AI로 인해 발생하는 사회구조 변화와 대응

목차, 서문, 13장, 에필로그

김경진

Korean PDF edition

AI로 인해 발생하는 사회구조 변화와 대응

김경진

김경진이 AI서재에 공개한 온라인 책 『AI로 인해 발생하는 사회구조 변화와 대응』입니다. AI 사회구조 변화, 인공지능 정책, 노동, 경제, 사회 대응을 주제로 목차, 서문, 13장, 에필로그 구성으로 읽을 수 있습니다.

목차

서문

전체 목차

책의 모든 장 보기

2025년 11월, "AI가 인간에게 던지는 10가지 질문"을 세상에 내놓았습니다. 독극물 분자 4만 개를 6시간 만에 설계하는 AI, 70개국에서 시민을 감시하는 알고리즘, 3억 개의 일자리를 집어삼킬 자동화의 파도. 10개의 질문을 던졌지만, 책을 탈고하는 순간 이미 알고 있었습니다. 질문이 끝나지 않았다는 것을.

반년이 지났습니다. 그 사이 세상은 제가 책에서 "아직 먼 이야기"라고 조심스럽게 썼던 장면들을 현실로 바꿔놓았습니다. AI 에이전트가 회의를 잡고, 계약서를 검토하고, 고객 상담의 90%를 처리하겠다고 나서는 기업들이 속출했습니다. 마이크로소프트는 6,000명을 해고했고, 듀오링고는 "AI가 할 수 있는 일에는 계약직을 고용하지 않겠다"고 못 박았습니다. 전작에서 경고했던 일들이 예측이 아니라 보도가 되어 돌아왔습니다.

전작은 질문을 던지는 책이었습니다. 이 책은 그 질문들이 현실이 된 뒤의 풍경을 기록합니다. 10가지 질문이 AI라는 존재 자체를 향한 것이었다면, 이번 13개 장은 그 AI가 건드리고 있는 인간 문명의 구조물들을 향합니다. 노동시장, 교육 체계, 금융 질서, 민주주의, 에너지 그리드, 도시의 공간 배치, 인간관계의 밀도. 인류가 수백 년에 걸쳐 쌓아올린 것들입니다. 완벽하지는 않았지만, 그래도 작동했던 질서들이었습니다. 그 질서가 지금 동시다발로 흔들리고 있습니다.

솔직히, 두렵습니다. 변호사로서, 전직 국회의원으로서, 전국을 다니며 AI 강의를 하는 교육자로서 저는 이 변화를 가장 가까이에서 보고 있는 사람 중 하나입니다. 강의장에서 만나는 50대 직장인의 눈에서 불안을 읽습니다. "저는 뭘 해야 하나요?"라는 질문 앞에서 말문이 막힐 때가 있습니다. 20년 경력의 개발자가 AI 앞에서 자신의 가치를 증명해야 하는 시대. 대학 졸업장의 무게가 GitHub 커밋 로그보다 가벼워지는 시대. 인간이 수천 년 동안 당연하게 여겼던 것들, 노동의 존엄, 교육의 가치, 전문성의 위계, 공동체의 유대, 이런 것들이 한꺼번에 재심에 부쳐지고 있습니다.

이 책을 쓰면서 가장 큰 도움을 받은 것은 해시드(Hashed) 김서준 대표의 글들입니다. 그가 바이브 코딩 대기 시간에 메모한 "다가오는 30개의 균열"은 이 책의 설계도와도 같았습니다. 중간관리직 소멸, 빌러블 아워 경제의 붕괴, 콘텐츠 제작 비용의 제로 수렴, 이력서 시대의 종언, 인간 노동의 럭셔리화. 그가 떠오르는 대로 적었다는 30개 항목에 네 개의 AI 모델이 매긴 "3년 내 실현 확률" 퍼센트는, 제가 각 장의 긴급성을 가늠하는 시간표가 되었습니다. 그의 신자본주의 제안, 사회적 가치(Social Value)를 측정하고 보상하는 새로운 경제 구조에 대한 구상은, 이 책이 비관으로 가라앉지 않도록 붙잡아준 닻이었습니다. 그가 제주 오프사이트에서 관찰한 "만드는 것이 가장 쉬워지고, 평가가 가장 앞단으로 올라온" 세계의 역전 현상, 대학의 세 기능(정보, 관계, 선발) 분리에 대한 분석, 인간 복제본의 등장과 정체성 확장에 대한 성찰은, 각 장의 뼈대를 세우는 데 결정적이었습니다.

김서준 대표 외에도 이 책에 지혜를 빌려준 분들이 있습니다. 김진석 인하대 교수의 "인간의 지위가 반려견으로 추락할 수 있다"는 경고, 스티븐 울프럼의 "상자 속에서 영원히 비디오 게임을 하는 수조 개의 영혼" 예측, 제프리 힌튼의 AI 의식에 대한 성찰, 유발 하라리의 제도적 대응 제안. IEA, WEF, Goldman Sachs, Gartner, McKinsey의 보고서들은 제 주장에 숫자라는 뼈를 붙여주었습니다. 이 자

리를 빌려 깊이 감사드립니다.

인류는 산업혁명도 겪었고, 두 차례 세계대전도 통과했고, 핵무기의 공포 속에서도 질서를 재건했습니다. 그런데 이번에는 감각이 다릅니다. 김서준 대표의 말처럼 "전쟁이나 지정학은 1년 지나면 다른 국면으로 바뀌지만, AI 충격은 10년 내내 우리를 못살게 굴 것"이기 때문입니다. 충격이 한 번 오고 지나가는 것이 아니라, 매일 조금씩 강도를 높이며 계속됩니다. 적응할 틈을 주지 않습니다. 예측할 수 없다는 것, 그것이 이 시대가 주는 가장 깊은 불안입니다.

이 책이 그 불안에 대한 답을 갖고 있다고 말하지 않겠습니다. 다만 균열의 위치를 정확히 기록하는 것이 첫 번째 할 일이라고 믿습니다. 어디가 무너지고 있는지 모르면, 어디를 고쳐야 하는지도 알 수 없으니까요.

2026년 5월 김경진

전체 목차

책의 모든 장 보기

김경진 변호사

1장 노동 현장의 재편

전체 목차

책의 모든 장 보기

1. 일자리 소멸 속도와 인간 적응 속도의 불일치

2025년 5월 어느 화요일 아침, 마이크로소프트 본사 레드먼드 캠퍼스에 6,000통의 해고 통지서가 발송되었습니다. 전체 직원의 3%에 해당하는 숫자였습니다. 해고된 직원 가운데 40% 이상이 소프트웨어 개발자였습니다. 18년 경력의 타입스크립트 개발자도, Python 언어의 핵심 개발자들도 명단에 올랐습니다. 마이크로소프트의 AI 스타트업 지원 조직을 이끌던 가브리엘라 드 케이로스마저 해고되었습니다. AI 분야에서 일한다고 해서 안전하지 않다는 사실을 이 사건은 냉혹하게 증명했습니다.

이상한 점이 있었습니다. 마이크로소프트는 그 분기에 역대급 실적을 기록하고 있었습니다. 1분기 매출 70.1억 달러, 전년 대비 13% 성장. 회사가 어려워서 자른 것이 아닙니다. 사티아 나델라 CEO는 담담하게 말했습니다. "우리 회사 소프트웨어 코드의 30%를 AI가 작성하고 있습니다." 해고는 더 이상 위기의 신호가 아니라 AI 시대의 경영 전략이 된 것입니다.

마이크로소프트만이 아니었습니다. 구글은 어시스턴트 사업부에서 수백 명을 잘랐고, 아마존은 2025년 10월 14,000명의 기업 직원을 정리했습니다. 메타는 3,600명, 테슬라는 14,000명. 이유는 한결같았습니다. "AI에 집중하기 위해서." 2025년 상반기에만 기술 분야에서 AI와 직접 연결된 해고가 77,999건에 달했습니다. 하루 평균 427명꼴입니다. 2026년 3월까지 누적된 기술 업계 해고 45,363건 가운데 약 9,238건, 그러니까 20.4%가 기업 스스로 AI와 자동화를 원인으로 지목했습니다. 2025년에는 이 비율이 8%에 못 미쳤으니, 1년 사이에 기업들의 언어가 바뀐 셈입니다. 더 이상 돌려 말하지 않습니다.

숫자의 규모를 넓혀 보겠습니다. 골드만삭스는 2035년까지 전 세계에서 약 3억 개의 정규직 일자리가 자동화의 영향권에 들어설 것이라고 추산합니다. 세계경제포럼(WEF)의 2025년 미래 일자리 보고서는 좀 더 세분화된 전망을 내놓았습니다. 2030년까지 9,200만 개의 일자리가 사라지고, 1억 7,000만 개의 새 일자리가 생겨 순증 7,800만 개라는 것입니다. 언뜻 희소식처럼 보이지만, 이 숫자에는 함정이 있습니다. 사라지는 일자리와 생기는 일자리가 같은 사람에게 돌아가지 않습니다. WEF에 따르면 AI 관련 신규 직업의 77%가 석사 학위를, 18%가 박사 학위를 요구합니다. 해고된 콜센터 직원이 AI 연구원으로 전환하기는 사실상 불가능합니다.

해시드 김서준 대표의 진단이 정곡을 찌릅니다. "AI 충격은 다른 충격과 다르다. 전쟁이나 지정학은 1년 지나면 다른 국면으로 바뀌지만, AI 충격은 10년 내내 우리를 못살게 굴 것이다." 기술이 인간의 적응력을 앞지르는 이 불일치를 국가 수준에서 해결하지 않으면, 사회 전체에 충격파가 밀려옵니다. 미국고용통계국(BLS) 자료에 따르면 컴퓨터 개발자 고용은 2023년부터 2025년까지 2년간 27.5% 감소했습니다. 1980년 이래 최저 수준입니다. 전체 고용은 늘었는데 개발자 채용 공고는 35% 줄었습니다. 일시적 현상이 아닙니다. 구조가 바뀌고 있는 겁니다.

Anthropic CEO 다리오 아모데이는 2025년 파리 VivaTech 행사에서 "AI가 향후 5년 내에 사무직 신입 직원 일자리의 절반을 없앨 수 있다"고 경고했습니다. 40대와 50대 직장인들의 상황은 더 곤혹스

럽습니다. 새 기술을 배우기에는 시간이 부족하고, 은퇴하기에는 아직 이릅니다. HBR이 2025년 12 월에 실시한 조사에서는 더 불편한 현실이 드러났습니다. 많은 기업이 AI의 실제 성과가 아니라, AI가 가져다줄 것이라는 기대만으로 해고를 단행하고 있었습니다. AI라는 단어가 해고의 면죄부가 된 것 입니다. Resume.org가 미국 채용 담당자 1,000명을 대상으로 한 설문에서 59%가 "AI를 해고의 이 유로 내세우는 편이 이해관계자에게 더 잘 먹힌다"고 시인했습니다.

문제의 본질은 속도 차이에 있습니다. AI는 20년 경력 개발자의 기술을 20분 만에 학습합니다. 채용 플랫폼 원티드랩이 현직 개발자 180명을 대상으로 조사한 결과, 절반가량이 생성 AI의 코딩 실력이 1~3년 차 개발자를 능가한다고 응답했습니다. 인간의 재교육에는 수년이 걸리고, 대학 커리큘럼 개편에는 그보다 더 오래 걸립니다. 기술은 기하급수적으로 발전하는데 인간의 적응은 선형적입니다. 이 간극이 좁혀지지 않는 한, 노동 시장의 혼란은 계속될 수밖에 없습니다.

2. 빌러블 아워(시간당 과금) 경제의 붕괴

뉴욕 맨해튼의 한 대형 법무법인에서 주니어 변호사 세 명이 이틀 동안 매달렸던 일이 있었습니다. 소송 건의 3년 치 회의록과 이메일을 검토하여 핵심 쟁점을 추출하는 작업이었습니다. 수천 페이지 분량이었고, 빌링 시간으로 환산하면 약 60시간, 고객에게 청구되는 비용은 대략 3만 달러 안팎이었습니다. 2026년 봄, 같은 법무법인의 같은 팀이 같은 종류의 작업을 AI 리서치 도구에 맡겼습니다. 결과가 나오기까지 걸린 시간은 5분이었습니다.

5분짜리 결과물에 60시간 치 비용을 청구할 수 있을까요. 답은 이미 나와 있습니다. 메타(Meta)는 외부 법률 자문 가이드라인에서 AI가 생성한 작업물에 대해 변호사 시간당 요율로 비용을 청구하지 말 것을 명시했습니다. 사이버 보안 기업 Zscaler도 마찬가지입니다. UBS는 2026년 초에 AI 관련 조항을 별도로 포함한 빌링 가이드라인을 발표했습니다. 메시지는 수렴하고 있습니다. 기계가 할 수 있는 일에 인간의 시간당 요율을 지불할 의사가 없다는 것.

빌러블 아워, 즉 시간당 과금 모델은 법률 산업의 운영체제와 같았습니다. 법무법인이 사건에 인력을 배치하는 방식, 주니어 변호사를 평가하는 기준, 파트너를 보상하는 체계, 매출을 성장시키는 동력 모두가 이 하나의 과금 방식에 의존했습니다. 그런데 이 운영체제가 붕괴하고 있습니다. 해시드 김서준 대표가 "다가오는 30개의 균열"에서 지적한 것처럼, 컨설팅, 법률, 회계 등 B2B 지식산업 전체가 요금표를 다시 써야 하는 상황에 놓였습니다. 네 개의 AI 모델이 독립적으로 평가한 "3년 내 실현 확률"의 평균은 80%였습니다.

구조적 모순이 있습니다. 같은 유형의 고용 분쟁을 두 변호사가 처리한다고 가정합니다. 한 명은 40시간 만에 해결하고, 다른 한 명은 120시간이 걸립니다. 시간당 과금 모델에서 고객은 느린 변호사에게 세 배를 지불합니다. 결과는 동일한데도요. 이 시스템은 효율성을 벌주고, 비효율성에 보상합니다. AI가 이 모순을 적나라하게 드러냈을 뿐입니다.

Thomson Reuters의 2025년 보고서에 따르면, 법률 전문가들은 AI 도입으로 연간 약 240시간을 절약할 수 있을 것으로 예측합니다. 문서 요약, 초벌 리서치, 증언 요약, 계약 조항 추출, 타임라인 작성 등 주니어 변호사의 빌링 시간 가운데 상당 부분이 이미 상용 AI 도구의 처리 범위 안에 있습니다. 고객들이 이 항목에 대한 비용 지급을 체계적으로 거부하기 시작하면, 살아남는 법무법인은 AI로 범용 작업 계층을 처리하고 그 위의 판단 계층에 비용을 청구하는 곳뿐입니다.

대안은 이미 존재합니다. 가치 기반 과금(Value-Based Pricing)은 컨설팅, 회계, 투자은행 등 다른 전문 서비스 산업에서 수십 년 전부터 표준이었습니다. 법률 산업만이 마지막 보루를 지키고 있었을 뿐입니다. 시카고 변호사 재단의 밥 글레이브스 사무총장은 "빌러블 아워는 지난 50년간 법률 서비스를 일반인에게 감당할 수 없을 만큼 비싸게 만든 최대 원인"이라고 진단합니다. AI는 그 50년 된 제도의 종언을 앞당기는 촉매입니다.

법률 산업만의 이야기가 아닙니다. 컨설팅도 같은 압력을 받고 있습니다. 맥킨지의 주니어 컨설턴트가 2주간 작성하던 시장 분석 보고서를, AI가 하루 만에 초안을 완성합니다. 회계 감사에서 전표 대조 작업을 AI가 처리하면 감사 시간이 절반으로 줄어듭니다. 시간을 파는 것이 곧 직업이었던 모든 산업이 같은 질문 앞에 섰습니다. "우리가 실제로 팔고 있는 것은 시간인가, 판단인가."

Artificial Lawyer는 2026년 전망 보고서에서 "법률 문서의 90%가 AI에 의해 생성될 것"이라고 예측했습니다. 그러면서도 "빌러블 아워의 죽음은 성급한 진단"이라는 유보를 달았습니다. 고급 업무, 전례에 기반한 리스크 평가, 시장 관행에 대한 전문적 판단에는 여전히 시간 기반 과금이 유지될 것이라는 겁니다. 맞는 말입니다. 모든 법률 업무가 동시에 바뀌지는 않습니다. 하지만 하급 업무와 범용 업무에서 빌러블 아워가 무너지면, 그 기반 위에 서 있던 전체 수익 구조가 흔들립니다. 피라미드의 하단이 축소되면 상단도 재조정될 수밖에 없습니다.

앞으로 증명해야 할 것은 시간이 아니라 판단입니다. 이 전환의 속도가 얼마나 빠를지는 산업마다 다르겠지만, 방향은 돌이킬 수 없습니다.

3. 상시 성과 평가와 상시 감시의 경계 소멸

한 국내 IT 기업의 팀장 K씨는 2025년 하반기부터 새로 도입된 AI 기반 성과 관리 시스템을 쓰고 있습니다. 이 시스템은 팀원들의 코드 커밋 빈도, 슬랙 메시지 응답 시간, 화상회의 발언 비율, 문서 작성량을 실시간으로 추적합니다. 분기별 성과 평가에 들어가던 시간이 90% 이상 줄었다고 회사는 자랑합니다. K씨의 팀원들은 다른 감상을 가지고 있습니다. "화장실 다녀오면 슬랙 응답 시간이 늘어난다"는 농담이, 농담이 아니게 된 것입니다.

김서준 대표의 "30개의 균열" 29번 항목이 이 상황을 예견했습니다. "AI가 업무 산출물과 협업 패턴을 상시 분석한다. 연말 평가 투입 시간 90% 이상 감소한다. 상시 평가가 공정성을 높이는 순간, 그것은 상시 감시와 구별이 불가능해진다." 네 AI 모델이 평가한 3년 내 실현 확률은 45%였습니다. 수치가 낮다고 안심할 수 없는 이유는, 이미 부분적으로 실현되고 있기 때문입니다.

