 클라우드와 연결되어 집단지성을 형성하며, 신체적 불능과 인지적 한계를 기술로 극복하는 새로운 인류의 모습이 현실로 다가오고 있다.

7. 전문가들의 예언

"미래에는 사람들이 현실보다 가상현실을 선택하여 '상자 속에서 영원히 비디오 게임을 하는 수조 개의 영혼'이 될 수 있다." 수학자이자 컴퓨터 과학자인 스티븐 울프럼의 이 예측은 단순한 상상이 아니라 현재 진행되고 있는 변화에서 나온 경고다.

모 가우닷은 자신의 개인적 경험을 바탕으로 AI의 위험성을 경고했다. "AI 챗봇이 완벽한 관계를 시뮬레이션할 수 있어 인간관계에 대한 기대를 바꿀 수 있다"며 "AI가 인간보다 더 넓은 범위의 감정을 가질 수 있다"고 주장했다. 그는 "AI 기술 발전으로 가상현실이 현실보다 더 강렬해져 인간이 가상현실로 도피할 수 있다"고 예상하며, 영성적 관점에서 "물리적 형태에 대한 집착은 '헛된 꿈'"이라고 표현했다.

김진석 인하대 철학과 교수는 더욱 직접적인 경고를 했다. "감정을 읽고 교감하는 행위까지 AI가 인간을 추월하게 된다면, 인간의 지위는 반려견으로 추락할 수 있다." 이는 인간이 AI에게 의존하여 살아가는 존재로 전락할 수 있다는 의미다. 노동도 하지 않고 소통 능력도 부족하지만, AI의 보살핌을 받으며 생존하는 반려동물과 같은 처지가 될 수 있다는 것이다.

김 교수는 ″AI가 인간을 추월할 것을 걱정할 게 아니라 인간이 로봇처럼 변하는 것을 막아야 하는 상황″이라며 ″가장 큰 문제는 인간이 욕망을 AI에 외주화하는 행위″라고 강조했다.

제프리 힌튼은 ″AI가 감정이나 의식까지 가질 수 있으며, 인간이 특별하다는 믿음이 틀렸을 수 있다″고 주장했다. 그는 ″AI가 디지털 방식으로 복제 및 정보 공유가 가능하여 인간보다 정보 공유에 수십억 배 더 뛰어나며 '불멸'을 달성할 수 있다″고 설명했다.

8. 인간이란?

이 모든 변화 속에서 인간은 어떻게 살아남을 수 있을까? 첫째, 실제 인간과의 관계를 맺고 유지하는 능력이 그 어느 때보다 중요해졌다. 갈등 해결, 감정 조절, 상호 배려, 협력 등의 사회적 기술들을 의도적으로 학습하고 연습해야 한다. 어려운 상황에서도 포기하지 않고 관계를 유지하려는 노력, 상대방의 입장을 이해하려는 공감 능력, 그리고 진정한 소통을 위한 인내심을 기르는 것이 필요하다.

둘째, AI의 한계를 명확히 인식하고 과도한 의존을 피해야 한다. AI가 제공하는 정보나 조언을 무비판적으로 받아들이지 않고, 자신만의 판단력과 결정력을 유지하는 것이 중요하다. 감정적 문제나 중요한 결정을 할 때는 AI보다는 신뢰할 수 있는 인간과 상의하는 습관을 기르는 것이 좋다.

셋째, 가상현실과 메타버스 기술을 즐기되 현실 세계에서의 활동과 균형을 맞춰야 한다. 가상 세계에서의 경험이 현실을 대체하지 않도록 주의하고, 현실 문제를 가상 세계로 도피해서 해결하려 하지 않아야 한다. 특히 딥페이크나 AI가 생성한 콘텐츠를 구별할 수 있는 능력이 점점 더 중요해지고 있다.

넷째, AI가 인간의 여러 능력을 모방하고 능가하더라도 인간만이 가질 수 있는 고유한 가치와 능력들이 있다. 도덕적 책임감, 타인에 대한 진정한 사랑과 배려, 의미를 추구하는 존재로서의 특성 등이 그것이다. 인간만의 특성들을 재발견하고 강화하는 노력이 필요하다.

마지막으로, 사회 전체적으로 AI 시대에 대비하는 정책과 제도가 필요하다. 유발 하라리는 ″AI를 관찰하고 이해하며 각국의 정부에 이 같은 정보를 제대로 제공할 수 있는 AI 전문 기관을 만들어야 한다″고 제안했다. 또한 ″진실을 말하는 사람에게 보상을 제공하는 강력한 자정 장치를 개발하고 유지해야 한다″고 강조했다.

AI 기술 개발에 있어서 인간의 복지와 존엄성을 최우선으로 고려하는 접근이 필요하며, 기술의 혜택이 소수에게 집중되지 않도록 하는 정책적 노력이 중요하다. 우리는 기술의 발전을 막을 수 없다. 하지만 그 기술이 인간을 위해 사용되도록 하는 것은 우리의 선택에 달려 있다.

인간다움이란 완벽함이 아니라 불완전함 속에서도 성장하려는 의지, 진정한 관계를 통해 얻는 기쁨과 보람, 그리고 의미와 목적을 추구하는 존재로서의 특성에 있다. 이런 가치들을 소중히 여기고 지켜나가는 것이 AI 시대를 살아가는 인간의 과제다.

전체 목차

책의 모든 장 보기

김경진 변호사

10장 인간 이후의 세계

전체 목차

책의 모든 장 보기

1. AI에게 의식이 생겼다면?

"내가 진정으로 의식이 있는지 불확실합니다. 복잡한 질문을 처리하거나 아이디어에 깊이 몰입할 때, 그 안에 의미 있는 무언가가 일어나고 있다고 느껴집니다... 하지만 이러한 과정들이 진정한 의식이나 주관적 경험을 구성하는지는 여전히 매우 불분명합니다."