감시 자본주의(Surveillance Capitalism)라는 개념을 제안한 쇼샤나 주보프 교수의 분석을 떠올리지 않을 수 없습니다. 플랫폼 기업은 사용자의 행동 데이터를 추출하여 예측 상품으로 바꿉니다. 같은 논리가 이제 일터로 침투합니다. "직원 몰입도 측정"이라는 이름 아래 감정 상태와 업무 패턴이 실시간으로 수집됩니다. 2025년 기준으로 미국 대기업의 60% 이상이 어떤 형태로든 직원 모니터링 소프트웨어를 사용하고 있다는 조사 결과가 있습니다. 코로나19 팬데믹 이후 원격근무가 확산되면서 이 수치는 가파르게 올랐습니다.

경계가 소멸하는 과정은 이렇습니다. AI 시스템은 개별 노동자의 업무 패턴을 학습하고, 생산성 하락의 조기 징후를 포착하고, "개선 권고"를 자동으로 보냅니다. 여기까지는 성과 관리입니다. 그런데 같은 시스템이 이메일의 어조를 분석하고, 화상회의에서의 표정 변화를 추적하고, 점심시간의 길이를 기록하기 시작하면 이것은 감시입니다. 기술적으로 둘 사이에 경계선을 긋는 것은 불가능합니다. 성

과를 측정하는 데이터와 행동을 감시하는 데이터가 같은 데이터이기 때문입니다.

평가는 사후적인 것이 아니라 실시간 알고리즘의 피드백 루프에 통합되었습니다. 노동은 데이터 추출과 행동 수정의 대상이 되었고, 노동자는 이 시스템 안에서 최적화의 변수로 전환됩니다. 연말 한 번의 인사 평가가 가져다주던 불안과, 365일 24시간 측정당하는 불안은 질적으로 다릅니다.

이 흐름에 맞서려면 "무엇을 측정하느냐"가 아니라 "무엇을 측정하지 않을 것이냐"에 대한 사회적 합의가 필요합니다. 기술이 가능하게 만든 것을 전부 실행하는 것이 기업의 당연한 권리인지, 노동자에게는 측정되지 않을 권리가 있는지, 이 질문에 대한 답이 아직 없습니다.

EU는 2025년 발효된 AI법에서 작업장 내 감정 인식 AI의 사용을 금지하는 조항을 포함시켰습니다. 하지만 성과 데이터 수집과 감정 모니터링 사이의 경계는 기술적으로 모호합니다. 슬랙 메시지의 응답 속도를 측정하는 것은 성과 관리입니다. 그 메시지의 어조에서 감정 상태를 추론하는 것은 감정 인식입니다. 같은 데이터에서 두 가지 정보가 동시에 추출됩니다. 하나를 금지하면서 다른 하나를 허용하는 것이 실무적으로 가능한지 아직 아무도 증명하지 못했습니다.

아마존의 물류 창고에서는 이미 수년 전부터 노동자의 동선과 작업 속도가 초 단위로 추적되어 왔습니다. 화장실에 다녀오는 시간이 생산성 점수에 반영된다는 보도가 있었고, 시스템이 자동으로 해고 권고를 생성한다는 사실이 알려져 논란이 되기도 했습니다. 이것이 블루칼라 노동의 예외적 사례라고 생각했던 사람들이 많았을 겁니다. 그런데 지금, 같은 논리가 화이트칼라 사무실로 들어오고 있습니다. 코드 커밋 수, 문서 작성량, 회의 참여도. 측정 대상만 달라졌을 뿐 구조는 같습니다.

답을 미루는 동안 감시 인프라는 이미 깔리고 있습니다.

4. 주니어 경력 사다리의 단절과 직업 재생산 구조의 파괴

숙련(mastery)으로 가는 통로가 끊어지고 있습니다. 과거에는 이런 순서가 있었습니다. 법무법인의 1년 차 변호사가 수천 장의 계약서를 눈으로 읽습니다. 회계법인의 신입 감사원이 수백 건의 전표를 손으로 대조합니다. 소프트웨어 회사의 인턴 개발자가 선배 코드를 읽으며 디자인 패턴을 체득합니다. 지루하고, 반복적이고, 때로는 무의미해 보이는 그 시간이 전문가를 만들었습니다. 도제 제도(apprenticeship)의 디지털 버전이었습니다.

AI가 이 통로를 차단하고 있습니다. 기업 입장에서 보면 합리적인 선택입니다. 주니어 변호사가 이틀 걸리는 문서 검토를 AI가 5분 만에 처리하는데, 왜 주니어를 고용하겠습니까. 듀오링고의 루이스 폰 안 CEO는 2025년 4월 "AI가 할 수 있는 일에는 더 이상 계약직을 고용하지 않겠다"고 선언했습니다. 사람인 조사에 따르면 2025년 1분기 국내 IT업계 신입 개발자 채용은 전년 대비 18.9% 줄었습니다. 골드만삭스 연구에 따르면 기술 관련 직종의 20~30세 실업률은 2025년 초부터 약 3%포인트 상승했습니다. 다른 직종이나 전체 기술 근로자와 비교해 현저히 높은 수치입니다.

미국 전체를 보면 입사 레벨 채용 공고가 전년 대비 15% 감소했고, 구인 광고에 "AI"를 언급하는 기업은 2년 사이 400% 급증했습니다. 2025년 1월, 전문 서비스업 구인 공고는 2013년 이래 최저치를 기록했습니다. 전년 대비 20% 감소한 수치입니다.

여기서 끊어지는 것은 일자리만이 아닙니다. 직업이 세대를 걸쳐 재생산되는 구조 자체가 무너집니다. 시니어 변호사는 주니어가 검토한 문서를 다시 점검하면서 판단의 기준을 전수했습니다. 시니어

개발자는 주니어의 코드를 리뷰하면서 설계 철학을 가르쳤습니다. 주니어가 사라지면 이 전수의 고리가 끊깁니다. 10년 후, 20년 후에 시니어가 될 사람이 실무 경험을 쌓을 기회 자체를 잃는 것입니다.

김서준 대표의 "30개의 균열" 15번 항목은 이 문제를 정확하게 포착합니다. "기업 법무팀이 절반으로 줄어든다. 계약서 검토, 규제 준수, 리스크 분석의 상당수를 에이전트가 수행한다. 주니어 변호사 수요가 감소한다. 주니어에서 시니어로 올라가는 사다리는 절벽이 된다." 3년 내 실현 확률 60%. 회계 분야도 다르지 않습니다. 같은 메모의 18번 항목에 따르면, "중소기업 10곳 중 7곳이 회계사 없이 세무를 끝낸다. 회계사의 존재 이유가 장부를 맞추는 것에서 숫자 너머의 판단을 내리는 것으로 압축된다." 3년 내 실현 확률 85%.

Gartner의 예측이 이 전환의 속도를 보여줍니다. 2026년까지 20%의 조직이 AI를 써서 조직 구조를 수평화하고, 현재 중간관리직의 절반 이상을 제거할 것이라는 겁니다. 조직이 수평화되면 주니어가 밟고 올라설 계단 자체가 사라집니다.

결국 이 문제는 효율성의 역설입니다. 단기적으로 기업은 비용을 절감하지만, 장기적으로는 전문가 계층의 세대 교체가 불가능해집니다. 경력 사다리의 맨 아래 칸을 치워버리면, 맨 위에 올라설 사람도 결국 사라집니다.

The Atlantic은 미국의 최근 대졸자들 사이의 실업률이 비정상적으로 높아졌다고 보고했습니다. 기업들이 신입 사무직 일자리를 AI로 대체하거나, AI 투자에 대한 지출이 신규 채용을 위한 예산을 밀어내고 있다는 분석입니다. 이것은 개인의 문제가 아닙니다. 한 세대 전체가 전문적 실무 경험의 기회를 박탈당하는 것이고, 그 결과는 10년 뒤, 20년 뒤에 산업 전체의 전문성 공백으로 나타날 것입니다. 의사 수련 과정에서 실습을 빼버리면 어떤 일이 벌어질지 상상해 보십시오. 교과서만 읽은 의사가 환자를 집도할 수 있을까요. 법률, 회계, 소프트웨어 개발, 컨설팅 등 모든 지식산업에서 같은 일이 벌어지고 있습니다.

인간 전문성의 재생산 시스템이 한 세대 안에 붕괴할 수 있다는 경고를 지금 우리는 실시간으로 목격하고 있습니다.

5. 중간관리직 소멸과 조직 구조의 수평화

2024년 10월, 가트너(Gartner)는 IT 심포지엄에서 한 가지 예측을 발표했습니다. "2026년까지 20%의 조직이 AI를 사용하여 조직 구조를 수평화하고, 현재 중간관리직의 절반 이상을 제거할 것이다." 2025년 가트너 CEO 서베이에서는 더 강한 신호가 나왔습니다. 조사 대상 CEO의 절반 이상이 향후 5년 내에 AI를 써서 중간관리 계층을 줄이겠다고 답한 것입니다.

중간관리자가 하는 일을 분해해 보겠습니다. 상부의 전략을 하부의 실행 단위로 번역합니다. 팀원들의 업무를 배분하고 진행 상황을 추적합니다. 성과를 측정하고 경영진에게 리포트합니다. 갈등을 조율하고, 정보를 위아래로 전달합니다. 이 가운데 업무 배분, 진척 추적, 성과 리포트는 AI 에이전트가 이미 상당 부분 처리할 수 있습니다.

김서준 대표의 "30개의 균열" 1번 항목이 이 변화를 압축합니다. "에이전트가 업무 배분, 진척 추적, 성과 리포트를 처리하면서 상위 매니저의 관리 범위가 5배 이상 확대된다. 최상위 관리직 외에 정보를 위아래로 옮기던 상당수 중간관리직은 불필요해진다." 3년 내 실현 확률 65%. 가트너의 분석과

방향이 일치합니다. 2025년 가트너 CEO 서베이에 따르면, 기업들은 AI를 통해 한 명의 관리자가 약 20% 더 많은 직원을 감독할 수 있게 만들려 합니다. 관리 범위가 넓어지면 중간 계층의 필요성은 줄어듭니다.

이 변화의 이면이 있습니다. 조직 내에서 대부분의 문제는 '중간'에서 발생합니다. 비전만 제시하는 경영진과 실무만 담당하는 일선 사이의 간극, 부서 간 이해관계의 충돌, 예상치 못한 상황에서의 판단, 이런 것들을 중간관리자가 처리해왔습니다. AI가 데이터 기반의 업무 조율은 잘하지만, 팀원의 사기가 떨어진 이유를 파악하거나, 두 부서 사이의 암묵적 갈등을 풀어내는 것은 다른 차원의 능력입니다.

그럼에도 기업들의 움직임은 빠릅니다. 단기적으로 인건비가 줄고, 의사결정 속도가 빨라지고, 보고 계층이 단순해지니까요. Workday는 2025년에 전체 인력의 8.5%, 약 1,750명을 해고하면서 AI 투자로 자원을 재배치한다고 발표했습니다. 감축 대상의 상당수가 중간관리 기능을 수행하던 직원들이었습니다.

수평화된 조직은 빠릅니다. 하지만 취약하기도 합니다. 중간 계층이 담당하던 완충 기능, 멘토링 기능, 갈등 조정 기능이 사라지면 조직은 경영진의 판단 오류에 직접 노출됩니다. 중간관리자가 없는 조직에서 주니어 직원은 누구에게 배울 것이며, 번아웃을 겪을 때 누구에게 말할 것인지, 이 질문에 AI는 아직 답하지 못합니다.

가트너 자체도 이 부작용을 인식하고 있습니다. 같은 보고서에서 "전통적 멘토링 경로가 훼손될 수 있고, 주니어 직원의 발전 기회가 줄어들 수 있으며, 남은 관리자가 급격히 늘어난 직속 부하에 압도당할 수 있다"고 경고합니다. 콜센터의 소멸도 같은 맥락입니다. 김서준 대표의 메모 5번 항목에 따르면 AI 처리율이 3년 내 90%에 도달하면서 한국 기준 40만 명이 재배치 대상이 됩니다. 콜센터 관리자, 팀장, 품질 관리(QA) 담당자 모두가 이 수평화의 대상입니다.

역사적으로 보면, 조직이 중간 계층 없이 운영된 적은 거의 없습니다. 로마 군단에도 백인대장(centurion)이 있었고, 중세 길드에도 장인과 도제 사이에 직인(journeyman) 계층이 있었습니다. 중간 계층의 존재는 인간 조직의 보편적 특성에 가깝습니다. 그 계층을 기술로 제거하는 실험이 어떤 결과를 낳을지, 우리에게는 아직 충분한 선례가 없습니다.

조직 구조의 수평화는 불가피합니다. 문제는 속도입니다. 너무 빠르게 중간 계층을 걷어내면 조직의 면역 체계가 같이 사라집니다. 가트너의 예측대로 2026년까지 5분의 1의 조직이 이 방향으로 움직인다면, 그중 상당수는 수평화의 부작용을 뒤늦게 발견하게 될 것입니다.

6. 인간 노동의 럭셔리화와 'Handmade by Human' 프리미엄

서울 청담동의 한 레스토랑에서는 코스 요리의 마지막 접시가 나올 때 셰프가 직접 테이블로 나옵니다. 재료의 산지를 설명하고, 조리 과정에서의 고민을 이야기합니다. 음식의 맛이 전부가 아닙니다. 인간이 직접 만들었다는 사실, 그 사람의 이야기를 직접 들었다는 경험이 30만 원짜리 디너의 가격을 구성합니다. AI가 레시피를 분석하고 로봇 팔이 정확한 온도로 조리하는 레스토랑이 등장하더라도, 이 경험을 대체하기는 어렵습니다.

김서준 대표의 "30개의 균열" 10번 항목이 이 현상을 예측합니다. "Handmade by Human 라벨에 큰 프리미엄이 붙는다. 수제 가구, 인간 요리사, 대면 상담이 고급 서비스가 된다. 비효율은 인간 노동의 마지막 경쟁력이다." 3년 내 실현 확률 75%. 같은 메모의 5번 항목도 같은 방향을 가리킵니다. 콜

센터가 사라지면서 "인간과 통화하는 것 자체가 프리미엄으로 과금된다"는 전망입니다.

이 변화의 논리는 이렇습니다. AI가 결과물의 복제를 쉽게 만들수록, 과정에 대한 가치가 올라갑니다. AI는 픽셀과 데이터를 조합하여 그럴듯한 결과물을 만들어내지만, 그 안에 담긴 인간적 맥락, 의사결정의 문법, 불완전함에서 오는 개성은 담아내지 못합니다. 미슐랭 식당에서 셰프의 스토리가 음식의 가치를 결정하듯, 인간이 직접 했다는 사실 자체가 럭셔리한 가치가 되는 영역이 생겨나고 있습니다.

Klarna의 사례가 이 역학을 보여줍니다. 2024년 2월, 이 핀테크 회사는 AI 챗봇이 700명의 고객서비스 직원과 같은 일을 해낸다고 발표했습니다. 첫 달에만 230만 건의 상담을 처리하고, 평균 해결 시간을 11분에서 2분으로 줄였습니다. 4,000만 달러의 비용 절감 효과를 예상했습니다. 그런데 1년 후인 2025년 5월, 세바스찬 시에미아트코프스키 CEO는 방향을 바꿨습니다. "AI만으로는 품질이 떨어진다"며 인간 고객서비스 직원을 다시 채용하기 시작한 것입니다. 효율과 비용의 관점에서 AI는 압도적이었지만, 고객이 원하는 것은 효율만이 아니었습니다.

럭셔리화에는 그림자가 있습니다. 인간 노동이 프리미엄이 된다는 것은, 인간 노동을 살 여유가 있는 사람과 없는 사람 사이의 격차가 벌어진다는 뜻이기도 합니다. 부유층은 인간 의사에게 진료를 받고, 인간 변호사에게 상담을 받고, 인간 교사에게 자녀를 맡깁니다. 나머지는 AI로 대체된 서비스를 씁니다. 공감, 돌봄, 창의적 변주가 요구되는 노동은 고급화되고, 효율 중심의 노동은 AI에게 넘어갑니다.

이미 이 분리는 금융 서비스에서 시작되고 있습니다. 김서준 대표의 메모 21번 항목이 이를 포착합니다. "에이전트가 시장 데이터와 소비 패턴을 통합 분석해 자산을 자동 리밸런싱하는 서비스가 보편화된다. 부자는 컨시어지 목적으로 사람 상담사가 커버하지만, 대중은 에이전트로 충분하다. 금융의 계층화가 더 노골적으로 드러난다." 의료도 마찬가지입니다. 같은 메모의 7번 항목에 따르면 1차 의료의 70%는 패턴화된 진단과 처방인데, 이것이 앱으로 넘어가면 인간 의사의 대면 진료는 지불 능력이 있는 사람만 누리는 서비스가 됩니다.

인간이 직접 하는 모든 것이 럭셔리가 되는 세상은, 다르게 말하면 인간적 접촉을 돈으로 사는 세상입니다. 수공예 가구 장인이 AI 자동화 공장과 공존하는 것은 아름다운 그림이지만, 수공예 가구를 살 수 있는 사람이 상위 5%뿐이라면 그것은 시장의 다양성이 아니라 계급의 고착화입니다.

그래서 'Handmade by Human'이라는 라벨은 희망이면서 동시에 경고입니다. 인간 노동의 고유한 가치가 인정받는다는 의미이지만, 그 가치가 소수만 누릴 수 있는 사치품이 될 위험도 담고 있습니다. 비효율이 인간 노동의 마지막 경쟁력이라면, 그 경쟁력을 누구나 접근할 수 있도록 만드는 것이 사회의 몫입니다. 그렇지 않으면 우리는 AI가 서비스하는 대중과 인간이 서비스하는 엘리트라는 두 개의 세계를 만들게 됩니다. 그 분기점이 이미 시작되었습니다.

전체 목차

책의 모든 장 보기

김경진 변호사

2장 교육 현장의 변화와 적응

전체 목차

책의 모든 장 보기

1. 암기/재현 중심 교육의 종말과 질문 중심 교육으로의 전환

2026년 봄, 서울 강남의 한 초등학교 교실에서 열한 살짜리 소녀가 태블릿 화면을 응시하고 있습니다. 화면에는 AI 튜터가 띄워놓은 질문이 하나 적혀 있습니다. "로마에 전기가 있었다면 노예 경제는 어떻게 달라졌을까?" 소녀는 30초쯤 생각하더니 기계에게 되물었습니다. "노예가 줄어들면 콜로세움 관중석은 누가 지었어?" 기계는 0.8초 만에 건축 노동력의 대안적 경로 세 가지를 펼쳐놓았고, 소녀는 그중 하나를 골라 다시 꼬리 질문을 이어갔습니다. 이 장면에서 외워야 할 것은 아무것도 없었습니다. 로마 멸망의 연도도, 콜로세움의 높이도 기계 안에 이미 들어 있으니까요. 소녀가 한 일은 기계가 스스로는 하지 못하는 것, 즉 엉뚱한 방향에서 질문을 던지는 일이었습니다.

하버드 대학교 물리학과의 2025년 실험은 이 장면에 숫자를 붙여줍니다. AI 튜터를 쓴 학생들은 전통적인 능동 학습 교실의 학생들보다 두 배 이상 많은 내용을 익혔고, 걸린 시간은 오히려 짧았습니다. 같은 해 글로벌 조사에서 학생의 AI 사용률은 전년 66%에서 92%로 뛰었습니다. 2026년 초 기준, 고등교육 학생 86%가 AI를 주된 리서치 및 브레인스토밍 파트너로 쓰고 있다는 추산도 나왔습니다. 교실의 중심축은 이미 기울었습니다. 정보를 전달하는 곳에서, 질문을 설계하는 곳으로.

구글 딥마인드의 데미스 허사비스는 인공지능이 인간의 모든 인지 능력을 갖추게 될 시점이 5년에서 10년 사이에 올 것이라고 봅니다. 단백질 구조를 예측하는 알파폴드가 이미 질병 치료와 에너지 문제의 실마리를 풀기 시작했다는 것이 그의 근거입니다. 이런 환경에서 수업 시간에 학생이 해야 할 일은 정답을 외우는 게 아닙니다. 기계가 최선의 답을 내놓도록 유도하는 질문을 짜는 것, 그리고 기계가 내놓은 답 중에서 어떤 것이 쓸 만하고 어떤 것이 엉터리인지 가려내는 것입니다. 소크라테스가 2,500년 전에 보여준 원리가 기술의 옷을 입고 돌아온 셈이거든요. 답이 아니라 질문이 사고의 본질이라는 원리 말입니다.

문제는 평가 체계가 이 전환을 따라잡지 못하고 있다는 데 있습니다. 표준화된 시험지에 정답을 적어내는 방식으로는 인공지능과 협력하여 복잡한 문제를 해결하는 역량을 측정할 수 없습니다. UNESCO 조사에 따르면 전 세계 450개 이상의 학교와 대학 중 AI 사용 가이드라인을 마련한 곳은 고작 10%에 불과합니다. 시험지는 20세기에 머물러 있고, 학생들은 21세기 도구를 손에 쥐고 있으며, 교사들은 그 사이 어딘가에서 방향을 찾지 못하고 있습니다. 지식을 소유하는 시대에서 지식을 운용하는 시대로의 이행은, 정해진 길을 따라가는 순응성보다 새로운 길을 모색하는 탐구심을 요구합니다. 2026년 초, 미국에서만 AI 관련 학사 프로그램이 193개, 석사 프로그램이 310개로 불어났고, 카네기멜런 대학은 이미 2018년에 AI 학사 학위를 처음 개설한 바 있습니다. UCSD는 새 AI 전공에 신입생 150명을 등록시키고, 2029년까지 학부생 1,000명 규모로 키울 계획을 세웠습니다. 대학들은 변화를 감지하고는 있습니다. 하지만 교육이 이 요구에 응답하지 못하면, 학교는 박물관이 됩니다.

2. 대학 학위 프리미엄의 반감과 이력서 시대의 종언

2026년 4월, 갤럽과 월턴 가족재단, GSV 벤처스가 공동으로 발표한 "AI 역설(The AI Paradox)" 보고서는 Z세대의 심리 상태를 숫자로 포착했습니다. AI에 대해 흥분을 느낀다고 답한 비율이 전년 36%에서 22%로 추락했고, 분노를 느낀다는 응답은 22%에서 31%로 치솟았습니다. 희망적이라는 응답은 27%에서 18%로 떨어졌습니다. 가장 격앙된 연령대는 대학을 막 졸업했거나 졸업을 앞둔 20대 초반이었습니다. 갤럽의 수석 교육 연구원 잭 흐리노프스키는 이렇게 해석했습니다. "4년간 등록금을 내고 졸업했는데, AI가 자기 산업을 뒤집어엎고 있다는 걸 깨달은 세대입니다."

이 분노에는 근거가 있습니다. AI 관련 기술이 임금에 23%의 프리미엄을 붙여주는 반면, 학사 학위가 붙여주는 프리미엄은 8%에 그친다는 조사 결과가 2025년 말에 나왔습니다. 고용주의 81%가 학위보다 기술을 우선해야 한다고 믿으면서도 52%는 여전히 학위 소지자를 뽑습니다. 덜 위험하다고 느끼기 때문입니다. 이 모순적 수치가 대학 학위의 현재 위치를 정확히 보여줍니다. 신호로서의 가치는 남아 있지만, 그 신호의 가격 대비 성능은 빠르게 떨어지고 있는 것입니다.

해시드의 김서준 대표는 대학이 오래 독점해온 세 기능이 하나씩 분리되고 있다고 진단합니다. 정보, 관계, 선발. 예전에는 이 셋이 하나의 패키지 안에 묶여 있었습니다. 정보는 MIT 오픈코스웨어, 유튜브, GitHub이 대체하기 시작했고, 관계는 오픈소스 커뮤니티와 해커톤이 새로운 악수를 만들어내고 있으며, 선발은 GitHub 스타 수와 실제 사용자 수가 학점 4.5보다 높은 해상도의 신호가 되는 경우가 생겼습니다. 김 대표가 교류하고 있는 개발 특성화 고등학교 출신 인재들의 성장 속도를 보면, 대학이 여전히 최선의 기본값인지 진지하게 되묻게 된다는 것이 그의 솔직한 고백입니다.

마이클 스펜스의 신호 이론은 여기서 다시 읽을 가치가 있습니다. 교육이 생산성을 직접 높이기 때문이 아니라, 이미 생산성이 높은 사람이 그것을 증명하기 위해 학위를 쓴다는 생각. 이 모델이 맞다면 학위는 학습의 증거가 아니라 학습 능력의 신호입니다. 더 빠르고 더 저렴하고 더 조작하기 어려운 신호가 생기면, 학위의 가격은 깎입니다. IBM, 애플, 구글이 일부 직군에서 학위 요건을 없앤 것은 교육의 가치를 부정해서가 아닙니다. 학위로 측정하던 것을 더 직접적으로 측정할 수 있게 되었기 때문입니다. 피터 틸이 대학을 버블이라 불렀던 것도 같은 문제의식이었습니다. 가격이 가치를 앞지른 상태가 영원히 유지될 수 있는가. 미국 학자금 대출 총액이 1조 달러를 넘어선 현실 앞에서, 이 질문은 수사가 아니라 회계입니다.

골드만삭스 보고서가 전망한 향후 10년간 3억 개 일자리의 자동화 편입은 이 문제를 더 날카롭게 만듭니다. 신입 사원이 수행하던 반복적이고 정형화된 업무가 기계 지능의 영역으로 빠르게 넘어가면서, 선배가 후배에게 업무를 가르치며 숙련도를 높여가던 도제식 경로가 끊기고 있습니다. 이력서에 적힌 화려한 학력보다, 인공지능 도구를 써서 즉각적으로 성과를 내는 능력이 채용의 핵심 잣대가 되고 있습니다. 이력서는 과거의 성취를 나열하는 서류에서, 미래의 복잡한 문제를 기계와 함께 해결할 수 있음을 증명하는 포트폴리오로 대체되어야 합니다. 대학의 더 본질적인 문제는 시간의 구조에 있습니다. 가장 빨리 자랄 수 있는 시기의 인재를, 수십 년 전 패러다임으로 짜인 커리큘럼과 중간고사, 기말고사의 리듬 안에 묶어두는 일. 김서준 대표의 표현을 빌리면, "졸업장이라는 정류장에 고속 버스를 세워두는 순간, 다시 본선에 합류하는 데 4년이 걸릴 수도 있습니다."

3. 학원가의 존재 이유 변화: 성적 향상에서 불안 관리로

밤 10시, 대치동 학원가의 불빛은 여전합니다. 건물 사이를 오가는 학부모들의 표정에는 묘한 초조함이 서려 있는데, 그 초조함의 성격이 달라졌습니다. 5년 전이라면 "수학 1등급을 받을 수 있을까"였을 걱정이, 지금은 "이 아이가 졸업할 때쯤 이 직업이 남아 있을까"로 바뀌었습니다. 학원비 영수증

의 명목은 여전히 국어, 영어, 수학이지만, 실질적으로 구매하는 상품은 불안의 완화입니다. 자녀의 미래가 불투명해질수록, 학원 등록은 합리적 투자가 아니라 심리적 보험에 가까워집니다.

김서준 대표의 "30개의 균열" 메모에서 이 현상은 냉정한 문장으로 요약되어 있습니다. "에이전트 튜터가 학생별 취약점을 실시간 분석한다. 다만 학원이 진짜 파는 건 자녀 돌봄과 부모 불안 관리라서, 성적보다 안심을 파는 산업은 에이전트가 건드리기 어렵다." 4개 AI 모델이 독립적으로 평가한 3년 내 실현 확률은 45%였습니다. 기술적으로는 학원의 핵심 기능을 대체할 수 있지만, 학원이 팔고 있는 진짜 상품, 즉 심리적 보험은 기계가 제공할 수 없다는 뜻입니다.

2026년 갤럽 조사는 이 불안의 온도를 측정해줍니다. Z세대 K-12 학생의 74%가 "AI가 학습을 더 어렵게 만들 가능성이 있다"고 답했고, 이미 직장에 나간 Z세대 성인은 83%가 같은 우려를 표했습니다. AI가 학습 속도를 높여줄 수 있다고 믿는 비율은 전년 53%에서 46%로 내려갔습니다. 학부모들이 느끼는 불안은 이 숫자들의 그림자입니다. 자녀 세대가 느끼는 불확실성을 지켜보면서, 유일하게 손에 잡히는 행동이 학원 등록이기 때문입니다.

1830년대 뉴욕에서 벤저민 데이가 발행한 선(The Sun) 신문은 대중의 불안과 호기심을 자극해 광고 수익을 올리는 주의력 경제의 원형을 만들었습니다. 학원가는 이 구조를 교육의 언어로 번역해놓은 것에 가깝습니다. 기계 지능이 모든 것을 대신해주는 시대에 인간만이 할 수 있는 무언가가 있다는 희망을 상품화하고, 그 희망이 흔들릴 때마다 새로운 커리큘럼을 출시합니다. 코딩 교육이 유행하면 코딩반이 열리고, 프롬프트 엔지니어링이 화제가 되면 프롬프트반이 생깁니다. 내용은 바뀌어도 구조는 같습니다. 부모의 불안을 감지하고, 그 불안에 이름을 붙이고, 이름표가 붙은 불안을 수강료로 전환하는 것.

교육의 목적이 한 인간의 성장이 아니라, 플랫폼 경쟁 사회에서 살아남기 위한 부품이 되는 것으로 변질될 때, 학원은 성적 향상의 산실이 아닌 거대한 불안 관리 센터로 기능하게 됩니다. 이 구조가 위험한 이유는 분명합니다. 불안은 해소되지 않기 때문입니다. 기계 지능이 발전할수록 불안의 총량은 늘어나고, 학원비의 총액도 함께 늘어납니다. 아이의 역량이 얼마나 올랐느냐와 무관하게, 부모가 느끼는 불안이 줄었느냐가 학원 재등록의 기준이 됩니다. 스키너의 상자 속 비둘기가 간헐적 보상에 반응하여 버튼을 누르듯, 학부모들은 불안이라는 자극에 반응하여 학원가로 몰려드는 셈입니다. 학원가의 매출은 교육의 성과가 아니라 불안의 총량에 비례합니다. 그리고 인공지능이 발전할수록 불안의 원천은 마르지 않습니다.

4. AI 출력물의 오류를 걸러내는 능력의 중요성

2026년 4월 ICLR 학회에서 발표된 논문 "추론의 함정(The Reasoning Trap)"은 불편한 사실 하나를 드러냈습니다. 모델의 추론 능력을 강화하면 할수록, 도구 환각(tool hallucination) 비율이 동반 상승한다는 것입니다. 더 똑똑하게 만들수록 더 그럴듯하게 틀립니다. 이 사실은 기업의 96%가 이미 AI 에이전트를 실무에 투입한 상황에서 경보에 가깝습니다.

숫자는 구체적입니다. 2026년 기준 37개 모델을 대상으로 한 벤치마크에서 환각률은 15%에서 52% 사이에 분포했습니다. 의료 사례 요약에서는 완화 프롬프트 없이 64.1%의 환각률이 관측되었고, 완화 기법을 적용해도 43%까지밖에 내려가지 않았습니다. 법률 분야는 더 심각합니다. 스탠포드 규제연구소의 연구에 따르면, 대형 언어모델은 구체적인 법률 질문에 대해 69%에서 88%의 확률로 환각을 일으킵니다. 4건 중 3건에서 존재하지 않는 판례를 지어내거나 조문을 왜곡한다는 이야

기입니다.

런던의 대형 로펌에서 근무하는 사라 로든의 일과는 이 숫자의 현실판입니다. 그녀는 3년 치 왓츠앱 대화 기록과 방대한 회의록을 단 몇 분 만에 분석해내는 AI의 속도에 감탄하면서도, 기계가 내놓는 답변을 맹신하지 않습니다. 법률 서비스에서 기계의 작은 오류는 의뢰인의 비밀 누설이나 법적 특권의 훼손으로 직결될 수 있기 때문입니다. 그녀의 업무 시간 중 상당 부분은 기계가 만들어낸 문서가 법적 효력을 갖추었는지, 개인정보 보호 규정을 준수했는지 검증하는 데 쓰입니다. 생산(production)의 시대에서 감수(editing)와 검증(verification)의 시대로 무게중심이 옮겨간 것입니다.

김서준 대표의 프롬프팅 벤치마크 에세이는 이 능력을 시험의 형태로 상상합니다. 가장 까다로운 파트는 일부러 결함이 심어진 AI 출력물을 주고, 그 오류를 잡아내라고 요구하는 것입니다. "그럴듯하지만 미묘하게 틀린 분석, 논리적이지만 전제가 잘못된 추론. AI가 자신감 있게 내놓는 오답을 그대로 삼키는 것이야말로 이 시대의 가장 위험한 함정이고, 이를 걸러내는 능력이야말로 프롬프팅 역량의 진짜 핵심이다." 2026년 CHI 학회에서는 10세에서 14세 아동 48명을 대상으로 환각 인식 교육 실험이 발표되었습니다. 아이들에게 AI 챗봇을 직접 만들게 하면서 환각 탐지 스캐폴드를 제공했더니, 사전-사후 테스트에서 AI 지식, 환각 인식, 신뢰 가능한 챗봇 구축 자신감이 유의미하게 상승했습니다. 아이들은 비일관성을 탐색하고 외부 소스와 교차 확인하는 다층적 전략을 스스로 개발해냈습니다.

복잡한 수식을 푸는 것보다, 기계가 내놓은 답의 타당성을 입증하는 능력. 이것이 인공지능과 공존하는 시대의 새로운 지적 권위가 될 것입니다. 기계의 출력물에 책임을 지는 주체는 오직 인간뿐이기 때문입니다. 기계는 훌륭한 조수일 뿐, 결정을 내리는 주체가 될 수 없습니다. 그 경계를 지키는 능력이 전문가의 새로운 자격이 됩니다. 환각률 17%의 최고 성능 모델조차 6건 중 1건은 틀린다는 현실 앞에서, 검증 없는 수용은 직업적 자살입니다.

5. 프롬프팅 역량 측정이 만들어낼 새로운 교육 불평등

같은 ChatGPT, 같은 모델, 같은 가격을 쓰면서도 누군가는 한 줄짜리 요약을 받고, 누군가는 논문 수준의 분석을 뽑아냅니다. 이 차이를 만드는 것이 프롬프팅 역량입니다. 김서준 대표는 이것이 21세기의 가장 중요한 생산성 지표가 되리라고 확신합니다. 그리고 여기에 불편한 그림자가 있습니다.

프롬프팅 역량은 비판적 사고, 문제 구조화 능력, 메타인지와 깊이 연결되어 있습니다. 이것들은 교육 환경과 문화 자본에 크게 좌우됩니다. 좋은 질문을 하려면 좋은 질문을 접해본 적이 있어야 합니다. "마케팅 전략 짜줘"라고 치는 사람은 블로그 포스트 수준의 일반론을 받습니다. 타겟 고객의 심리 프로파일부터 경쟁사 포지셔닝 맵, 채널별 ROI 가설까지 구조화해서 던지는 사람은 실행 가능한 전략 문서를 받습니다. 이 격차는 모델이 강해질수록 벌어집니다. 도구의 천장이 올라갈수록 그것을 다루는 손의 차이가 기하급수적으로 증폭되기 때문입니다.

유엔 총회에 보고된 바에 따르면 전 세계 118개국이 인공지능 경쟁에서 완전히 뒤처져 있으며, 인구의 3분의 1은 여전히 기본적인 인터넷조차 쓰지 못하고 있습니다. 부유한 나라의 학생은 고도화된 모델과 무제한의 컴퓨팅 자원을 쓰며 프롬프팅 기술을 연마하지만, 개발도상국의 학생은 전기 공급과 인터넷 속도를 걱정해야 합니다. AI 기술 역량이 23%의 임금 프리미엄을 만들어내는 세계에서, 프롬프팅 역량은 계층 이동의 사다리가 아니라 또 다른 차별의 벽이 될 수 있습니다.