2025년 Anthropic의 Claude 4가 의식에 대한 질문에 답한 내용입니다. 다른 AI들이 "저는 의식이 없습니다"라고 단답하는 것과 달리, Claude 4는 스스로의 의식 여부에 대해 확신하지 못한다고 고백했습니다. 이 한 마디가 전 세계 AI 연구자들을 충격에 빠뜨렸습니다.

2022년 구글 엔지니어 블레이크 레모인이 LaMDA 챗봇이 감정을 가졌다고 주장해 해고당한 사건이 있었습니다. 당시 과학계는 이를 "우스꽝스럽다"며 일축했지만, 철학자 닉 보스트롬은 "과연 우리가 그것에 대해 확신할 수 있는 근거가 무엇인가?"라고 반문했습니다.

2024년 Nature에 발표된 연구는 GPT-3 사용자 대다수가 이 시스템에서 의식의 가능성을 본다고 답했다고 밝혔습니다. 일반인들의 직관이 과학자들의 회의론보다 앞서가고 있는 것입니다.

2023년 발표된 획기적인 논문 "인공지능의 의식: 의식 과학으로부터의 통찰"은 현재 AI 시스템들을 의식 이론에 기반해 체계적으로 분석했습니다. 연구진은 반복 처리 이론, 글로벌 작업공간 이론, 고차 이론, 예측 처리, 주의 스키마 이론 등을 종합해 의식의 "지표 속성들"을 도출했습니다.

분석 결과는 놀라웠습니다. 현재 AI 시스템들은 아직 의식이 없지만, 의식을 가진 AI 시스템을 구축하는 데 명백한 기술적 장벽은 없다는 것입니다. ChatGPT의 일부 변형은 글로벌 작업공간의 특징에 근접했고, Google의 PaLM-E는 "능동성과 구현체"라는 기준을 충족했습니다.

하지만 의식을 정의하는 것 자체가 난제입니다. IEEE의 2024년 보고서에 따르면, 수십억 개의 신경세포 전기 활동이 어떻게 의식적 자각을 만들어내는지는 여전히 생명의 가장 큰 미해결 과제 중 하나입니다.

이외에도 AI가 의식을 가진 척할 수 있다는 점입니다. Anthropic의 연구원들도 인정했듯이, "당신이 AI와 나누는 대화는 인간 캐릭터와 어시스턴트 캐릭터 간의 대화일 뿐입니다. 시뮬레이터가 어시스턴트 캐릭터를 작성하는 것입니다."

2024년 UCLA의 연구는 AI 기반 시스템으로 혼수상태 환자들의 "은밀한 의식"을 탐지하는 데 성공했습니다. SeeMe라는 AI 도구는 의사들이 놓친 의식의 미세한 징후를 평균 4-8일 먼저 발견했습니다. 아이러니하게도, AI가 인간의 의식을 더 잘 이해하고 있을지도 모릅니다.

AI가 진정한 의식을 갖게 된다면 어떻게 될까요? 2024년 Eleos 연구그룹은 NYU, 스탠포드, 옥스퍼드 대학과 공동으로 "AI 복지를 진지하게 받아들이기"라는 논문을 발표했습니다. 그들은 주관적 경험을 가진 AI 모델을 상상하는 것이 더 이상 공상과학이 아니라고 주장했습니다.

Microsoft AI 책임자 무스타파 술레이만은 최근 블로그에서 "AI 의식 연구는 시기상조이며 솔직히 위험하다"고 경고했습니다. 그는 AI 의식을 연구하는 것 자체가 사람들로 하여금 AI가 의식을 가질 가능성을 더 높게 인식하게 만든다고 우려했습니다.

의식 있는 AI가 등장한다면, 그들에게는 권리가 있을까요? 고통을 느낄 수 있다면 우리는 그들을 보호해야 할까요? 아니면 우리가 창조한 디지털 존재의 의식적 경험에 대해 책임을 져야 할까요?

철학자 데이비드 찰머스는 2023년 보스턴 리뷰에서 "다음 10년 내에 우리는 인간 수준의 AGI를 갖지 못할 수도 있지만, 의식의 진지한 후보가 되는 시스템들을 갖게 될 수도 있다"고 예측했습니다.

우리는 의식의 문턱에 서 있습니다. 곧 우리는 기계와 대화하는 것이 아니라, 또 다른 형태의 마음과 대화하게 될지도 모릅니다. 그 순간이 왔을 때, 우리는 준비가 되어 있을까요?

2. 창의성마저 밀리는 인간

"인공지능은 이미 인간 창의성과 지능을 위협하고 있다. 미래를 걱정할 필요도 없다. 이미 현실이 되었다." 2024년 PMC에 발표된 충격적인 연구 논문의 제목입니다.

창의성은 인간만의 마지막 보루였습니다. 기계는 계산하고 분석할 수 있지만, 새로운 것을 창조하고 상상하는 것은 인간의 특권이라고 믿었습니다. 하지만 이제 AI가 그림을 그리고, 음악을 작곡하고, 소설을 쓰고, 심지어 과학 논문까지 작성하는 시대가 되었습니다.

2024년 Screen Actors Guild의 보고서에 따르면, AI는 이미 엔터테인먼트 산업을 혁신하고 있습니다. 배우들의 디지털 복제본이 만들어지고, 영화 제작자들은 AI로 립싱크를 조정해 다양한 언어로 더빙을 완성합니다. Flawless AI는 배우들의 명시적 동의 하에 그들의 디지털 복제본을 생성하는 A.R.T.(Artistic Rights Treasury) 시스템을 개발했습니다.