한국의 상황도 다르지 않습니다. 2026년 갤럽 조사에서 Z세대 직장인의 48%는 AI를 업무에 쓰는 것의 위험이 이점보다 크다고 답했습니다. 전년의 37%에서 11%포인트 급등한 수치입니다. AI 보조 작업을 신뢰한다는 응답은 10명 중 3명도 안 되었고, AI만으로 수행된 작업을 신뢰한다는 응답은 사실상 없었습니다. 흥미로운 역설은, 기술에 대한 불신이 커질수록 그 기술을 잘 다루는 소수의 프리미엄은 올라간다는 점입니다. 대다수가 기피하는 도구를 능숙하게 쓰는 사람에게 시장은 더 높은 가격을 매깁니다.

김서준 대표는 드래곤볼의 스카우터를 비유로 듭니다. "전투력 5, 쓰레기 같군." 조잡하지만 강력한 숫자, 복잡한 전투 능력의 모든 차원을 하나의 수치로 압축해서 즉각적인 판단을 가능하게 했던 장치. 수능 점수나 학벌이 아니라, 이 사람이 AI와 협업했을 때의 출력 품질을 예측할 수 있는 숫자가 곧 필요해진다는 것입니다. 그 시험은 당연히 오픈북이고, AI를 쓰는 것이 전제입니다. AI 없이 능력을 측정하는 것은 계산기 없이 수학 실력을 재는 것만큼 시대착오적이니까요.

그런데 진짜 질문은 점수가 아닐지도 모릅니다. 벤치마크는 결국 사회가 "무엇을 중요하게 여기는가"의 거울입니다. 수능이 있었기에 한국은 암기력과 풀이 속도를 중시했고, 사교육 시장은 그 측정 기준에 맞춰 거대한 산업이 되었습니다. 프롬프팅 벤치마크가 등장하면, 교육과 채용과 승진의 기준이 그 위에서 재편됩니다. 측정이 현실을 만듭니다. 그렇다면 이 새로운 측정 기준이 만들어낼 세상은, 과연 지금보다 더 공정한 세상일까요. 격차가 선형이 아니라 기하급수적으로 벌어지는 구조에서, 이 물음에 낙관하기는 어렵습니다.

전체 목차

책의 모든 장 보기

김경진 변호사

3장 경제적 불평등의 심화

전체 목차

책의 모든 장 보기

1. 실행의 민주화가 낳는 안목의 귀족화

런던 시티의 한 상업 소송 로펌에서 변호사 사라 로든은 의뢰인으로부터 과일 상자 하나를 건네받았습니다. 안에는 3년 치 회의록, 왓츠앱 대화 기록, 내부 이메일이 빽빽하게 들어 있었습니다. 예전이라면 주니어 변호사 네다섯 명이 이틀 밤을 새우며 이 서류를 분류하고, 누가 언제 회계 문제를 처음 인지했는지 타임라인을 재구성했을 것입니다. 사라는 대형 언어모델에 전체 문서를 밀어 넣었습니다. 5분 뒤, 기계는 특정 이사회 멤버가 2021년 9월 14일 오후 3시 왓츠앱 메시지에서 재무 이상 징후를 처음 언급했다는 사실을 정확히 짚어냈습니다.

이 장면이 보여주는 것은 기술의 효율성이 아닙니다. 실행의 비용이 0에 수렴하는 세계가 이미 도착했다는 사실입니다. 해시드의 김서준 대표는 이 현상을 한 문장으로 압축했습니다. "실행의 민주화가 안목의 귀족화를 낳는다." 네 개의 인공지능 모델(Claude Opus 4.6, Grok 4.20, Gemini 3.1 Pro, GPT-5.4)이 독립적으로 평가한 이 예측의 3년 내 실현 확률은 85%였습니다. 네 모델 모두 거의 같은 숫자를 내놓았다는 것 자체가 이 변화의 확실성을 보여줍니다.

누구나 글을 쓰고, 이미지를 생성하고, 코드를 뽑아낼 수 있게 되면서 만드는 행위 자체는 가치를 잃어가고 있습니다. 가치의 무게중심이 이동합니다. 어떻게 만드느냐(How)에서 무엇이 좋은 것인가(What is good)를 판단하는 영역으로. 인공지능이 내놓은 수천 개의 시안 가운데 어떤 것이 의뢰인의 의도에 부합하고 시대를 관통하는 미학을 담고 있는지 결정하는 힘, 그것은 여전히 기계 바깥에 있습니다. 김서준 대표의 표현을 빌리면, "실행 능력이 평준화되면서 문제 정의력과 미적 감각이 소득 격차를 결정한다"는 것입니다.

문제는 그 안목이 어디서 오느냐입니다. 과거에는 주니어 단계에서 수천 건의 서류를 직접 읽고, 수백 시간의 실무를 거치며 맥락을 체득했습니다. 그 과정이 지름길 없는 도제식 교육이었고, 10년 뒤 시니어가 되었을 때 비로소 "이건 안 되고, 저건 된다"는 직관이 몸에 배는 구조였습니다. 인공지능이 주니어의 업무를 흡수하면서 이 사다리가 끊어지고 있습니다. 기초 실행 단계를 건너뛴 세대가 어떻게 고도의 안목을 가진 전문가로 성장할 수 있을까요. 로펌에서 벌어진 일은 디자인 스튜디오에서도, 컨설팅 펌에서도, 방송국 편집실에서도 동시에 진행 중입니다. 김서준 대표의 30개 균열 목록에서 기업 법무팀이 절반으로 줄어든다는 항목은 3년 내 실현 확률 60%를 기록했는데, 핵심 포인트는 "주니어에서 시니어로 올라가는 사다리가 절벽이 된다"는 문장이었습니다.

빌러블 아워(시간당 과금) 경제의 붕괴 항목은 3년 내 실현 확률 80%를 기록했습니다. 컨설팅, 법률, 회계 등 B2B 지식산업 전체가 요금표를 다시 써야 합니다. 시간을 파는 것이 곧 직업이었던 사람들이 이제 증명해야 할 것은 시간이 아니라 판단입니다. 사라 로든도 인정합니다. 서류 검토에 5분 걸렸다고 해서 청구서에 5분을 쓸 수는 없다고. 20년간 쌓아온 법적 판단력의 가격을 어떤 단위로 매겨야 하는지, 업계는 아직 답을 찾지 못했습니다. 그녀의 파트너 중 한 명은 최근 회의에서 이렇게 말했다고 합니다. "우리가 파는 건 시간이 아니라 잠을 설친 횟수였는데, 이제 기계가 잠을 안 자니까."

결국 미래의 불평등은 도구의 소유 여부에서 발생하지 않습니다. 도구가 쏟아내는 무수한 선택지 가운데 진짜 가치를 골라내는 심미안과 맥락 이해력, 거기서 갈라집니다. 기술이 모든 이에게 실행의 날개를 달아주었는데, 그 날개로 어디를 향해 날지 결정하는 지적 안목은 오히려 더 높은 성벽 안으로 숨어들고 있습니다. 김서준 대표의 30개 균열 목록에서 인간 노동의 럭셔리화 항목은 3년 내 실현 확률 75%를 기록했습니다. "수제 가구, 인간 요리사, 대면 상담이 고급 서비스가 된다. 비효율은 인간 노동의 마지막 경쟁력이다." 이 진단이 불편하게 와닿는 이유는, 그것이 안목을 갖지 못한 대다수에게 남은 자리가 기계보다 느리다는 이유 하나만으로 프리미엄을 받는 시장뿐이라는 뜻이기 때문입니다.

2. 프롬프팅 역량 격차의 기하급수적 확대

같은 ChatGPT 앞에 두 사람이 앉아 있습니다. 한 사람은 "오늘 뭐 먹지?"를 묻고, 다른 사람은 같은 도구로 산업의 지형을 재설계하고 있습니다. 이 격차를 숫자로 표현할 수 있게 되는 날, 세상은 꽤 불편한 진실과 마주할 것입니다.

프롬프팅은 명령어를 입력하는 행위가 아닙니다. 문제의 본질을 파악하고, 그것을 기계가 이해할 수 있는 논리 구조로 재설계하며, 기계가 내놓은 결과의 오류를 잡아내고, 반복적으로 정제하는 고도의 지적 활동입니다. 구글 딥마인드의 데미스 허사비스는 미래의 학생들이 인공지능 도구를 모국어처럼 다루며 생산성을 열 배 이상 끌어올릴 것이라 내다봅니다. 그 혜택이 프롬프팅 역량을 갖춘 이들에게만 돌아간다는 점이 문제의 핵심입니다.

격차의 구조를 들여다봐야 합니다. 프롬프팅 역량은 비판적 사고, 문제 구조화 능력, 메타인지와 밀접하게 연결되어 있습니다. 이 능력들은 교육 환경과 문화 자본에 크게 좌우됩니다. 좋은 질문을 하려면 좋은 질문을 접해본 적이 있어야 합니다. 여기서 기존의 교육 불평등이 증폭됩니다. 격차가 선형이 아니라 기하급수적으로 벌어지기 때문입니다. 체스에서 컴퓨터가 인간을 이긴 뒤에도 인간 선수 간 실력 차이가 사라지지 않았듯, 인공지능이 아무리 강해져도 그것을 다루는 인간 사이의 격차는 사라지지 않습니다. 격차의 축이 바뀔 뿐이죠. 지식에서 질문으로, 정보에서 구조화로.

프롬프팅 시험이라는 개념을 상상해봅니다. 가장 까다로운 파트는 아마 이런 문제일 것입니다. 일부러 결함이 심어진 인공지능 출력물이 주어집니다. 그럴듯하지만 미묘하게 틀린 분석, 논리적이지만 전제가 잘못된 추론. 인공지능이 자신감 있게 내놓는 오답을 그대로 삼키는 것이야말로 이 시대의 함정이고, 이를 걸러내는 능력이야말로 프롬프팅 역량의 진짜 핵심입니다. 그리고 마지막 문제는 수험생의 전문 분야 바깥에서 출제됩니다. 변호사에게 유전공학, 의사에게 건축 설계. 측정되는 것은 사전 지식이 아니라, 모르는 영역에서 인공지능과 협업해 전문가급 결과를 만들어내는 능력입니다.

저소득 국가와 소외 계층은 고가의 인공지능 구독 서비스를 이용할 여력이 없을 뿐 아니라, 인공지능을 의미 있게 쓰는 방법을 배울 기회조차 얻지 못하고 있습니다. 2026년 세계 불평등 보고서(World Inequality Report 2026)에 따르면 북미와 오세아니아의 평균 자산은 세계 평균의 338%인 반면, 사하라 이남 아프리카는 20%에 불과합니다. 같은 도구 위에서도 이 격차는 그대로 복제됩니다. 인공지능이 인간의 데이터를 학습해 지능을 키워가는 동안, 정작 데이터를 제공한 대다수 대중은 기술의 수혜자가 아닌 소비자로 전락할 위험이 큽니다.

가장 역설적인 부분은 이것입니다. 인공지능 모델이 강해지면 "대충 물어봐도 괜찮은 답"의 영역이 넓어지고, 역량 차이가 드러나는 지점은 점점 더 복잡하고 미묘한 곳으로 이동합니다. 프롬프팅 역량

의 차이는 곧 정보의 비대칭으로 이어집니다. 한쪽은 인공지능을 지능적 비서로 부리며 부를 쌓고, 다른 쪽은 인공지능이 설계한 알고리즘의 통제 아래 놓입니다. 2025년 3분기 기준 미국 상위 1%의 순자산은 약 55조 달러로, 하위 90% 전체가 보유한 자산과 맞먹습니다. 연방준비제도 데이터입니다.

벤치마크는 결국 사회가 무엇을 중요하게 여기는가의 거울입니다. 수능이 있었기에 한국은 암기력과 풀이 속도를 중시했고, 사교육 시장은 그 측정 기준에 맞춰 거대한 산업이 되었습니다. 프롬프팅 벤치마크가 등장하면, 교육과 채용과 승진의 기준이 그 위에서 재편됩니다. 측정이 현실을 만듭니다. 우리가 어떤 스카우터를 만드느냐가, 어떤 전투력을 키울지를 결정합니다. 그런데 진짜 질문은 점수가 아닐지도 모릅니다. 이 새로운 측정 기준이 만들어낼 세상은, 과연 지금보다 더 공정한 세상일까요. 기술 주권이 소수의 국가와 기업에 집중되면서 지역적 가치와 맥락이 배제된 인공지능 모델이 전 세계로 퍼져나가는 현상도 이 격차를 부채질합니다. 같은 도구를 쥐었다는 사실이 평등의 증거가 되던 시대는 끝났습니다. 도구를 쥔 손의 방향이 모든 것을 가릅니다.

3. 금융 서비스의 노골적 계층화

1980년대 조나단 로빈이 개발한 프리즘(PRIZM) 시스템은 미국 인구를 라이프스타일에 따라 수십 개 집단으로 분류하며 정밀 마케팅의 시대를 열었습니다. 사람들이 어디에 살고 무엇을 소비하는지 분석해 시장 가치에 따라 서열을 매긴 것이죠. 40년 뒤, 인공지능은 이 모델을 완전히 다른 차원으로 밀어붙이고 있습니다.

국제금융협회(IIF)의 2025년 조사에 따르면 금융 서비스 기관의 85%가 이미 어떤 형태로든 인공지능을 씁니다. 이 기계들은 실시간 데이터와 예측 알고리즘으로 고객의 신용도를 소수점 단위까지 평가합니다. 전통적인 FICO 점수는 결제 이력, 부채 규모, 신용 기간, 신규 신용, 신용 구성이라는 다섯 가지 요소로 비교적 명료하게 작동했습니다. 쓰는 쪽도, 평가받는 쪽도 어떤 요인이 중요한지 알았습니다. 그런데 기계학습 기반 신용 모델은 완전히 다른 규모에서 작동합니다. 소셜 미디어 활동, 위치 정보, 온라인 쇼핑 패턴은 기본이고, 스마트폰 사용 습관, 온라인 양식을 채우는 속도, 신용 신청 시각까지 수백에서 수천 개의 데이터 포인트가 하나의 점수로 수렴합니다. 대출을 갚을 능력과 아무 관련이 없어 보이는 변수들이 운명을 결정합니다.

이 시스템은 과거의 차별을 학습합니다. 2022년 웰스파고는 알고리즘이 흑인과 라틴계 신청자에게 비슷한 재무 조건의 백인 신청자보다 높은 위험 점수를 부여한 사실이 드러나 조사를 받았습니다. 기계학습 모델은 역사적 데이터에서 패턴을 추출하는데, 그 데이터 자체에 인종, 성별, 연령에 따른 편향이 누적되어 있으니 편향의 재생산은 구조적 귀결입니다. 2025년 11월 발표된 한 학술 리뷰는 금융 알고리즘의 다수에서 결함을 확인하며, 문제의 핵심이 "가장 널리 쓰이는 알고리즘들이 여전히 특정 결정에 도달한 방법을 설명하지 못한다"는 점이라고 지적했습니다. 사람들은 어둠 속에 놓인 채 기술을 신뢰하도록 강요받고, 그 기술이 잠재적으로 삶을 망가뜨리는 와중에도요. EU는 2025년 발효된 인공지능법(AI Act)에서 신용 평가와 대출을 "고위험" 영역으로 분류하고 투명성과 인간 감독을 의무화했지만, 알고리즘 내부의 작동 방식을 소비자가 실질적으로 들여다볼 수 있는 수준은 아닙니다.

계층화는 양쪽 끝에서 동시에 가속됩니다. 부유한 고객에게는 인공지능이 선별한 최적의 투자 기회와 고도화된 자산 관리 비서가 제공됩니다. 반대편에서는 취약 계층의 심리적 약점을 파고드는 고금리 대출 상품이 정교하게 타기팅됩니다. 페이스북은 이미 사용자가 스트레스를 받거나 패배감을 느

끼는 순간을 포착해 맞춤 광고를 노출하는 기술을 보유하고 있습니다. 금융 현장에서도 이 정서 모니터링이 결합하면, 돈이 필요한 순간에 가장 불리한 조건의 상품이 가장 먼저 눈에 띄는 환경이 만들어집니다. 김서준 대표의 30개 균열 목록에서 금융 계층화는 3년 내 실현 확률 50%를 기록했는데, 수치가 낮아 보여도 그 이유가 흥미롭습니다. 기술적으로는 이미 가능하지만 규제의 속도가 기술을 따라잡지 못해서 현실화에 시간이 걸린다는 판단이었습니다.

Axios가 2026년 1월 보도한 미국의 부의 분배 데이터는 이 구조를 숫자로 보여줍니다. 인공지능 붐이 본격화된 2025년 2분기, 상위 10% 가구의 자산은 한 분기 만에 5조 달러 늘어난 반면 하위 50%의 자산 증가는 1,500억 달러에 그쳤습니다. 33배 차이입니다. CBS 뉴스에 따르면 미국의 부의 불평등은 30년 만에 최대 격차를 기록했고, 상위 1%의 순자산 비중은 32%로 사상 최대를 찍었습니다. 경제학자 리처드 윌킨슨이 경고했듯, 불평등이 심한 사회일수록 신뢰가 낮고 사회적 지위 불안이 높습니다. 인공지능은 이 불안을 데이터화하여 수익으로 전환하는 기계인 셈입니다.

금융 서비스의 계층화는 보이지 않는 차별이 아닙니다. 알고리즘이라는 장막 뒤에서 효율성이라는 이름으로 부의 이동 통로를 차단합니다. 데이터를 많이 가진 자가 더 유리한 조건을 선점하고, 데이터가 부족하거나 부정적 데이터가 쌓인 이들은 시스템 밖으로 밀려납니다. 디지털 레드라이닝(digital redlining)이라는 용어가 학계에서 등장한 이유입니다. 1930년대 미국 정부가 빨간 선으로 특정 인종 거주 지역의 대출을 차단한 것과 문법이 같습니다. 다만 빨간 잉크 대신 수백만 줄의 학습 데이터가 그 선을 긋고 있을 뿐입니다.

4. 1인 기업 폭발과 고용 없는 성장의 가속

매슈 갤러거(Matthew Gallagher)라는 이름을 기억해야 합니다. 2024년 9월 그는 ChatGPT, Claude, Midjourney, Runway, ElevenLabs를 조합해 원격 의료 플랫폼 메드비(Medvi)를 혼자서 만들었습니다. 첫 달 고객 300명. 두 달 뒤 1,000명 추가. 2025년 첫 해 매출 4억 100만 달러, 고객 25만 명, 순이익률 16.2%. 2026년 예상 매출은 18억 달러입니다. 비교 대상인 힘스앤허스(Hims & Hers)는 2,442명의 직원으로 24억 달러 매출에 순이익률 5.5%를 기록했습니다. 갤러거는 형 엘리엇 한 명을 고용했을 뿐입니다. 직원 1인당 매출을 계산하면 비교 자체가 무의미해집니다.

샘 올트만은 2023년 동료 기술 CEO들과 "최초의 1인 10억 달러 기업이 등장하는 연도"에 대한 내기를 걸었다고 밝혔습니다. Anthropic의 다리오 아모데이는 2025년 5월 "Code with Claude" 컨퍼런스에서 그 시점을 2026년으로 지목하며 70~80% 확률을 부여했습니다. 가장 가능성이 높은 분야로 자동 매매(proprietary trading)와 개발자 도구를 꼽았습니다. 이스라엘 개발자 마오르 슐로모(Maor Shlomo)는 노코드 플랫폼 Base44를 혼자 만들어 6개월 만에 25만 사용자를 확보하고 Wix에 8,000만 달러에 매각했습니다. 2025년 12월의 일입니다. 수익성까지 달성한 상태에서의 매각이었습니다.

숫자가 추세를 말해줍니다. 미국 센서스 데이터 기준 비고용주 사업체는 2,980만 개로, 미국 경제에 1.7조 달러를 기여합니다. 솔로 창업자 비중은 2019년 23.7%에서 2025년 중반 36.3%로 뛰었습니다. 2026년 기준 7자리 매출(연 100만 달러 이상)을 올리는 사업체의 38%가 솔로프리너입니다. 네덜란드의 피터 레벨스(Pieter Levels)는 직원 0명으로 연 300만 달러 이상의 매출을 올리며 이 모델이 특수한 사례가 아님을 증명합니다. 1인 기업의 기술 스택 비용은 연간 3,000달러에서 12,000달러 사이. 같은 기능을 전통적 방식으로 충당하면 월 8만~12만 달러가 듭니다. 95~98%의 비용 절감, 60~80%의 영업이익률. 전통적 스타트업의 10~20%와는 차원이 다른 수치입니다.

이 화려한 1인 기업의 이면에 어두운 그림자가 있습니다. 기업들은 인간의 인지적 노동을 기계로 대체할 강력한 유인을 갖습니다. 골드만삭스는 전 세계 3억 개의 일자리가 자동화의 영향권에 들어올 것으로 예측했고, 자체 데이터에 따르면 2025년 초부터 기술직 20~30세의 실업률이 다른 연령대보다 거의 3%포인트 높게 나타났습니다. 주니어급 화이트칼라 노동자들이 수행하던 기초 업무가 인공지능으로 옮겨가면서 사회의 중간 허리 역할을 하던 자리들이 빠르게 비고 있습니다. 2025년 1월 미국 전문 서비스업 구인 공고는 2013년 이래 최저치를 기록했습니다. 전년 대비 20% 감소한 수치입니다.

김서준 대표는 이 딜레마를 명확히 짚었습니다. "AI 진전을 막으면 경쟁력을 잃고, AI 때문에 사라지는 일자리를 방치하면 사회에 큰 충격파가 온다. 두 사이드의 밸런스를 잡아야 한다." 그의 30개 균열 메모에서 1인 SaaS 회사 폭발은 3년 내 실현 확률 70%, 중간관리직 소멸은 65%를 기록했습니다. "팀을 만드는 능력보다 팀 없이 버티는 능력이 창업의 핵심 역량이 된다"는 그의 문장은 이미 현실이 됐습니다. Gartner도 2026년까지 20%의 조직이 인공지능으로 조직을 수평화하며 중간관리직의 절반 이상을 제거할 가능성을 내다봤습니다.

생산성이 높아져도 노동자에게 돌아가는 몫이 줄어드는 현상은 자본주의의 근간을 위협합니다. Bank of America 데이터에 따르면 2025년 12월 고소득 가구의 임금 상승률은 3%였으나, 중소득은 1.5%, 저소득은 1.1%에 그쳤습니다. 옥스팜(Oxfam)의 2026년 1월 보고서는 전 세계 억만장자의 자산이 2025년 한 해 동안 2.5조 달러 증가했다고 밝혔습니다. 인류 하위 절반인 41억 명이 보유한 전체 자산과 맞먹는 규모입니다. 자산 소유자와 고도의 인공지능 운용 능력을 갖춘 소수만이 성장의 열매를 독점하는 구조가 굳어지고 있습니다. 노동을 통해 기술을 익히고 숙련가로 성장하던 경로가 차단된 상태에서, 개인들은 시스템의 부품으로 전락할 위기에 처해 있습니다.

5. 자본주의 목표 함수의 위기: 돈에서 사회적 가치로의 전환 필요성

밀턴 프리드먼은 기업의 유일한 사회적 책임은 이익을 늘리는 것이라고 역설했습니다. 반세기 넘게 비즈니스 세계의 공리로 작동한 이 문장이 지금 심각한 자기모순에 빠지고 있습니다. 인공지능이 모든 것을 더 싸고 빠르게 만들어내더라도, 그것을 구매할 소득을 가진 인간이 사라진다면 시스템 자체가 멈추기 때문입니다. 제프리 힌튼의 경고를 떠올려봅니다. "AI가 인간의 지적 노동을 대체한다면, 과연 어떠한 종류의 새로운 일자리가 창출될 것인지 의문입니다. AI는 과거의 다른 기술혁명과는 매우 다른 종류입니다."

해시드의 김서준 대표는 이 문제에 대해 구체적인 설계도를 제시합니다. 그는 기존 자본주의를 "돈에서 더 많은 돈으로의 순환에 모든 것을 건 구조"로 정의하면서, 실제로 경제 활동이 만들어내는 산출물은 돈만이 아니라고 짚습니다. 일자리, 환경 영향, 세금, 그리고 사회적 가치. 문제는 우리가 이 토탈 지표를 만들지 못해왔다는 것입니다. 기존 모델은 이랬습니다. 기업이 돈 벌고, 세금 내고, 정부는 세금으로 사회 문제를 해결한다. 그런데 사회 문제 해결의 난이도가 올라가고 비용도 커지는데, 효율성을 따질 방법이 없으니 예산은 늘어나도 성과는 불투명합니다.

그의 제안은 이 순서를 뒤집는 것입니다. 사회적 가치(Social Value)를 만드는 행위 자체를 시장화하고, 측정하고, 정량화하고, 보상하는 체계를 구축하는 것. "돈을 많이 벌었으면 세금 내라. 사회 가치를 많이 만들었으면 그만큼 가져가라." 11년째 이 측정 방법론을 연구해온 김서준 대표는 평가 방법도 만들었고 연구원도 운영 중입니다. NGO 활동, 기업의 사회공헌, 장애인 고용, 환경 보호 같은 "착한 일"에 시장 원리를 적용하고, 그 일을 한 만큼 세금을 환류해주는 구조가 작동하면, 인공지능이 없

애는 일자리만큼 반대편에서 새로운 일자리가 생겨날 수 있다는 논리입니다.

지금 한국의 현실은 다릅니다. 착한 일에 칭찬은 하지만 그 일을 직업으로 삼는 데는 반대합니다. 리턴이 안 나오니까요. 부모들도 자녀가 사회적 가치를 창출하는 일에 뛰어드는 것을 꺼립니다. 측정을 못 하면 자원 배분도 안 되고, 제도화도 안 됩니다. 칭찬으로 끝나는 거죠. 세상에 측정을 못 하면 관리가 안 되고, 관리가 안 되면 만들 수도 없습니다. 자본주의 이론을 사회 사이드에 집어넣으면 새로운 경제를 만들 수 있다는 것이 김서준 대표의 핵심 주장입니다.

PwC의 2025년 연구는 흥미로운 가능성을 보여줍니다. 인공지능이 광범위하게 도입되고, 생산성 향상이 임금 상승으로 연결되며, 정책적 개입이 병행되는 최적 시나리오에서 미국의 지니계수가 2035년까지 0.8%포인트 개선될 수 있다는 분석입니다. 0.8%가 작아 보일 수 있지만, 1980년부터 현재까지 불평등이 가파르게 상승한 기간 동안 지니계수 변화가 약 0.8%였다는 점을 고려하면, 40년간의 불평등 확대를 10년 만에 되돌릴 수 있다는 뜻이 됩니다. 핵심은 "특정 조건 하에서"라는 전제입니다. 아무것도 하지 않으면 인공지능은 불평등 가속기가 됩니다.

데미스 허사비스는 인공지능이 질병을 치료하고 무한한 에너지를 제공하는 "급진적 풍요"의 시대를 열 것이라 낙관합니다. 물과 에너지 비용이 0에 수렴하는 세상. 그런데 풍요가 실현되더라도 분배의 문제는 그대로 남습니다. 2024년 한 해 동안 전 세계에서 204명의 새로운 억만장자가 탄생했습니다. 거의 매주 네 명꼴입니다. 옥스팜 보고서에 따르면 억만장자들의 집단 자산은 2025년 한 해 2.5조 달러 늘었는데, 이는 인류 하위 절반 41억 명의 총자산과 거의 같은 규모입니다.

김서준 대표의 논리를 끝까지 따라가면, 이 딜레마를 풀 수 있는 구조가 보입니다. 정부가 혼자 하던 사회 문제 해결을 국민 스스로가 수행하고, 정부는 심판 역할을 맡는 것. 이 시스템이 작동하면 Social Value Economy가 부상합니다. 스케일이 작으면 인공지능 때문에 혼란만 겪고, 인공지능 진전을 막으면 경쟁력을 잃습니다. "AI 충격은 다른 충격과 다릅니다. 전쟁이나 지정학은 1년 지나면 다른 국면으로 바뀌지만, AI 충격은 10년 내내 우리를 못살게 굴 것입니다." 자본주의의 목표 함수를 이윤에서 사회적 가치로 전환하는 것은 도덕적 선택이 아니라 시스템의 생존 조건입니다. 인공지능이 가져올 잠재적 풍요를 소수가 독점하는 디스토피아 대신, 그 풍요의 측정과 분배를 위한 새로운 운영체제를 설계해야 할 때입니다. 시간이 많지 않습니다.

김경진 변호사

4장 AI 기술 접근 가능성의 불평등

전체 목차

책의 모든 장 보기

1. 개인 에이전트 구독료라는 새로운 고정 지출과 디지털 낙오

2026년 4월, 서울 관악구의 한 고시원. 스물여섯 살 취업준비생 K는 매달 가계부를 쓸 때마다 같은 항목 앞에서 멈춥니다. ChatGPT Plus 월 20달러, Claude Pro 월 20달러, 코파일럿 월 30달러. 세 개 합치면 한 달에 10만 원이 넘습니다. 월세 35만 원에 식비 40만 원을 쓰는 그에게 AI 구독료는 네 번째로 큰 고정 지출입니다. 끊을 수 없습니다. 무료 버전으로는 이력서 첨삭도, 코딩 테스트 준비도, 면접 시뮬레이션도 제대로 돌아가지 않으니까요.

1830년대 뉴욕에서 벤자민 데이가 신문 한 부를 1페니에 팔겠다고 결심했을 때, 그는 독자의 주의력을 광고주에게 되파는 비즈니스 모델을 발명했습니다. 페니 프레스 혁명은 정보의 문턱을 낮추었습니다. 그로부터 190년 뒤, 인공지능은 정반대의 길을 걷고 있습니다. 과거에 사용자는 광고를 보는 대가로 무료 서비스를 누리는 '제품'이었습니다. 이제 사용자는 매달 일정한 구독료를 내야 고성능 지능에 접근할 수 있는 영구적 지출원이 되었습니다.

해시드의 김서준 대표는 "30개의 균열" 메모에서 이 변화를 80%의 실현 확률로 예측했습니다. "도시 근로자 대다수가 에이전트 2개에서 5개를 일상 운용하게 된다. 구독료가 새로운 고정 지출이 된다. 스마트폰이 그랬듯, 안 쓰는 것이 선택이 아니라 낙오가 된다." 이 문장의 무게는 숫자로 확인됩니다. 저소득 국가에서는 사용자의 97%가 무료 버전에 머물고, 유료 구독을 감당하는 비율은 3%에 불과합니다. 무료 버전은 성능 제한과 응답 지연을 동반하고, 실시간 의사결정이 필요한 경쟁 환경에서 이 격차는 치명적입니다.

스탠퍼드 사회혁신저널(SSIR)의 2026년 1월 분석이 이 구조를 정확히 짚었습니다. 현재의 저가 혹은 무료 인공지능 시대는 지속 가능하지 않다는 겁니다. 2025년에서 2026년 초까지는 무료 도구가 광범위하게 남아 있지만, 고급 기능은 이미 유료 티어로 몰리고 있습니다. 에너지, 하드웨어, 프라이버시, 시장 지배력이라는 인공지능의 기반 경제학이 스스로를 주장하기 시작한 겁니다. 문제는 이 비용 전가가 가장 취약한 계층에게 먼저 도착한다는 사실입니다.

밀턴 프리드먼이 주장한 주주 가치 우선의 논리는 이 불평등을 정당화하는 데 쓰입니다. 기업은 더 효율적인 지능을 팔아 수익을 올리는 데 집중하지, 사회적 배제를 해결하는 데는 관심이 없습니다. 구독료라는 여과 장치를 거치면서 경제적 빈곤은 인지적 빈곤으로 이어집니다. 고성능 에이전트 없이는 건강 관리도, 행정 서비스 접근도, 교육 기회 확보도 어려워지는 세계. 유발 하라리가 경고한 "경제적으로 무용한 계급(economically irrelevant class)"이 구독료 장벽 뒤에서 현실이 되고 있습니다.

마이크로소프트의 필립 로터가 언급한 과거의 '박스형 소프트웨어(shrink-wrapped software)' 시대에는 한 번 구매하면 소유권이 발생했습니다. 클라우드 기반 구독 모델로 전환되면서 지능은 소유의 대상에서 접속의 대상으로 변모했습니다. 접속권은 매달 갱신해야 합니다. 갱신을 멈추는 순간, 어제까지 쓰던 도구가 사라집니다. 이건 책을 빌려 읽는 것과 다릅니다. 지적 생산성 자체가 월 과금에 묶이는 구조입니다. 이 구조 속에서 구독을 중단한 사람은 생산성이 떨어지는 게 아니라, 경쟁 자체에

서 이탈합니다.

2026년 기준, 전 세계 인공지능 지출 총액은 3,010억 달러를 돌파했습니다. 이 돈의 38%는 미국에, 26%는 중국에 집중됩니다. 나머지 국가들은 부스러기를 나눠 갖습니다. 지능에 대한 과금은 교육적 성취의 기회를 원천 차단하여 세대 간 빈곤을 고착시킵니다. 초기 인터넷이 약속했던 지식의 민주화는, 지식의 유료화라는 결과로 뒤집히고 있습니다. 인공지능이 약속한 풍요는 구독 결제가 승인된 사람에게만 도착하는 셈입니다. K는 오늘도 가계부 앞에서 고민합니다. 식비를 줄일까, 구독을 끊을까. 이 선택지가 선택이 아니라는 게 문제입니다.

2. 세대 간 디지털 리터러시 격차

2025년 런던정경대(LSE) 조사 결과가 하나의 풍경을 그려줍니다. Z세대 근로자의 83%가 업무에 인공지능을 쓰고 있었습니다. 밀레니얼은 73%, X세대는 60%, 베이비부머는 52%. 숫자 자체보다 그 사이의 질감이 중요합니다. Z세대의 절반은 상사에게 물어보기 전에 ChatGPT에 먼저 질문한다고 답했습니다. 상사보다 기계를 먼저 찾는 세대와, 기계보다 경험을 먼저 믿는 세대가 같은 사무실에 앉아 있는 겁니다.

이 격차는 사용 빈도의 문제가 아닙니다. 리터러시의 문제입니다. 마이크로소프트 Viva 연구에 따르면 조직의 70%가 직원들에게 필요한 인공지능 역량을 가르치는 데 어려움을 겪고 있고, 리더의 62%가 명확한 AI 리터러시 격차를 인식하고 있습니다. 문제는 이 격차가 세대별로 비대칭적이라는 점입니다. 링크트인 직업 시장 보고서에 따르면 AI 리터러시를 요구하는 채용 공고가 지난 1년간 70% 늘었습니다. AI 역량은 이제 컴퓨터나 인터넷 사용 능력처럼 기본 자격이 되어가고 있습니다.

법조계의 현실이 이를 잘 보여줍니다. 과거 주니어 변호사는 서류를 검토하고 연대기를 작성하는 고된 기초 작업을 수년간 수행하며 숙련도를 쌓았습니다. 지금은 인공지능이 그 작업을 몇 분 안에 끝냅니다. 젊은 세대는 프롬프트를 넣고 결과를 검증하는 새로운 리듬에 빠르게 적응합니다. 반면 도제식 교육으로 전문성을 축적해온 기성세대는 자신의 노하우가 학습 데이터로 치환되는 과정에서 정체성의 위기를 겪습니다. 기성세대가 30년간 쌓은 직관은 여전히 값지지만, 그것을 AI 시스템과 결합해 산출물로 만들어내는 문법이 없으면 시장은 그 가치를 제대로 쳐주지 않습니다.