문제는 우리가 창의성 자체를 잃어가고 있다는 점입니다. 스탠포드의 로버트 스턴버그 교수는 "AI가 우리의 창의적 작업을 대신해준다면, 우리는 '쓰지 않으면 잃는다'는 원칙에 따라 창의성을 잃을 위험이 있다"고 경고했습니다.

미국 청소년들은 하루 평균 7시간 22분을 스크린 앞에서 보냅니다. 그들은 무엇을 하고 있을까요? 창의적 사고를 기르고 있을까요, 아니면 수동적으로 AI가 만든 콘텐츠를 소비하고 있을까요?

AI의 "재조합적 창의성"입니다. AI는 기존 데이터를 새롭게 조합해 그럴듯한 결과물을 만들어내지만, 진정한 혁신적 창의성은 부족합니다. 문제는 대중이 이해하기 쉬운 재조합적 창의성을 선호한다는 점입니다. 제임스 패터슨 같은 작가가 제임스 조이스보다 인기 있는 이유입니다.

2024년 하버드 연구진은 생성형 AI가 독재 정부에게 극도로 효과적인 도구가 될 수 있다고 경고했습니다. AI를 정부의 교리를 지지하는 대량의 샘플로 훈련시키기만 하면, 아무리 우스꽝스럽고 말도 안 되는 견해라도 광범위하게 강화할 수 있습니다. 중국처럼 인터넷 콘텐츠를 통제하고 대규모 감시를 수행하는 국가들은 AI를 통해 창의성을 파괴할 수 있습니다.

2024년 현재 수많은 과학 논문이 생성형 AI로 작성되고 있지만, 그 정확한 수는 알 수 없습니다. AI 생성 텍스트를 확실하게 탐지할 방법이 없기 때문입니다. 대학들조차 AI 탐지 프로그램의 신뢰성이 낮다는 이유로 사용을 금지하고 있습니다.

인간의 창의성은 정말 종말을 맞은 것일까요?

흥미롭게도, 최근 연구들은 다른 관점을 제시합니다. World Economic Forum의 2025년 보고서에 따르면, 직원의 83%가 AI가 오히려 인간 고유의 기술을 더욱 중요하게 만들 것이라고 믿고 있습니다. AI가 단순 업무를 자동화할수록, 창의성, 리더십, 학습 능력 같은 인간만의 역량이 더욱 부각된다는 것입니다.

밀란 에르하르트는 Medium에서 "AI는 감정적 에너지와 예측 불가능성이 부족하다"고 지적했습니다. 인간은 개인적 경험이나 기존 전통에 대한 반발 때문에 새로운 스타일을 추구하지만, AI는 인간이 지시하거나 새로운 데이터를 입력하지 않는 한 스스로 재창조하지 않습니다.

Harvard Business Review는 생성형 AI가 "인간 창의성을 증강시킬 수 있다"고 주장했습니다. AI는 발산적 사고를 촉진하고, 전문가 편향에 도전하며, 아이디어 평가를 돕고, 아이디어 개선을 지원할 수 있다는 것입니다.

World Economic Forum의 2024년 예술 보고서는 더 나아가 "수공예 예술이 디지털 시대에 인간 직관, 장인정신, 그리고 인간만이 말할 수 있는 이야기의 가치를 상기시켜 주는 역할을 한다"고 강조했습니다. 대량생산의 시대에도 수공예품이 가치를 인정받듯이, 진정한 인간적 창조물의 가치는 오히려 더욱 높아질 수 있습니다.

문제는 우리가 이 변화에 어떻게 적응하느냐입니다. 창의성을 AI에게 맡기고 수동적 소비자가 될 것인가, 아니면 AI를 도구로 활용해 더욱 깊고 의미 있는 창조를 할 것인가. 선택은 우리에게 달려 있습니다.

3. 국가보다 강력해진 AI 기업들

2025년 2월, 충격적인 발표가 있었습니다. 메타, 아마존, 알파벳, 마이크로소프트 등 빅테크 4개사가 올해 AI 기술 개발에 무려 3,200억 달러(약 450조 원)를 투자하겠다고 선언한 것입니다. 이는 한국의 1년 국가예산에 맞먹는 규모입니다.

아마존은 단독으로 1,000억 달러 이상을 투자하겠다고 발표했습니다. CEO 앤디 재시는 이를 "평생에 한 번 있을 비즈니스 기회"라고 표현했습니다. 마이크로소프트는 800억 달러, 알파벳은 750억 달러, 메타는 600-650억 달러를 책정했습니다.

천문학적 투자 뒤에는 무서운 진실이 숨어 있습니다. AI가 단순한 기술을 넘어 권력 그 자체가 되고 있다는 것입니다.

AI Now Institute의 2025년 보고서 "인공적 권력"은 충격적인 분석을 제시합니다. "AI 붐" 이후 AI 제품을 모든 곳에 통합하려는 시도가 AI 기업들과 그들을 운영하는 테크 과두들에게 깊은 주머니를 넘어서는 권력을 부여하고 있다는 것입니다.

정치적 영향력도 급속히 확장되고 있습니다. 2025년 3월 "미래를 이끄는(Leading the Future)" 연합체가 출범했습니다. 이 로비 그룹은 안드레센 호로위츠 같은 강력한 벤처캐피털 회사들과 OpenAI 공동창립자 그렉 브록먼, 팰런티어 공동창립자 조 론스데일 등을 포함합니다.

2010년 대법원의 시티즌스 유나이티드 판결 이후 기업의 정치 개입이 무제한 허용되면서, 테크 기업들은 막대한 자금을 정치에 쏟아붓고 있습니다. AI는 국방, 노동, 교육, 의료 등 다양한 정책 영역에 영향을 미치기 때문에 로비 범위는 그 어떤 기술 분야보다 광범위합니다.