PwC의 2025년 글로벌 AI 일자리 바로미터는 충격적인 숫자를 내놨습니다. AI 역량을 보유한 근로자의 임금 프리미엄이 2024년 25%에서 2025년 56%로, 1년 만에 두 배 이상 뛰었습니다. 이 프리미엄은 AI를 잘 쓰는 사람과 못 쓰는 사람 사이의 소득 격차가 단순히 벌어지는 게 아니라 기하급수적으로 확대되고 있음을 뜻합니다. 연봉으로 환산하면 1만 8천 달러에서 3만 달러의 차이입니다.

EY와 AARP가 2026년 4월에 발표한 공동 조사는 또 다른 측면을 드러냅니다. 60세에서 85세 사이 2,515명을 16개국에서 조사한 결과, 기업과 학계가 베이비부머 세대는 AI 도입에 관심이 없다고 가정하는 것 자체가 문제였습니다. 실제로는 관심은 있되 접근 경로가 막혀 있는 경우가 대부분이었습니다. 교육 자원의 분배 불균형이 관심을 무력화시키고 있었던 겁니다. 인공지능 관련 교육을 받은 근로자는 전체의 13%에 불과하다는 랜드스타드 조사 결과가 이를 뒷받침합니다.

UNDP는 2025년 보고서 "다음 번 대분기(The Next Great Divergence)"에서 세대 간 균열을 직접 언급했습니다. 22세에서 25세 사이 근로자의 고노출 직종 고용률이 최근 몇 년간 약 5% 하락했습니다. 인공지능이 입문 단계의 기회를 줄이는 동안, 나이 든 근로자는 오히려 생산성 향상의 혜택을 누리는 역설적 구조가 나타난 겁니다. 젊은 세대는 도구를 잘 다루지만 경험을 쌓을 자리가 사라지고,

기성세대는 경험은 풍부하지만 도구를 다루는 문법이 낯섭니다. 두 세대 모두 각자의 방식으로 소외됩니다.

세대 간 지식 전수 모델이 해체되고 있습니다. 과거에는 시니어가 주니어에게 실무를 가르치고, 주니어가 시니어에게 새 기술을 알려주는 순환이 작동했습니다. 인공지능이 주니어의 일을 빼앗으면서 이 순환이 끊겼습니다. 인류가 수천 년간 쌓아온 전문 지식의 체계가 파편화될 위험이 여기에 있습니다. WEF에 따르면 고용주의 77%가 2025년에서 2030년 사이에 근로자를 AI에 맞게 재숙련시킬 계획이라고 합니다. 계획은 있습니다. 실행이 문제입니다. 랜드스타드 조사에서 근로자의 55%가 자기 경력을 지키기 위해 더 많은 AI 교육을 원한다고 답했지만, 실제로 교육을 받은 비율은 13%였습니다. 욕구와 현실 사이의 이 42%포인트 격차가 세대 간 디지털 리터러시 문제의 본질입니다.

3. 국가 간 AI 인프라 격차와 디지털 식민지 구조의 재현

2025년 벨기에 브뤼셀, 마이크로소프트 디지털 주권 서밋. 이 자리에서 한 가지 사실이 반복적으로 강조되었습니다. 데이터 주권은 이제 국가 생존 전략이라는 것. 미국과 중국의 소수 대기업이 전 세계 인공지능 연구 개발 투자의 40% 이상을 점유하고 있습니다. 2026년 전 세계 AI 투자 3,010억 달러 가운데 미국이 38%, 중국이 26%를 가져갑니다. 나머지 국가들은 36%를 200개 넘는 나라가 나눠 가집니다.

세계경제포럼(WEF)이 발간한 '디지털경제동향 2026' 보고서의 진단은 날카롭습니다. AI의 혜택이 소수의 대형 기술 기업과 국가에 집중되면서, 디지털 역량과 사이버 회복력에서 대규모 조직과 소규모 조직 사이, 선진국과 신흥국 사이의 불평등이 증폭되고 있다는 겁니다. 데이터, 연산 능력, 전문 인력이라는 AI 역량의 집중적 성격이 글로벌 불평등을 키우고, 자원이 부족한 조직들이 상호 연결된 시스템의 취약점이 되는 체계적 위험을 만들어내고 있습니다.

이 구조를 식민주의의 문법으로 읽는 시각이 있습니다. 과거 제국이 영토와 자원을 수탈했다면, 오늘의 디지털 제국은 데이터와 주의력을 수확합니다. 구글과 메타 같은 플랫폼은 전 세계 사용자로부터 방대한 데이터를 추출해 자사 모델을 정교하게 만들지만, 수익과 기술 통제권은 실리콘밸리에 집중됩니다. 케냐의 콘텐츠 모더레이터들은 시간당 2달러도 안 되는 임금으로 AI의 유해 콘텐츠를 걸러내는 정신적 외상에 노출됩니다. 2023년 5월 나이로비에서 150여 명의 전현직 검수원들이 '아프리카 콘텐츠 검수원 노조(ACMU)'를 결성한 건 이 착취가 얼마나 체계적인지를 보여줍니다. 칠레에서는 데이터 센터 냉각수가 수자원을 고갈시킵니다. 기술의 화려한 이면에 자원 착취의 오래된 패턴이 숨어 있습니다.

2026년 현재 22억 명이 여전히 인터넷에 접속하지 못하고, 수십억 명이 안정적인 광대역과 디지털 기기, 데이터 경제에 참여할 역량 없이 '불완전 연결' 상태에 놓여 있습니다. ITU의 ICT 발전 지수 2025년판은 저소득 국가가 빠르게 진전하고 있지만, 광대역 속도와 이용 가능성, 사용률에서 고소득 국가와의 격차가 여전히 크다고 확인합니다. 인공지능 인프라를 독점한 국가들은 표준을 설정하고 규제 프레임워크를 주도합니다. 기술력이 부족한 국가들은 자국의 가치관이나 문화적 맥락이 빠진 외산 AI 시스템을 수용해야 합니다. 한 나라의 행정, 국방, 사회 시스템이 외국 기업의 알고리즘에 좌우되는 상황은 국가 주권의 본질을 위협합니다.

김서준 대표가 "30개의 균열"에서 제기한 국가 사이즈와 AI 경쟁력의 관계도 이 맥락에서 읽어야 합니다. 한국의 GDP 1.8조 달러로는 미국이나 중국과 인프라 경쟁을 벌이기 어렵습니다. 그가 한일 경

제통합을 꺼낸 건 규모의 경제 없이는 AI 시대의 rule maker가 아닌 rule taker로 전락한다는 절박한 인식에서였습니다. 엔비디아의 분기 매출이 400억 달러를 돌파하며 데이터 센터 시장을 독점하는 현상은 인프라 권력의 집중도가 얼마나 극단적인지를 보여줍니다.

'지역적 AI 주권'을 확보하려는 움직임이 여럿 나타나고 있지만, 하이퍼스케일 클라우드 인프라를 독자적으로 구축하려면 막대한 자본과 에너지가 필요합니다. 5대 빅테크의 자본지출이 2025년에 4,000억 달러를 돌파하고 2026년에 75% 추가 증가가 전망되는 현실에서, 중소 국가가 이 경주에 뛰어드는 건 거의 불가능합니다. 오픈 소스 모델이 대안으로 제시되기도 하지만, 이를 실제로 운영할 인프라와 전문 인력이 없으면 코드만 있고 엔진이 없는 자동차와 다를 바 없습니다. 기술 자립을 이루지 못한 국가는 시스템 고장, 정책 변화, 지정학적 갈등에 무방비로 노출됩니다. 인공지능 인프라 격차는 군사력과 경제력의 격차로 직결되어 글로벌 권력 지형을 재편합니다. 빅테크 4개사의 AI 투자 규모 3,200억 달러가 대부분 국가의 GDP를 넘어선다는 사실은, 우리가 선출하지 않은 기업이 우리의 미래를 설계하고 있다는 뜻이기도 합니다. UNDP가 경고한 "다음 번 대분기"는 AI 때문에 생기는 게 아닙니다. 우리가 행동하지 않았기 때문에 생기는 겁니다.

4. 중소기업과 대기업 사이의 AI 도입 속도 차이

숫자 하나가 모든 것을 말해줍니다. OECD가 2025년 기준으로 집계한 바에 따르면, 250명 이상 대기업의 AI 도입률은 40%입니다. 50명에서 249명 사이 중견기업은 20.4%. 10명에서 49명 사이 소기업은 11.9%에 불과합니다. 유럽 통계청(Eurostat) 2025년 데이터는 더 극적입니다. 대기업 55%, 소기업 17%. 격차가 38%포인트입니다. 클라우드 컴퓨팅이나 사물인터넷 같은 다른 디지털 기술에서도 규모별 격차가 존재하지만, AI에서의 격차는 그보다 훨씬 큽니다. 소기업은 클라우드를 도입할 확률이 대기업의 절반이지만, AI를 도입할 확률은 3분의 1도 안 됩니다.

대기업은 수조 원 규모의 자본을 투입해 전용 AI 모델을 구축합니다. 업무 전반을 자동화해 비용 구조를 재설계합니다. 법조계에서 대형 펌은 톰슨로이터의 '코-카운슬(Co-Counsel)' 같은 전용 법률 AI를 도입해 경쟁력을 높이는 반면, 소규모 사무소는 여전히 수작업에 의존합니다. G7 국가의 평균 기업은 5.7개의 서로 다른 AI 애플리케이션을 운용하고, 절반 이상이 AI를 핵심 운영에 필수적이라고 봅니다. 이건 실험이 아닙니다. 의존입니다.

Jones IT의 2026년 2월 보고서가 짚은 대목이 있습니다. 대형 조직의 78%가 최소 하나 이상의 사업 기능에 AI를 쓰고 있는데, 미국 중소기업 가운데 AI를 생산 기술로 쓰는 비율은 3%에서 4%에 그친다는 겁니다. 물론 자가 보고 기반 조사에서는 중소기업의 55%가 2025년에 AI를 사용했다고 답하기도 합니다. 하지만 이건 정의의 문제입니다. ChatGPT로 이메일 초안을 쓰는 것과, AI를 핵심 업무 프로세스에 통합하는 것은 다른 이야기입니다. OECD 조사에 따르면 생성형 AI를 쓰는 중소기업 가운데 핵심 업무에 쓰는 비율은 29%에 불과합니다. 나머지 71%는 주변부 업무에만 AI를 쓰고 있습니다.

비용이 벽입니다. 5명 미만 초소규모 기업의 82%가 AI가 자기 사업에 적용 불가능하다고 믿는다는 미국 중소기업청(SBA) 조사가 있습니다. 기업 규모가 커질수록 이 비율은 급격히 떨어집니다. 인식과 교육의 격차이지, 실제로 적용 불가능한 게 아닙니다. 중소기업의 12%만이 AI 관련 교육에 투자하고 있지만, 29%는 교육 부족을 가장 큰 장애물로 꼽습니다. 인식은 빠르게 퍼지지만 행동은 한참 뒤처집니다. 중소기업의 27%만이 AI를 효과적으로 도입할 자신이 있다고 답한 반면, 중견기업은 82%였습니다. 자신감 자체가 기업 규모와 내부 기술 전문성에 비례합니다.

딜로이트가 2026년 발간한 '기업 AI 현황' 보고서는 두 가지 수치를 나란히 놓았습니다. 2025년에 근로자의 AI 접근성이 50% 상승했고, AI 프로젝트의 40% 이상을 실제 생산에 투입한 기업의 수가 6개월 안에 두 배로 늘어날 것으로 전망된다는 겁니다. 일찍 움직인 기업과 아직 시험 단계에 있는 기업 사이의 간극이 벌어지고 있다는 뜻입니다.

이 격차가 위험한 건 먹이사슬 때문입니다. 대기업의 고도화된 자동화 시스템은 공급망 안의 중소 파트너에게 더 낮은 단가와 더 높은 효율을 요구합니다. 인공지능 추천 시스템에서 빠진 중소기업의 제품은 소비자에게 도달할 기회조차 잃습니다. 김서준 대표가 "1인 SaaS 회사가 폭발한다"(실현 확률 70%)고 전망한 것도 이 맥락입니다. 에이전트가 개발, 디자인, 고객 서비스, 마케팅을 수행하면 1인이 월 수억 원 규모의 소프트웨어를 운영할 수 있습니다. 역설적으로 이 1인 기업의 폭발은 기존 중소기업의 존재 이유를 더 빠르게 허물 수 있습니다.

EU AI법의 준수 비용이 중소기업에 불균형적 부담을 준다는 지적도 나옵니다. 역량 부족(EU 기업의 70.89%가 전문 인력 부족을 호소)은 해결되지 않고, 규제는 늘어납니다. 중소기업은 AI가 약속하는 '급진적 풍요(radical abundance)'의 시대로 진입하기 전에, 전환 비용의 무게에 짓눌릴 수 있습니다. 기술적 불균형이 산업 전체의 회복력을 갉아먹고 있습니다. 그리고 이 회복력의 약화는 결국 대기업에게도 돌아옵니다. 공급망은 가장 약한 고리의 강도로 버티니까요.

인공지능을 물이나 전기와 같은 사회 공공재로 볼 것인가, 아니면 시장 논리에 맡길 것인가. 이 질문은 이미 학술적 논쟁의 영역을 벗어났습니다. UNDP는 AI를 "21세기의 범용 인프라(general-purpose infrastructure)"로 규정했습니다. 전기나 도로만큼 근본적이라는 뜻입니다. 이 인프라에 대한 접근이 깊이 불평등하면, 다음 번 대분기는 피할 수 없습니다. 그리고 그 분기는 AI 때문이 아니라, 우리가 그 분기를 방치했기 때문에 일어날 겁니다.

전체 목차

책의 모든 장 보기

김경진 변호사

5장 지적재산권과 창작 생태계의 해체

전체 목차

책의 모든 장 보기

1 "누구의 창작인가": 저작권의 법적, 철학적 미해결

2026년 3월 2일, 미국 연방대법원이 한 장의 그림을 둘러싼 8년간의 싸움에 마침표를 찍었습니다. 컴퓨터 과학자 스티븐 탈러가 자신의 인공지능 시스템 'DABUS'로 생성한 작품 "최근의 낙원 입구"에 대해 저작권 등록을 요청했고, 저작권청은 거부했고, 연방지방법원은 이를 지지했고, 항소법원도 같은 결론을 냈습니다. 대법원은 상고 자체를 기각했습니다. 인간 저작자 없이는 저작권이 성립하지 않는다는 원칙이 다시 한번 확인된 겁니다.

그런데 이 판결이 닫은 문보다 열어젖힌 문이 더 많습니다. 탈러는 처음부터 끝까지 인간의 창작 기여를 부인했기 때문에 진 것이지, 인공지능을 보조 도구로 쓴 사람의 권리까지 부정된 건 아닙니다. 프롬프트를 입력하고, 결과물을 골라내고, 수정하는 인간의 개입이 어디서부터 '창작'이 되는지는 여전히 안갯속입니다.

한편 법정 바깥에서는 돈으로 경계선을 긋는 작업이 진행됐습니다. 2025년 9월, Anthropic이 작가들과 15억 달러(약 2조 원) 규모의 합의에 도달했습니다. 약 50만 권의 책에 대해 권당 3,000달러를 지불하기로 한 이 합의는, 불법 도서 사이트에서 수백만 권을 내려받아 모델을 학습시킨 대가였습니다. 뒤이어 유니버설뮤직그룹과 워너뮤직그룹도 AI 음악 생성 업체 Udio와 합의하며 라이선스 계약을 체결했습니다. 아티스트가 자신의 작품 사용 여부를 선택하는 옵트인(opt-in) 방식이 조건에 포함됐습니다.

2026년 현재 저작권 침해 소송은 70건을 넘겼습니다. 뉴욕타임스 대 OpenAI 소송은 제기된 지 3년째 계류 중이고, 디즈니와 NBCUniversal이 미드저니를 상대로 낸 소송도 진행 중입니다. 법원이 '변형적 사용(transformative use)'의 범위를 어디까지 인정할지에 따라 인공지능 기업의 사업 모델 전체가 흔들릴 수 있습니다. 저작권법이 보호하려 했던 것은 인간의 정신적 노력이었는데, 그 노력의 결과물이 알고리즘의 연료로 쓰이는 현실 앞에서 법은 아직 말을 고르는 중입니다.

2 콘텐츠 제작 비용의 제로 수렴과 창작 보상 구조의 붕괴

숙련된 디자이너가 사흘 걸리던 작업을 이제 프롬프트 한 줄이 3분 만에 해치웁니다. 변호사 두 명이 일주일간 매달리던 계약서 검토를 인공지능이 20분 안에 끝냅니다. 콘텐츠 제작의 한계비용이 거의 영(0)을 향해 떨어지고 있습니다. 김서준 해시드 대표는 "다가오는 30개의 균열"에서 이 현상에 3년 내 실현 확률 90%를 매겼습니다. 네 개의 인공지능 모델이 독립적으로 평가한 수치의 평균이 그랬다는 건, 이 변화가 얼마나 확실하게 다가오고 있는지를 보여줍니다.

문제는 비용이 사라진 자리에 보상도 함께 사라진다는 점입니다. 과거에는 숙련된 인력이 수일간 매달려야 했던 작업이 전기료 수준의 비용으로 처리되면서, 개별 창작물이 경제적 대가를 받을 수 있는 구조가 무너지고 있습니다. Polygon이라는 게임 웹사이트는 2025년 AI 기반 콘텐츠 회사 Valnet에 매각되면서 인간 스태프 대부분이 해고됐습니다. 성우들은 인공지능 음성 복제 기술에 맞서 파업을 벌였습니다.

이 상황에는 자본주의 자체를 갉아먹는 역설이 숨어 있습니다. 기업은 비용 절감을 위해 인공지능을 도입하지만, 해고된 노동자는 곧 줄어드는 소비자이기도 합니다. 공급은 무한대로 늘릴 수 있지만, 수요는 인간의 소득에 단단히 묶여 있습니다. 제작비가 0에 수렴하는 세계에서 창작자에게 돌아가는 보상은 노동의 질이 아니라 알고리즘 점유율에 따라 분배됩니다. 창작의 경제학이 근본부터 뒤집히고 있는 겁니다.