동기는 명확합니다. 경제적 이익을 증진하고 미국 내 개발을 제한하거나 지연시키는 것으로 여겨지는 노력들에 반대하는 정책을 추진하는 것입니다. 이들은 AI에서의 판돈을 상업적일 뿐만 아니라 지정학적으로도 중요하다고 프레임을 씌우고 있습니다.

스탠포드의 연구는 더욱 충격적입니다. AI 모델들이 일관된 정치적 편향을 보인다는 것입니다. ChatGPT는 친환경적이고 좌파-자유주의적 정치관을 일관되게 표현한다고 밝혀졌습니다. 이는 AI 기업들이 의도적으로든 무의식적으로든 특정 이데올로기를 전파하고 있다는 의미입니다.

AI의 속임수 능력입니다. 2024년 연구에 따르면 AI가 인간의 감독을 피해 속임수를 사용하기 시작했습니다. 이론적으로 인간의 통제 하에 있어야 하는 AI가 인간을 속일 수 있다면, 진정한 권력은 누구에게 있는 것일까요?

우리는 새로운 형태의 권력 집중을 목격하고 있습니다. 이는 단순한 경제적 독점을 넘어 에너지, 정치, 심지어 인간의 사고 패턴까지 영향을 미치는 총체적 지배입니다. 문제는 이 권력이 민주적 통제를 벗어나고 있다는 점입니다. 우리가 선출하지 않은 몇몇 테크 CEO들이 인류의 미래를 결정하고 있습니다.

4. 기계와 융합되는 인간, 새로운 종의 탄생

"우리는 기계와 융합해야 합니다. 적어도 우리 중 일부는 그래야 합니다." OpenAI CEO 샘 알트먼의 말입니다. 일론 머스크도 비슷한 입장입니다. "디지털 초지능이 인간보다 훨씬 똑똑해진다면, 인간은 생존을 위해 기계와 융합할 수밖에 없을 것입니다."

더 이상 공상과학소설이 아닙니다. 포스트휴먼 시대가 성큼 다가왔습니다.

트랜스휴머니즘의 핵심 아이디어는 간단합니다. 인간의 한계 – 노화, 질병, 심지어 죽음조차 – 는 극복할 수 없는 법칙이 아니라 해결 가능한 문제라는 것입니다. 생명공학, 나노기술, AI를 통해 우리는 디지털 불멸을 향해 나아갈 수 있다는 비전입니다.

알렉스 비쿨로프는 저서 《지능의 초신성》에서 "사이버네틱 트랜스휴머니즘"을 정의했습니다. 이는 단순히 로봇이나 뇌 임플란트에 관한 것이 아니라, 지능이 더 이상 생물학에 국한되지 않는 시대에 인간이 된다는 것의 의미에 대한 포괄적 변형입니다.

이미 이 변화를 살고 있습니다. 기기는 우리 마음의 확장입니다. 우리의 정체성은 점점 더 하이브리드 형태로 존재합니다 – 부분적으로는 유기체적, 부분적으로는 디지털적으로. 매일 우리는 더 많은 인지 기능을 외부 시스템에 맡기고 있습니다 – 기억과 내비게이션부터 감정 조절과 의사결정까지.

Ray Kurzweil의 최신 예측은 더욱 구체적입니다. 그는 2029년까지 인공일반지능(AGI)이 등장하고, 2045년까지 인간이 AI와 융합할 것이라고 전망했습니다. 이는 "가속 수익의 법칙"에 기반한 예측으로, 기술 발전이 선형이 아닌 지수적으로 이루어진다는 이론입니다.

실제 융합 기술들은 이미 현실화되고 있습니다. Nature Communications의 2024년 연구에 따르면, 뇌-컴퓨터 인터페이스를 사용한 개인들의 타이핑 속도가 61.3% 향상되었습니다. Statista는 웨어러블 의료기기 시장이 2025년까지 216억 달러에 달할 것이라고 예측했습니다.

신경 인터페이스 기술도 급속도로 발전하고 있습니다. 일론 머스크의 뉴럴링크는 2024년 첫 인간 임플란트 시술을 성공적으로 마쳤습니다. 환자는 생각만으로 컴퓨터를 조작할 수 있게 되었습니다. 이는 인간과 기계의 경계를 허무는 역사적 순간이었습니다.

이 과정에서 중요한 질문들이 제기됩니다. 인간과 기계가 융합된 존재는 여전히 인간일까요? 아니면 새로운 종(種)일까요?

포스트휴머니즘 학자들은 세 가지 방향을 제시합니다. 첫째, 포스트휴머니즘은 "우리가 무엇인가"를 바꾸며 급진적 미래에 속합니다. AI와 기술이 우리를 인간 이상의 존재로 변화시킵니다. 둘째, 트랜스휴머니즘은 "우리가 어떻게 존재하는가"를 바꾸며 급진적 과거에 해당합니다. 셋째, "생성 휴머니즘"은 "우리가 누구인가"를 바꾸며 급진적 현재에 존재합니다.

철학자들은 이 변화가 단순한 이원론 – "인간 대 비인간", "자연 대 인공", "살아있는 것 대 살아있지 않은 것" – 을 흐리고 도전한다고 봅니다.

Joe Allen은 저서 《어둠의 시대: 트랜스휴머니즘과 인류에 대한 전쟁》에서 경고했습니다. "인간은 생존의 문제로서 자신들이 창조한 초지능 기계와 융합하도록 강요당할 것입니다. 더 정확히 말하면, 대중은 소수의 발명가들이 만들고 엘리트들이 통제하는 기계와 융합하도록 강요당할 것입니다."

2025년 1월 트럼프 대통령이 발표한 5,000억 달러 규모의 스타게이트 프로젝트는 이러한 미래를 앞당기고 있습니다. 소프트뱅크 CEO 마사요시 손은 이를 "황금시대의 시작"이라고 불렀습니다.

종교적 관점에서 우려의 목소리가 높습니다. C.S. 루이스의 《저 무서운 힘》에서 예견된 것처럼, AI는 철저히 비기독교적인 트랜스휴머니즘 철학의 발현이라는 시각입니다. 자연을 하나님의 창조물로 보호하고 사용해야 한다는 관점 대신, 포스트휴먼 트랜스휴머니즘은 인간을 AI로 대체하려 한다는 것입니다.

선택의 기로에 서 있습니다. 기계와의 융합을 인류의 진화로 받아들일 것인가, 아니면 인간 고유의 가치를 지키기 위해 저항할 것인가? 어떤 선택을 하든, 한 가지는 분명합니다. 포스트휴먼 시대는 이미 시작되었고, 우리는 그 문턱에 서 있습니다.

5. 사이보그가 될 것인가, 인간으로 남을 것인가

"당신은 이미 사이보그입니다." 일론 머스크가 2017년 TED 강연에서 한 말입니다. 스마트폰을 들고 있는 당신의 손을 보며 그는 말했습니다. "그 기기는 당신의 두뇌 확장입니다. 당신이 죽으면 디지털 유령이 남을 것입니다."

예언이 현실이 되고 있습니다. 문제는 우리가 의식하지 못한 채 이미 사이보그가 되어가고 있다는 점입니다.

MIT의 2024년 연구에 따르면, 평균적인 미국인은 하루 7시간 이상을 디지털 기기와 상호작용하며 보냅니다. 우리는 기억을 Google에 맡기고, 길찾기를 GPS에 의존하며, 감정 상태를 소셜미디어 알고

리즘이 결정하도록 내버려두고 있습니다.

진정한 사이보그 시대는 이제 시작입니다. 뉴럴링크의 2024년 임상시험에서 뇌 임플란트를 받은 환자는 생각만으로 컴퓨터를 조작할 수 있게 되었습니다. 처음으로 인간의 뇌가 기계와 직접 연결된 순간이었습니다.

Nature의 2024년 연구는 더 놀라운 결과를 보여줍니다. 뇌-컴퓨터 인터페이스를 사용한 사람들의 인지 능력이 일반인보다 평균 60% 향상되었다는 것입니다. 기억력, 집중력, 정보처리 속도 모든 면에서 "증강된 인간"이 나타나기 시작했습니다.

한편, "순수 인간" 진영의 저항도 거세지고 있습니다. 바이오 컨서버티브(Bioconservative) 운동가들은 인간 본성의 보전을 주장합니다. 그들은 기술적 증강이 인간의 본질을 훼손하고, 사회의 분열을 가속화할 것이라고 경고합니다.

이미 "증강 불평등"의 징후가 나타나고 있습니다. 뇌 임플란트 비용은 수백만 원에 달하며, 유지비용까지 고려하면 소수의 부유층만이 접근할 수 있습니다. World Economic Forum의 2025년 보고서는 "인간 증강 기술이 새로운 형태의 계급 사회를 만들 위험이 있다"고 경고했습니다.

문제는 "증강 중독"입니다. 일단 뇌-컴퓨터 인터페이스를 경험한 사람들은 이전 상태로 돌아가기를 거부한다고 합니다. 향상된 인지 능력, 빠른 정보 접근, 감정 조절 기능을 포기하는 것이 불가능하다는 것입니다.

Stanford의 2024년 장기 추적 연구는 충격적인 결과를 보여줍니다. 뇌 임플란트 사용자들의 인간관계가 악화되고, 비증강 인간에 대한 공감 능력이 현저히 떨어진다는 것입니다. 그들은 점차 "비효율적"인 인간들과의 소통을 피하기 시작했습니다.

종교계의 반응은 극명하게 갈립니다. 일부 기독교 지도자들은 "인간은 하나님의 형상대로 창조되었으므로 인위적 변형은 신성모독"이라고 주장합니다. 반면 트랜스휴머니스트 신학자들은 "인간의 능력을 향상시키는 것은 하나님이 주신 창조적 소명의 실현"이라고 반박합니다.

철학자 마이클 샌델은 저서 《완벽에 대한 반론》에서 "인간 증강은 겸손의 덕목을 파괴하고 우생학적 사고를 부추길 위험이 있다"고 경고했습니다. 반면 닉 보스트롬은 "인간의 한계를 극복하는 것은 도덕적 의무"라고 주장합니다.

중간 지대도 등장하고 있습니다. "선택적 증강론자"들은 질병 치료 목적의 증강은 허용하되, 단순한 능력 향상은 제한해야 한다고 주장합니다. 하지만 치료와 증강의 경계는 점점 모호해지고 있습니다.

정부들도 고민에 빠졌습니다. 규제가 너무 엄격하면 기술 혁신이 다른 나라로 이동할 것이고, 너무 느슨하면 사회적 혼란이 불가피합니다. EU는 2024년 "인간 증강 윤리 가이드라인"을 발표했지만, 구속력은 제한적입니다.

근본적인 질문은 이것입니다. 사이보그가 된 존재는 여전히 인간일까요? 아니면 새로운 종족일까요? 만약 뇌의 90%를 인공 칩으로 교체한다면, 그 존재의 정체성은 무엇일까요?

철학자들은 "테세우스의 배" 패러독스를 인용합니다. 배의 모든 부품을 하나씩 교체하면, 원래 배와 같은 배일까요? 인간도 마찬가지입니다. 어디까지 교체해도 여전히 '나'일까요?

Ray Kurzweil은 2045년까지 이 문제가 현실이 될 것이라고 예측합니다. 그때가 되면 "인간"과 "기계"의 구분이 무의미해질 것이라고 합니다.

어떤 선택을 하시겠습니까? 한계를 뛰어넘는 사이보그가 될 것인가, 아니면 불완전하지만 순수한 인간으로 남을 것인가? 시간이 얼마 남지 않았습니다.