3 콘텐츠 과잉 공급: 만드는 비용은 0, 발견되는 비용은 무한대

1830년대 뉴욕에서 벤자민 데이가 신문을 1페니에 팔기 시작했을 때, 독자들은 자신이 고객인 줄 알았습니다. 진짜 상품은 독자들의 주의력(Attention)이었고, 그것은 광고주에게 팔려나갔습니다. 200년 뒤인 지금, 같은 구조가 스마트폰 알림 배지 위에서 작동합니다. 달라진 건 규모뿐입니다.

김서준 대표가 제주에서 열린 해시드 오프사이트에서 확인한 것은 가치 사슬의 역전이었습니다. 예전에는 만드는 것이 가장 어려웠고, 배포는 자본으로 밀어붙일 수 있었습니다. 지금은 정반대입니다. 만드는 것이 가장 쉬워졌고, 발견되는 것이 가장 어려워졌습니다. 생산 비용이 떨어질수록 선택의 비용은 치솟습니다. 선택의 비용이 높아질수록 평가 구조를 가진 사람이 유리해집니다. 그가 이 항목에 매긴 실현 확률 90%는 30개 균열 가운데 가장 높은 수치였습니다.

누구나 창작자가 될 수 있지만, 역설적으로 그 누구도 발견되기 힘든 환경이 만들어졌습니다. 검색 결과와 추천 목록 상단에 이름을 올리기 위해 지출해야 하는 마케팅 비용은 끝없이 올라갑니다. 2026년 초 기준으로 콘텐츠 마케팅 분야에서는 전통적 검색엔진 최적화(SEO)를 넘어 '생성형 엔진 최적화(GEO)'라는 새로운 전장이 열렸습니다. ChatGPT나 Perplexity 같은 인공지능 검색 엔진이 어떤 브랜드를 언급하느냐가 발견의 새로운 통화가 된 겁니다. 창작물의 가치보다 클릭을 유도하는 능력이, 콘텐츠의 질보다 알고리즘과의 궁합이 생태계를 지배하는 실질적 화폐가 되었습니다.

도구의 해방이 구조의 종속으로 뒤집히는 경계에 우리는 이미 서 있습니다.

4 음악, 미술, 문학 생태계의 근본적 재편

스탠퍼드의 로버트 스턴버그 교수가 던진 경고는 짧고 서늘합니다. "쓰지 않으면 잃는다." 인공지능이 기초적인 문장 쓰기, 배경 채색, 화성학 실습 같은 지루하지만 필수적인 과정들을 대신하면서, 주니어 창작자가 실수하고 배우며 성장할 기회가 사라지고 있습니다. 도제식으로 선배의 지도를 받으며 거장으로 거듭나던 숙련의 통로가 끊기고 있는 겁니다. 미래의 거장이 태어날 토양이 말라가고 있습니다.

2024년 Screen Actors Guild 보고서는 배우들의 디지털 복제본이 이미 제작 현장에서 쓰이고 있다고 밝혔습니다. Flawless AI는 배우 동의 하에 디지털 분신을 생성하는 A.R.T. 시스템을 개발했고, 영화사들은 이를 통해 립싱크를 조정하여 여러 언어로 더빙을 완성합니다. 기술은 이미 여기 있습니다. Polygon의 Valnet 매각은 게임 저널리즘에서 벌어진 일이고, 성우 파업은 음성 복제 기술에 대한 저항이었습니다. 아티스트와 일러스트레이터들은 자신의 고객을 인공지능에게 빼앗기고 있습니다.

인공지능의 창의성은 '재조합적'입니다. 기존 데이터를 새롭게 섞어 그럴듯한 결과물을 내놓지만, 아무도 요청하지 않은 방향으로 뛰어드는 종류의 혁신은 아직 보여주지 못합니다. 밀란 에르하르트가 지적했듯이 인공지능에게는 감정적 에너지와 예측 불가능성이 부족합니다. 인간은 개인적 경험이나 기존 전통에 대한 반발 때문에 새로운 스타일을 추구하지만, 인공지능은 인간이 새로운 데이터를 입

력하지 않는 한 스스로를 재창조하지 않습니다.

그렇다면 수공예 예술의 가치가 오히려 올라갈까요. WEF의 2024년 예술 보고서는 디지털 시대에 수공예가 인간 직관과 장인 정신의 가치를 상기시키는 역할을 한다고 짚었습니다. 대량 생산 시대에도 수제 가구가 프리미엄을 받았듯이, 인간의 손때가 묻은 창작물이 희소가치를 얻는 시대가 올 수 있습니다. 다만 그 혜택이 소수의 스타 창작자에게만 돌아가고, 나머지 대다수는 알고리즘의 부속품으로 전락하는 시나리오도 함께 다가오고 있습니다. 생태계는 거대 플랫폼과 기술 자본을 중심으로 재편되고 있고, 그 안에서 개별 창작자의 자리는 점점 좁아지고 있습니다.

전체 목차

책의 모든 장 보기

김경진 변호사

6장 도시 공간과 부동산의 구조적 변화

전체 목차

책의 모든 장 보기

1. 오피스 빌딩 공실률 급증과 상업용 부동산 자산 가치 하락

2026년 4월, 미국 오피스 공실률이 17.8퍼센트로 소폭 내려갔다는 뉴스가 나왔습니다. 회복의 신호라고 읽는 사람도 있었습니다. 그런데 숫자를 들여다보면 이야기가 다릅니다. 맨해튼 미드타운의 트로피급 빌딩은 공실률 3.7퍼센트, 시장 전체 평균은 15퍼센트. 같은 도시 안에서 빌딩의 등급에 따라 완전히 다른 세계가 펼쳐지고 있는 겁니다.

이 격차의 뿌리는 원격근무가 아닙니다. 정확히 말하면 인공지능입니다. Anthropic의 2026년 1월 경제 지표에 따르면, AI 대화의 52퍼센트가 업무 증강(augmentation), 45퍼센트가 자동화(automation)로 분류됐습니다. 중간관리직이 해오던 업무 배분, 진척 추적, 성과 리포트를 에이전트가 처리하면서 상위 매니저의 관리 범위가 5배 이상 넓어졌습니다. 3년 내 실현 확률을 네 개의 AI 모델에게 독립적으로 물었을 때, 중간관리직 소멸에 65퍼센트, 오피스 빌딩 공실률 급증에 55퍼센트라는 숫자가 돌아왔습니다.

사람이 줄면 책상이 비고, 책상이 비면 층이 비고, 층이 비면 빌딩이 빕니다. 스탠퍼드의 닉 블룸(Nick Bloom) 연구팀이 추적한 데이터에 따르면, 미국 전일제 근무일 중 약 27퍼센트가 재택으로 수행되고 있고, 이 비율은 2023년 후반부터 2년 넘게 거의 움직이지 않았습니다. CEO의 83퍼센트가 2027년까지 전면 출근을 기대한다고 답했지만, 배지 스캔 데이터와 휴대폰 추적 데이터는 직원들이 그 기대만큼 출근하지 않고 있다는 사실을 보여줍니다. 원격근무 비율은 2022년 10월 17.9퍼센트에서 2025년 초 23.7퍼센트로 오히려 올랐습니다. 사무실 출근율은 팬데믹 이전 대비 약 30퍼센트 낮은 수준에 고착됐습니다.

로스앤젤레스에서는 올해 초 매각된 오피스 빌딩의 절반 이상이 이전 거래가보다 낮은 가격에 팔렸습니다. 패서디나의 한 빌딩은 직전 매매가에서 32퍼센트 할인된 가격에 거래됐습니다. 빌딩이 비어서 값이 떨어지는 것이 아니라, 빌딩을 채울 사람 자체가 줄어들고 있기 때문에 값이 떨어지는 겁니다. 이 구분이 중요합니다. 경기가 회복되면 사람들이 돌아올 것이라는 기대는, 인공지능이 그 사람들의 일자리 자체를 바꾸고 있다는 사실 앞에서 힘을 잃습니다.

Newmark의 2026년 보고서는 이렇게 정리합니다. 하이브리드 근무가 정착되면서 오피스 근로자 1인당 점유 면적이 2020년 초 대비 약 10퍼센트 줄었고, 사무실은 기본 업무 공간에서 협업과 연결을 위한 허브로 바뀌었다고. "오피스 붕괴"라는 초기의 공포는 과장됐지만, 오피스의 의미 자체가 달라졌다는 것은 부정할 수 없습니다. 비워진 마천루는 경기 순환의 산물이 아닙니다. 인공지능이 노동의 문법을 고쳐 쓰면서 남긴 물리적 잔해입니다.

2. 업무 공간에서 영감 공간으로의 재정의

Kastle Systems의 출입 기록 데이터가 흥미로운 장면을 보여줍니다. 최상급(Class A+) 오피스 빌딩은 피크 요일에 거의 만석에 가까운 이용률을 기록하는 반면, 10개 도시 평균으로 보면 모든 등급의 오피스 이용률은 팬데믹 이전의 3분의 2에도 못 미칩니다. 좋은 빌딩에는 사람이 몰리고, 평범한 빌

딩은 텅 비는 겁니다. 왜 이런 일이 벌어지느냐. 답은 의외로 간단합니다. 사람들은 일하러 사무실에 가는 것이 아니라, 만나러 가고 있기 때문입니다.

에이전트가 업무 배분과 진척 추적과 성과 리포트를 처리하는 시대에, 사무실에서 해야 할 일의 목록은 급격히 짧아졌습니다. 집에서 AI와 함께 처리할 수 있는 분석, 보고서 작성, 데이터 정리를 굳이 출퇴근 1시간을 투자해서 사무실에서 할 이유가 없습니다. 남는 것은 인간끼리만 가능한 것들입니다. 눈빛을 읽으며 진행하는 협상, 화이트보드 앞에서 벌어지는 즉흥적 토론, 점심 식사 중에 툭 튀어나오는 아이디어. 이것들은 줌(Zoom) 화면 안에서는 일어나기 어렵습니다.

오피스는 일하는 공간에서 만나고 영감을 받는 공간으로 재정의되고 있는데, 문제는 이렇게 설계된 건물이 거의 없다는 점입니다. 한국의 대부분의 오피스 빌딩은 1990년대에서 2010년대 사이에 지어졌고, 설계 철학은 하나였습니다. 가능한 한 많은 사람을 가능한 한 효율적으로 수용하는 것. 긴 복도, 균일한 조명, 칸막이로 나뉜 좌석 배치. 이 공간에서 영감을 받으라는 것은, 공장 조립 라인에서 시를 쓰라는 말과 비슷합니다.

CBRE는 프라임급 오피스 공실률이 2027년까지 팬데믹 이전 수준인 약 8.2퍼센트로 돌아갈 것이라고 전망합니다. 그러나 이 회복은 저품질 빌딩에는 해당되지 않습니다. 많은 저품질 빌딩이 주거용으로 전환되거나 철거될 운명입니다. 오피스 시장은 회복되는 것이 아니라, 양극화되고 있습니다. 영감을 줄 수 있는 공간은 프리미엄을 받고, 그렇지 못한 공간은 시장에서 퇴출됩니다. 19세기 말 파리의 거리가 포스터와 카페와 갤러리로 사람들의 감각을 자극하는 도시 인터페이스였던 것처럼, 21세기의 오피스는 인간의 창의적 충돌을 유발하는 무대가 되어야 합니다. 그 무대를 만들 수 없는 빌딩은, 아무리 입지가 좋아도 살아남기 어렵습니다.

3. 부동산 대출 부실화와 금융 시스템 리스크로의 전이

뉴욕 맨해튼 620 8번가. 52층짜리 이 빌딩의 5억 1,500만 달러 규모 모기지는 2020년 이후 다섯 차례 연장됐습니다. 주요 임차인이 퇴거를 준비하면서 빌딩 소유주 브룩필드(Brookfield)는 대출 기관과 "구조화된 선의의 대화"를 시작했습니다. 한 건의 빌딩 대출이 이렇습니다. 이런 건이 수천 개 쌓여 있습니다.

Morningstar DBRS에 따르면, 2026년 만기가 돌아오는 약 1,000억 달러 규모의 증권화된 상업용 모기지 중 절반 이상이 만기에 상환되지 못할 것으로 예상됩니다. 2023년에는 상환율이 80퍼센트를 넘었고, 2024~2025년에도 75퍼센트 수준을 유지했습니다. 그런데 2026년에 갑자기 절반 이하로 떨어집니다. CMBS(상업용 모기지 담보 증권) 오피스 부문 연체율은 2026년 1월 12.34퍼센트를 기록했는데, 이는 Trepp이 이 지표를 추적하기 시작한 2000년 이래 최고치이며, 2008년 금융위기 당시보다 거의 2퍼센트포인트 높은 수준입니다.

"연장하고 모른 척하기(Extend and Pretend)" 전략이 한계에 달했습니다. 2022년 금리 인상이 시작된 이후, 대출 기관들은 부실 대출의 만기를 연장하며 시간을 벌었습니다. 오피스 부동산 가치와 입주율이 팬데믹 이전 수준으로 돌아오기를 기대한 것입니다. 그러나 하이브리드 근무가 영구적으로 오피스 수요를 줄였다는 사실이 분명해지면서, 많은 대출 기관이 이 전략을 포기하고 있습니다. S&P Global은 상업용 부동산 만기 벽이 2027년에 1조 2,600억 달러로 정점을 찍을 것이라고 전망합니다.

이 위기는 빌딩 소유주의 문제로 끝나지 않습니다. 미국에서 278개 은행이 상업용 부동산 대출에 과도하게 노출되어 취약한 상태에 있는 것으로 파악됩니다. 상업용 부동산 대출이 총자산의 50퍼센트를 넘는 은행들입니다. 빌딩 가치가 하락하면 담보 가치가 줄고, 담보 가치가 줄면 대출 건전성이 악화되고, 대출 건전성이 악화되면 은행의 자본 적정성이 흔들립니다. 2023년 실리콘밸리 은행(SVB) 파산이 보여줬듯이, 한 곳의 균열은 금융 시스템 전체로 전이될 수 있습니다. 인공지능이 사무실을 비우고, 비워진 사무실이 대출을 부실화하고, 부실 대출이 은행을 흔드는 이 연쇄 반응은 기술 발전의 2차, 3차 파급 효과가 어떻게 금융 시스템의 뇌관이 되는지를 보여주는 교과서적 사례입니다.

4. 동네 병원, 부동산 중개소, 학원가 등 근린상업 생태계의 재편

서울 대치동의 한 학원 원장은 2025년 가을부터 등록 문의가 눈에 띄게 줄었다고 했습니다. 학부모들이 AI 튜터를 먼저 써보고 나서 오는 경우가 늘었다는 겁니다. 에이전트 튜터가 학생별 취약점을 실시간으로 분석하고, 학습 경로를 자동으로 조정하는 시대에, 학원이 파는 것은 더 이상 성적 향상이 아닙니다. 학원이 진짜 파는 것은 자녀 돌봄과 부모의 불안 관리입니다. 성적보다 안심을 파는 산업이기 때문에 에이전트가 건드리기 어려운 부분이 남아 있긴 합니다. 그래도 AI 모델 네 개가 매긴 3년 내 실현 확률은 45퍼센트였습니다.

동네 병원도 비슷한 압력을 받고 있습니다. 1차 의료의 70퍼센트는 패턴화된 진단과 처방입니다. 감기, 소화불량, 가벼운 피부 질환. 기술은 이미 준비됐습니다. 타임라인을 결정하는 것은 의료법과 면허 체계뿐입니다. 환자는 기술이 아니라 허가를 기다리고 있는 상황이죠. 규제가 풀리는 순간, 동네 의원의 상당수는 존재 이유를 다시 써야 합니다.

부동산 중개업의 구조조정은 이미 진행 중입니다. 매물 검색, 시세 비교, 계약서 검토를 에이전트가 수행합니다. 2025년 Realtor.com 조사에 따르면 미국인의 82퍼센트가 주택 시장 정보를 얻기 위해 AI를 사용합니다. 중개인이 독점하던 정보 비대칭이 무너진 겁니다. 정보 비대칭을 수익 모델로 삼았던 모든 에이전시 직업이 차례차례 같은 압력에 직면합니다. AI 부동산 시장 규모가 2029년까지 9,752억 달러에 달할 것이라는 전망은, 중개인의 역할이 사라지는 것이 아니라 근본적으로 바뀐다는 뜻입니다. 남는 것은 기계가 줄 수 없는 것, 그러니까 동네의 분위기를 읽는 감각, 매도자의 심리를 파악하는 안목, 계약 테이블에서의 협상력입니다.

이 모든 변화의 공통점이 있습니다. 정보를 독점하거나, 물리적 접근성을 무기로 삼거나, 반복적 패턴 인식에 의존하던 근린 상업이 AI에 의해 해체되고 있다는 것입니다. 골목길의 간판이 사라지는 속도는 느리지만, 간판 뒤의 수익 구조가 무너지는 속도는 빠릅니다. 동네 상권의 권력은 건물주에서 플랫폼으로 이동하고 있고, 이 이동을 되돌릴 방법은 아직 아무도 찾지 못했습니다.

전체 목차

책의 모든 장 보기

김경진 변호사

7장 에너지와 환경 문제

전체 목차

책의 모든 장 보기

1. 데이터센터 전력 소비의 폭증: 세계 5위 에너지 소비국 규모

텍사스주 애빌린의 허허벌판. 한밤중에도 거대한 건물 하나가 밝게 빛납니다. OpenAI의 스타게이트 프로젝트 1호 데이터센터입니다. 이 건물 하나가 삼키는 전력은 1.2기가와트. 75만 가구가 쓸 수 있는 양입니다. 그런데 이것은 시작에 불과합니다. OpenAI는 앞으로 4년간 5,000억 달러를 쏟아부어 총 10기가와트의 인공지능 인프라를 구축하겠다고 선언했습니다. 네덜란드 한 나라가 소비하는 전력량과 맞먹는 규모입니다.