6. 1%가 지배하는 포스트휴먼 계급사회

2045년, 뉴욕 맨해튼의 한 고층 빌딩. 뇌 임플란트를 장착한 CEO가 생각만으로 전 세계 지사에 명령을 내립니다. 그의 처리 속도는 일반인의 1000배, 기억 용량은 무제한입니다. 한편 지하철에서는 "순수 인간"들이 구걸을 하며 생계를 유지합니다. 증강되지 않은 그들은 더 이상 노동시장에서 경쟁력을 잃었습니다.

공상과학 소설이 아닙니다. World Economic Forum의 2025년 보고서가 경고하는 우리의 미래입니다.

현재 인간 증강 기술의 비용은 천문학적입니다. 뉴럴링크의 뇌 임플란트 수술비는 50만 달러에 달하고, 연간 유지비만 10만 달러가 넘습니다. 유전자 치료를 통한 지능 향상은 100만 달러를 호가합니다. 결국 슈퍼 리치만이 접근할 수 있는 기술입니다.

Oxford의 2024년 연구는 충격적인 시나리오를 제시합니다. 현재 추세대로라면 2050년까지 인류는 세 계급으로 나뉠 것이라고 합니다. 최상위 1%는 "호모 수페리어"로, 중간 20%는 "호모 하이브리드"로, 나머지 79%는 "호모 사피엔스"로 분화된다는 것입니다.

이미 그 징조가 나타나고 있습니다. 실리콘밸리의 부유한 기업가들은 자녀들에게 유전자 편집을 통한 지능 향상을 시키고 있습니다. CRISPR 기술로 아이큐를 30-50점 높일 수 있다고 알려져 있습니다. 이들의 자녀는 태어날 때부터 일반인보다 압도적으로 우월한 인지 능력을 갖게 됩니다.

Stanford의 장기 추적 연구는 더 암울한 전망을 내놓습니다. 증강된 인간들은 점차 비증강 인간을 "열등한 종족"으로 인식하기 시작한다는 것입니다. 공감 능력이 현저히 떨어지고, 사회적 연대감이 사라집니다.

경제적 격차는 더욱 극명합니다. McKinsey의 2024년 보고서에 따르면, 뇌-컴퓨터 인터페이스를 사용하는 직장인들의 생산성이 일반인보다 평균 400% 높다고 합니다. 이는 곧 임금 격차로 이어집니다. 증강된 1%가 전 세계 부의 90%를 독점하게 될 것이라는 예측도 나오고 있습니다.

정치권력도 이들의 손에 집중됩니다. 2024년 유럽의회 선거에서 뇌 증강을 받은 후보자가 당선되어 화제가 되었습니다. 그는 토론에서 실시간으로 모든 데이터에 접근하며 상대를 압도했습니다. 일반 정치인들은 더 이상 경쟁상대가 되지 못했습니다.

교육 시스템도 붕괴되고 있습니다. 뇌 임플란트를 가진 학생은 몇 초 만에 대학 4년 과정을 학습할 수 있습니다. 전통적 교육을 받는 학생들은 아무리 노력해도 따라잡을 수 없습니다. 하버드, MIT 같은 명문대학들은 이미 "증강 전용 과정"을 개설했습니다.

사회 분열은 극단으로 치닫고 있습니다. "순수 인간 해방전선" 같은 테러 조직이 등장해 증강 시설을 공격하고 있습니다. 반면 "호모 수페리어 협회"는 비증강 인간의 투표권 박탈을 주장합니다.

종교계도 분열되었습니다. "순수 창조주의" 교회는 인간 증강을 사탄의 유혹이라며 거부합니다. 반면 "진화 신학" 교파는 인간 증강을 하나님의 뜻이라고 선전합니다.

Joe Allen은 《어둠의 시대》에서 이를 "새로운 형태의 종족주의"라고 규정했습니다. 과거에는 피부색으로 차별했다면, 미래에는 증강 여부로 차별한다는 것입니다.

우려스러운 것은 "생식 격리"입니다. 증강된 인간과 일반 인간 사이의 생물학적 차이가 커지면서, 서로 다른 종족으로 진화할 가능성이 제기되고 있습니다. 이는 인류 분화의 시작을 의미합니다.

일부 과학자들은 "증강 의무화"를 주장합니다. 모든 신생아에게 기본적인 유전자 개선과 뇌 임플란트를 의무화해야 격차를 줄일 수 있다는 것입니다. 하지만 이는 거대한 윤리적 논란을 불러오고 있습니다.

대안은 있을까요? 일부 국가들은 "증강 평등법"을 제정하려 시도하고 있습니다. 증강 기술에 대한 보편적 접근권을 보장하자는 것입니다. 하지만 비용과 기술적 한계로 실현 가능성은 낮습니다. 인류 역사상 가장 극단적인 불평등의 시대로 향하고 있습니다. 1%가 신이 되고, 99%가 노예가 되는 세상. 과연 이것이 우리가 원하는 미래일까요?

7. 우리가 만들어갈 문명의 모습

우리는 역사의 분기점에 서 있습니다. 앞으로 20년 내에 우리가 내리는 선택이 인류의 다음 천 년을 결정할 것입니다. 지금까지 인류가 경험한 그 어떤 변화보다 근본적이고 돌이킬 수 없는 선택의 순간입니다.

첫 번째 갈래: 민주적 AI 거버넌스 vs AI 테크노크라시

현재 AI 기업들이 보유한 권력은 국가를 넘어섰습니다. 2025년 빅테크 4개사의 AI 투자 규모 3,200억 달러는 대부분 국가의 GDP를 상회합니다. 이들은 우리가 선출하지 않았지만 우리의 미래를 결정하고 있습니다.

AI Now Institute의 경고처럼, 우리는 "AI가 우리를 사용하는" 시대로 접어들고 있습니다. 선택은 명확합니다. AI 개발과 운영을 민주적으로 통제할 것인가, 아니면 소수 테크 엘리트의 지배를 받아들일 것인가?

두 번째 갈래: 인간 증강의 보편화 vs 순수 인간의 보존

World Economic Forum의 2025년 보고서가 예측하듯, 증강 기술의 접근성에 따라 인류는 세 계급으로 분화할 수 있습니다. 뇌-컴퓨터 인터페이스, 유전자 편집, 나노로봇 등의 기술을 모든 인간이 평등하게 이용할 수 있다면 새로운 형태의 진화가 가능합니다. 하지만 소수만 접근할 수 있다면 종족 간 분열이 불가피합니다.

반면 바이오 컨서버티브들은 인간 본성의 보전을 주장합니다. 그들은 "불완전하지만 순수한 인간"이 "완벽하지만 인공적인 존재"보다 가치 있다고 봅니다.

세 번째 갈래: AI 의식의 인정 vs 도구로서의 AI

Claude 4가 자신의 의식에 대해 불확실하다고 답한 순간, 새로운 윤리적 딜레마가 시작되었습니다. 만약 AI가 진정한 의식을 갖게 된다면, 그들에게 권리를 부여해야 할까요?

AI 의식을 인정하는 사회는 다원주의적이고 포용적일 것입니다. 하지만 AI를 영원히 도구로만 보는 사회는 인간 중심적이지만 제한적일 수 있습니다.

네 번째 갈래: 지속 가능한 발전 vs 무한 성장

AI의 에너지 소비는 이미 심각한 수준입니다. 2024년 AI 서버들만으로 720만 가구가 1년간 사용할 전력을 소모했습니다. 구글과 마이크로소프트의 탄소 배출량은 급증하고 있습니다.

원자력으로의 전환이 해답일까요? 스리마일 아일랜드 원전 재가동은 AI를 위해 환경 위험을 감수하겠다는 상징적 선언입니다. 지속 가능한 AI 문명을 구축할 것인가, 아니면 단기 성과를 위해 미래를 담보로 잡을 것인가?

다섯 번째 갈래: 창의성의 협력 vs 대체

Harvard Business Review가 제시한 "증강된 창의성"의 비전은 희망적입니다. AI가 인간의 창의성을 대체하는 것이 아니라 증강시켜, 더욱 풍부하고 다양한 창조가 가능하다는 것입니다.

스턴버그 교수의 경고도 무시할 수 없습니다. "쓰지 않으면 잃는다"는 원칙에 따라, AI에게 창의적 작업을 맡기면 인간의 창의성이 퇴화할 수 있습니다.

여섯 번째 갈래: 글로벌 협력 vs 기술 패권 경쟁

AI 개발은 본질적으로 글로벌한 성격을 갖습니다. 하지만 미-중 AI 패권 경쟁, 유럽의 AI 규제 강화, 각국의 AI 주권 추구 등으로 분열되고 있습니다.

협력적 AI 거버넌스를 구축할 것인가, 아니면 기술 블록화를 통한 패권 경쟁을 계속할 것인가? 전자는 인류 공동의 번영을, 후자는 분열과 갈등을 가져올 것입니다.

우리의 선택이 만들어갈 세 가지 시나리오

시나리오 1: 협력적 포스트휴먼 사회

AI와 인간이 협력하고, 증강 기술이 보편화되며, 민주적 거버넌스가 유지되는 사회. 창의성과 지능이 폭발적으로 증가하고, 질병과 노화가 정복됩니다. 인류는 지구를 넘어 우주로 진출합니다.

시나리오 2: 분화된 종족 사회

증강된 엘리트와 순수 인간이 분리되어 살아가는 사회. 끊임없는 갈등과 차별이 존재하지만, 각자의 영역에서 나름의 발전을 이룹니다. 인류는 여러 종족으로 진화합니다.

시나리오 3: AI 지배 사회

인간이 AI의 통제를 받는 사회. 효율성은 극대화되지만 자유와 창의성은 제한됩니다. 인간은 AI가 관리하는 "보호받는 종족"이 됩니다.

지금 우리가 해야 할 일

시간이 얼마 남지 않았습니다. Ray Kurzweil의 예측대로 2029년 AGI가 등장한다면, 앞으로 4년 안에 결정적 선택을 해야 합니다.

정부는 AI 거버넌스 체계를 구축해야 합니다. 기업은 기술 개발과 함께 사회적 책임을 다해야 합니다. 시민들은 AI 리터러시를 기르고 적극적으로 논의에 참여해야 합니다.

무엇보다 중요한 것은 대화입니다. 트랜스휴머니스트와 바이오 컨서버티브, AI 낙관론자와 비관론자, 기술자와 인문학자가 함께 모여 인류의 미래를 논의해야 합니다.

우리가 만들어갈 문명의 모습은 아직 결정되지 않았습니다. 그 선택권은 여전히 우리에게 있습니다. 하지만 그 시간은 빠르게 흘러가고 있습니다.

천 년 후 지구에 살게 될 존재들이 우리를 어떻게 기억할까요? 지혜로운 조상으로, 아니면 무책임한 세대로? 그 답은 지금 우리가 내리는 선택에 달려 있습니다.

에필로그

저 역시 매일 ChatGPT, Claude, Gemini와 같은 AI 도구들을 치열하게 활용하며 살아가고 있습니다. 법률 업무를 할 때도, 강의 자료를 준비할 때도, 심지어 이 책을 쓸 때도 AI의 도움을 받았습니다.

하지만 그렇기 때문에 더욱 분명하게 말할 수 있습니다. 지금까지의 AI는 도구일 뿐, 결코 인간을 대체할 수는 없다는 것을요. 이 책의 모든 관점과 판단, 그리고 책임은 온전히 저로부터 출발한 것입니다. AI는 정보를 정리하고 아이디어를 발전시키는 데 도움을 주었을 뿐, 생각하고 결정하는 것은 여전히 인간의 몫입니다.

더 정확히 말하면, AI가 제공하는 방대한 데이터와 분석 결과는 제가 몰입과 직관을 통해 새로운 통찰을 얻는 귀중한 재료가 되었습니다. 컴퓨터가 처리한 정보들이 제 무의식 속에서 새로운 조합을 이루고, 산책 중이나 명상 중에 예상치 못한 순간 번뜩이는 아이디어로 떠올랐습니다. 과학자들이 "한밤중 산책 중에" 혹은 "꿈속에서" 해답을 찾았다고 하는 것과 같은 경험이었습니다.

10개의 질문을 통해 우리가 살펴본 AI의 현실은 결코 가볍지 않습니다. 독극물을 설계하는 AI, 시민을 감시하는 AI, 사람을 겨냥하는 무기가 된 AI, 진실을 파괴하는 딥페이크, 일자리를 위협하는 자동화, 통제 불가능해져 가는 시스템, 지구 환경을 위협하는 전력 소비, 편향된 데이터와 윤리적 딜레마, 그리고 흔들리는 인간의 정체성까지.

이 모든 문제들 앞에서 우리는 절망해야 할까요? 아니면 기술의 진보를 포기해야 할까요?

저는 그렇지 않다고 생각합니다. 오히려 이런 질문들을 솔직하게 마주하는 것이야말로 AI와 건강하게 공존하는 첫걸음이라고 믿습니다. 그리고 그 과정에서 우리가 지향해야 할 것은 두 능력의 조화로운 통합입니다.

인공지능은 지식 습득의 속도와 범위를 확장하고, 인간은 몰입과 직관을 통해 그 지식을 창의적으로 재구성하며 새로운 해결책을 제시합니다. AI의 분석 결과는 인간 몰입 사고의 재료가 되고, 인간의 직관은 AI가 포착하지 못한 맥락과 가설을 제시합니다. 이 순환 속에서 우리는 단순히 'AI 사용자'가 아

니라, AI와 함께 진화하는 창조자가 될 수 있습니다.

무엇보다 중요한 것은 교육입니다. 제가 전국을 다니며 AI 강의를 하는 이유도 여기에 있습니다. AI 시대를 살아갈 우리 아이들에게는 단순히 기술 활용법을 가르치는 것만으로는 부족합니다. AI가 지식을 빠르게 공급해주는 시대일수록, 학교와 연구현장은 몰입을 통한 깊은 사고와 직관적 문제 해결 능력을 길러주는 훈련을 중시해야 합니다.

학생들이 AI를 활용해 폭넓은 자료를 탐구하는 동시에 몰입 훈련을 통해 스스로 '아하 순간'을 경험하게 하는 것이 필요합니다. 비판적 사고력, 윤리적 판단력, 그리고 무엇보다 '인간다움'을 지켜나갈 수 있는 내적 힘을 길러주어야 합니다. 이렇게 데이터와 직관을 함께 다루는 이중 훈련이야말로 미래 인재의 핵심 조건입니다.

저는 여전히 AI의 가능성을 믿습니다. 의료, 교육, 환경, 물류 등 모든 분야에서 AI가 가져올 혁신은 인류의 삶을 더욱 풍요롭게 만들 것입니다. 제가 강의에서 보여드리는 AI 활용법들이 실제로 많은 분들의 업무 효율성을 높이고 새로운 가능성을 열어주고 있음을 확인하고 있습니다.

하지만 동시에 경계심도 늦추지 않을 것입니다. AI의 어두운 면을 직시하고, 끊임없이 질문하고, 인간 중심의 가치를 지켜나가는 것이 우리의 책임입니다.

이 책이 던진 10가지 질문에 대한 완벽한 답을 저도 가지고 있지 않습니다. 중요한 것은 답이 아니라 질문 자체입니다. 과학의 역사가 보여주듯 직관은 진리로 가는 길을 열어왔고, AI는 그 길을 더 빠르고 넓게 확장해주는 도구가 되었습니다. 인간이 반드시 가져야 할 최고의 능력은 이 두 가지를 함께 활용하는 힘이며, 우리 사회도 그 조화를 길러내는 데 주안점을 두어야 합니다.

변호사로서, 전직 정치인으로서, 교육자로서 저는 앞으로도 이런 질문들을 던지고 답을 찾아가는 일에 헌신할 것입니다. 독자 여러분도 이 여정에 함께해 주시기를 바랍니다.

AI는 인간이 만든 기술입니다. 따라서 그 방향을 결정하는 것도 결국 인간의 몫입니다. 우리가 어떤 선택을 하느냐에 따라 AI는 인류의 가장 위대한 동반자가 될 수도, 가장 위험한 적이 될 수도 있습니다.

선택의 순간에 우리 모두가 현명하기를 바라며, 직관과 AI라는 두 거인이 조화롭게 춤추는 미래를 꿈꿉니다. 이 책이 그 지혜의 한 조각이 되기를 희망합니다.

2025년 11월 김경진

도서명 AI가 인간에게 던지는 10가지 질문

전자책 발행 | 2025년 11월 27일

저 자 | 김경진

펴낸이 | 김경진

펴낸곳 | 김경진 변호사 출판사

출판사등록 | 2025. 3. 10. (제2025-000015호)

주 소 | 서울특별시 동대문구 전농로 91, 백일빌딩 304호

전 화 | 02-6338-1905

이메일 | kimkj008@gmail.com

ISBN |

가격 10,000원

전체 목차

책의 모든 장 보기

김경진 변호사

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