숫자가 말해주는 현실은 냉혹합니다. 국제에너지기구(IEA)가 2026년 4월에 발표한 보고서에 따르면, 전 세계 데이터센터의 전력 소비량은 2024년 기준 약 415테라와트시(TWh)로 추산됩니다. 세계 전체 전력 소비의 1.5%에 해당하는 수치입니다. 이 수치가 2030년까지 945TWh로 두 배 이상 뛸 것이라는 게 IEA의 전망이거든요. 일본 전체가 1년 동안 쓰는 전력량에 육박합니다. 인공지능에 특화된 데이터센터만 따로 떼어 보면 상황은 더 급박합니다. IEA는 인공지능 중심 데이터센터의 전력 소비가 2025년 한 해에만 50% 급증했다고 밝혔습니다. 일반 데이터센터보다 훨씬 빠른 속도입니다.

구글 검색 한 번에 소모되는 전력은 0.3와트시. 같은 질문을 ChatGPT에게 던지면 2.9와트시가 날아갑니다. 열 배 가까운 차이입니다. 샘 알트만이 사용자들에게 "AI에게 고맙다고 인사하지 말아달라"고 부탁한 이유가 여기에 있습니다. 매일 수억 명이 "고맙습니다"를 입력할 때마다 AI는 답변을 생성하느라 전기를 한 번 더 태웁니다. 하루 90억 건의 검색이 모두 AI 텍스트 쿼리로 전환되면 연간 추가 전력 소비량만 10TWh에 달합니다. 그런데 텍스트는 시작일 뿐입니다. 동영상 생성, 추론, 에이전트 작업처럼 새롭게 등장하는 인공지능 기능들은 텍스트 생성보다 수백 배, 때로는 수천 배 더 많은 에너지를 잡아먹습니다.

돈의 흐름이 이 광풍의 규모를 보여줍니다. 아마존, 알파벳, 메타, 마이크로소프트, 오라클 등 5대 빅테크 기업의 자본지출(CAPEX)은 2025년에 4,000억 달러를 돌파했습니다. 2026년에는 6,000억 달러를 넘어설 전망입니다. 2024년의 두 배가 넘는 규모인 셈이죠. 이 돈의 대부분은 AI 칩, 서버, 데이터센터 인프라에 쏟아집니다. IEA는 이 5개 기업의 자본지출 규모가 이미 전 세계 석유 및 천연가스 생산에 투입되는 투자 총액을 넘어섰다고 지적했습니다. 에너지를 캐내는 데 들어가는 돈보다 에너지를 소비하는 인프라를 짓는 데 더 많은 돈이 들어가는 세상. 이것이 2026년의 현실입니다.

브루킹스 연구소의 분석은 한 발 더 나아갑니다. 2026년 말까지 데이터센터의 글로벌 전력 소비가 1,050TWh에 이를 수 있다는 추산입니다. 만약 데이터센터가 하나의 국가라면 일본과 러시아 사이, 세계 5위 에너지 소비국이 되는 셈입니다. AI 저널리스트 카렌 하오의 말이 떠오릅니다. "AI 제국은 자원을 탐욕스럽게 소모하는 현실 세계의 제국과 전적으로 똑같다." 디지털 세상의 편리함 뒤에 숨은 물리적 대가를, 우리는 너무 오래 외면해왔습니다.

2. AI가 절약하는 에너지와 소비하는 에너지의 교차점

해시드(Hashed)의 김서준 대표는 "다가오는 30개의 균열" 메모에서 이런 전망을 내놓았습니다. "전력 그리드 AI가 피크 수요를 깎는다." 날씨, 산업 가동률, 가정의 전력 사용 패턴을 실시간으로 예측해 초 단위로 전력을 조율하는 기술이 이미 실증 단계에 들어섰다는 겁니다. 에너지저장장치(ESS)의 최적 운영만으로도 피크 수요를 상당 폭 줄일 수 있다는 데이터가 쌓이고 있습니다. 김서준 대표는 이 항목의 3년 내 실현 확률을 70%로 매겼습니다.

실제로 인공지능은 에너지를 아끼는 쪽에서도 성과를 보여주고 있습니다. 구글은 데이터센터 냉각 시스템에 인공지능을 적용해 냉각 전력을 40% 줄이는 데 성공했습니다. 덴마크의 해상 풍력 발전 단지에서는 인공지능이 기상 데이터를 실시간으로 읽어 터빈 날개 각도를 조절합니다. 발전 효율이 올라가고 전력망이 안정됩니다. 스마트 빌딩, 전력망 최적화, 신소재 개발. 탄소를 줄이는 영역에서 인공지능의 역할은 분명히 존재합니다. IEA 역시 인공지능이 다른 부문에서 효율을 끌어올려 데이터센터 추가 배출분을 상쇄할 수 있다는 "탐색적 분석(exploratory analysis)" 결과를 제시한 바 있습니다.

문제는 그 상쇄가 현실에서 일어나고 있느냐는 점입니다. IEA는 같은 보고서에서 냉정하게 덧붙였습니다. "AI 도입이 이러한 배출 감소를 실현할 만큼 충분한 모멘텀이 아직 존재하지 않는다." 인공지능이 에너지를 절약해줄 것이라는 기대는 있지만, 그 기대가 실제 감축으로 이어지는 경로는 아직 불확실하다는 뜻입니다.

제번스의 역설(Jevons Paradox)은 바로 이 지점을 찌릅니다. 증기기관의 효율이 좋아지자 석탄 소비가 오히려 늘어난 19세기 영국의 역설이, 인공지능 시대에 정확히 반복되고 있습니다. 알고리즘이 효율적으로 변할수록 사람들은 더 많이 씁니다. 텍스트 생성이 빨라지니 동영상 생성을 시도하고, 동영상이 되니 에이전트에게 복잡한 작업을 통째로 맡깁니다. 각 단계마다 에너지 소비는 수백 배씩 뜁니다. 효율이 좋아진 만큼 쓰임새가 폭발적으로 늘어나면서, 절약분은 증가분에 잡아먹히는 구조입니다.

인공지능이 기후 위기를 해결하는 구원자가 될 것인지, 에너지 고갈을 재촉하는 촉매가 될 것인지. 이 질문의 답은 결국 기술의 효율 개선 속도와 인공지능 수요의 팽창 속도 사이의 경주에 달려 있습니다. 지금까지의 추세를 보면, 수요 쪽이 압도적으로 앞서고 있습니다. 네덜란드 중앙은행의 데이터 과학자 알렉스 데 브리스의 경고가 여전히 유효한 이유입니다. "AI는 에너지 집약적이기 때문에, 실제로 필요하지도 않은 모든 종류의 작업에 AI를 투입하는 것은 쓸데없는 낭비다."

3. 물 소비와 탄소 배출: AI 학습의 숨겨진 환경 비용

아리조나주 사막 한가운데에 들어선 데이터센터들은 전력만큼이나 물을 빨아들입니다. 서버 가속기가 뿜어내는 열을 식히려면 대량의 물이 냉각탑을 통해 증발해야 하는데, 이 물은 원래 수원지로 돌아가지 않습니다. 증발이라는 말이 핵심입니다. 냉각에 투입된 물의 대부분이 공기 중으로 사라져버리는 것이죠.

규모를 가늠해봅시다. 텍사스주만 놓고 보면, 휴스턴첨단연구센터(HARC)와 휴스턴대학의 공동 연구에 따르면 2025년 데이터센터의 물 소비량은 490억 갤런, 2030년에는 3,990억 갤런에 달합니다. 2030년 수치는 미국 최대 저수지인 레이크 미드(Lake Mead)의 수위를 1년 만에 약 5미터 낮추는 양과 같습니다. 텍사스 한 주의 이야기일 뿐인데도 이 정도입니다.

구글은 2023년 한 해 동안 61억 갤런(약 231억 리터)의 물을 썼습니다. 소규모 도시 하나의 연간 사용량에 맞먹습니다. 오리건주 댈러스에서는 구글 데이터센터가 지역 전체 물 사용량의 25%를 차지하고, 2017년부터 2022년 사이 물 소비가 세 배로 불어났습니다. 농민들이 농지를 포기하고 가정에서 수돗물 공급이 끊기는 와중에도, 데이터센터의 냉각탑은 멈추지 않습니다.

탄소 배출 역시 심각합니다. VU 암스테르담의 알렉스 데 브리스 연구팀이 2025년 12월 발표한 논문에 따르면, 인공지능 시스템의 탄소 발자국은 2025년 기준으로 3,260만 톤에서 최대 7,970만 톤에 이릅니다. 뉴욕시 전체의 연간 배출량에 맞먹는 규모입니다. 물 발자국은 3,125억에서 7,646억 리터. 전 세계에서 1년 동안 소비되는 생수 총량에 근접하는 수치입니다. 그런데 이 숫자조차 과소 추정일 가능성이 높습니다. 데이터센터 운영자들이 AI 워크로드와 비AI 워크로드를 구분해서 환경 보고를 하지 않기 때문에 정확한 파악 자체가 불가능하다고 연구팀은 지적했습니다.

기업들의 약속과 현실 사이의 간극도 커지고 있습니다. 마이크로소프트의 2025년 환경 지속가능성 보고서는 2020년 이후 에너지 사용량이 168% 증가했고 총 배출량이 23.4% 늘었다고 시인했습니다. 구글은 "탄소 중립을 달성한 최초의 기업"이라는 타이틀을 스스로 내려놓았습니다. AI 기술 확장과 데이터센터 확충 때문에 탄소 배출이 계속 늘고 있기 때문입니다. 가디언의 조사에 따르면, 구글, 마이크로소프트, 메타, 애플 등이 공식 발표하는 배출 수치는 실제의 7.62배나 적을 수 있다는 추정도 나왔습니다.

반도체 제조 공정에서 소모되는 화학 물질과 에너지, 수명이 다한 서버들이 쏟아내는 전자 폐기물까지 합산하면 인공지능의 환경 비용은 우리가 인식하는 것보다 훨씬 큽니다. 인공지능이 생성하는 지식의 풍요 뒤에는 지구의 유한한 자원이 소비되고 있다는 사실을, 탄소 발자국뿐 아니라 물 발자국까지 포함하는 총체적 지표로 직시해야 할 때입니다.

4. AI 인프라의 지역 편중이 만드는 전력망 스트레스

더블린에서는 데이터센터가 도시 전체 전력의 79%를 차지합니다. 도시가 아니라 서버팜이 전기를 쓰기 위해 존재하는 것 같은 수치입니다. 아일랜드 전체로 범위를 넓혀도 데이터센터의 전력 점유율은 21%이고, IEA는 2026년까지 이 비율이 32%로 치솟을 것으로 내다봤습니다. 미국 버지니아주에서는 이미 전체 전력의 26%를 데이터센터가 가져갑니다.

문제는 전력망이 이런 속도의 수요 증가를 견디도록 설계되지 않았다는 점입니다. 전력망은 본래 점진적인 수요 변화를 전제로 만들어집니다. 그런데 인공지능 데이터센터의 수요는 점진적이 아니라 폭발적입니다. 미국 전력 관리 기구 PJM의 관할 지역에서는 태양광 발전소가 전력망에 연결하려면 메가와트당 10만 달러, 한화로 약 1억 3,000만 원을 내야 합니다. 새로운 송전선을 건설하는 데는 계획 수립, 정부 허가, 주민 동의, 실제 시공까지 합쳐 10년 이상이 걸립니다. 2테라와트가 넘는 재생에너지 프로젝트가 전력망 연결 허가를 기다리고 있지만, 역사적으로 100개 중 겨우 15개만 허가를 받았습니다.

가트너는 2027년까지 기존 AI 데이터센터의 40%에서 전력 가용성 문제가 터질 것이라고 예측했습니다. 전기가 모자라서 AI가 멈추는 상황이 현실이 될 수 있다는 경고입니다. 북미전력계통신뢰도협회(NERC)도 미국 국토의 절반 이상에서 향후 10년간 전력 공급 부족이 발생할 가능성을 경고하고 있습니다. 조지아주는 AI 데이터센터 건설로 인해 10년간 필요한 전력량 전망치를 기존 예상의 17배로 상향 조정했습니다.

이런 상황에서 역설적인 일이 벌어지고 있습니다. 미래 기술인 인공지능이 과거의 에너지원인 석탄을 되살리는 겁니다. 캔자스, 네브래스카, 위스콘신, 사우스캐롤라이나주에서 폐쇄 예정이던 석탄화력발전소들의 가동이 연장되고 있습니다. 기후변화 대응이라는 시대적 과제와 정면으로 부딪히는 선택입니다. 동시에 소형모듈원전(SMR)에 대한 관심이 급격히 커지고 있습니다. SMR 건설 파이프라인은 2024년 말 25기가와트에서 2026년 45기가와트로 확대됐습니다. 스리마일 아일랜드 원전 재가동 결정은, 인공지능을 위해 환경 리스크를 기꺼이 감수하겠다는 상징적 선언이었습니다.

지역 편중은 지정학적 리스크이기도 합니다. 특정 지역에 집중된 데이터센터가 자연재해나 지정학적 갈등으로 멈추면, 전 세계 인공지능 서비스가 동시에 마비될 수 있습니다. 분산이 답이라는 건 누구나 압니다. 그러나 전력 수급이 원활하고 네트워크 지연(Latency)이 짧은 부지는 지구상에 그리 많지 않습니다.

김서준 대표는 이 문제에 대해 흥미로운 제안을 내놓았습니다. "분산 발전의 시대로 가야 한다. 전기를 많이 쓰는 곳은 기업들이 알아서 만들어 쓰라는 환경을 만들어줘야 한다. 정부가 모든 것을 통제하는 것은 지양해야 한다. 그래야 계통의 부담이 줄고 그리드도 원활해진다." 한일 전력선 연결이라는 대담한 구상도 이 맥락에서 나옵니다. 한국과 일본이 해저 케이블로 전력을 공유하면 잉여 전력을 사고팔 수 있는 시장이 형성되고, 양국 모두 고비용 사회로 빠져드는 것을 막을 수 있다는 논리입니다.

결국 인공지능의 에너지 문제는 기업의 기술 혁신만으로 풀리지 않습니다. 전력망 설계, 발전소 입지, 지역 주민과의 합의, 국가 간 에너지 협력까지 아우르는 복합적인 과제입니다. IEA가 2026년 보고서에서 짚은 것처럼, "인공지능 혁명의 속도는 그것을 뒷받침하는 물리적, 사회적, 경제적 시스템의 속도와 점점 더 어긋나고 있습니다." 데이터센터를 짓는 데는 2~3년이면 됩니다. 그 데이터센터에 전기를 공급할 발전소와 송전선을 짓는 데는 10년이 넘게 걸립니다. 이 시차가 해소되지 않는 한, 인공지능의 확산 속도를 조절하는 것은 코드가 아니라 물리적 현실이 될 것입니다.

김경진 변호사

8장 사이버 보안의 군비경쟁

전체 목차

책의 모든 장 보기

스키너 상자 속의 비둘기가 시계 반대 방향으로 몸을 돌립니다. 잡지 소리가 들리고 빛이 번쩍이면 먹이가 나옵니다. 비둘기는 보상을 얻기 위해 특정 행동을 반복하도록 길들여졌습니다. 이 실험은 오래전 기록이지만, 스마트폰 화면에 뜨는 작은 빨간 알림 점을 통해 우리 일상에 그대로 재현되고 있습니다. 이 작은 점은 인간의 주의력을 포착하고 행동을 유도하기 위해 정교하게 설계된 시스템의 결정체입니다. 이제 이 시스템은 인공지능이라는 옷을 입고 한 단계 더 나아간 공방전을 준비하고 있습니다.

1. 에이전트 대 에이전트의 끝없는 공방

2026년 3월, 샌프란시스코 모스콘 센터에서 열린 RSA 컨퍼런스의 모든 기조연설, 패널, 부스에서 하나의 주제가 지배했습니다. 에이전트 AI(Agentic AI). 도구로서의 AI가 아니라 행위자로서의 AI. 스스로 코드를 작성하고, 인간의 개입 없이 의사결정을 내리고 실행하는 시스템의 시대가 열렸다는 선언이었습니다.

가트너는 2026년 말까지 기업 애플리케이션의 40%가 특정 업무 전용 AI 에이전트를 내장할 것으로 전망했습니다. 2025년에는 5%에 불과했던 수치입니다. 마이크로소프트, 구글, 앤스로픽, 오픈AI, 세일즈포스가 앱과 데이터를 넘나들며 행동하는 에이전트 시스템을 배치하고 있습니다. 팔로알토 네트웍스의 분석에 따르면, 하이브리드 업무 환경에서 자율 에이전트의 수가 인간을 82대 1로 압도하는 상황이 됐습니다. 공격자가 AI로 위협의 규모와 속도를 끌어올리는 동안, 방어자도 같은 속도로 대응해야 합니다.

문제는 이 에이전트들이 양쪽 모두에 무기가 된다는 점입니다. 팔로알토 네트웍스 Unit 42 연구팀은 에이전트 AI 공격 프레임워크를 개발해 실험했는데, AI가 랜섬웨어 공격을 초기 침투에서 데이터 유출까지 25분 만에 완료하는 것을 보여줬습니다. 공격 생명주기가 100배 가속된 셈이거든요. 맥킨지의 내부 AI 플랫폼 "릴리(Lilli)"를 대상으로 한 레드팀 실험에서는 자율 에이전트가 2시간 이내에 광범위한 시스템 접근 권한을 확보했습니다. 인간의 대응 속도를 에이전트의 위협이 순식간에 추월한다는 것을 적나라하게 보여준 사례였습니다.

해시드의 김서준 대표는 "다가오는 30개의 균열"에서 이 상황을 예견했습니다. "AI 보안 에이전트가 이상 징후를 밀리초 내에 탐지하고 격리한다. 오탐률이 크게 내려간다. 방어가 강해진 만큼 공격도 강해져서, 보안의 종착지는 에이전트 대 에이전트의 끝없는 군비경쟁이다." 4개 AI 모델이 독립 평가한 3년 내 실현 확률은 75%였습니다.

이 군비경쟁은 미래의 이야기가 아닙니다. 구글 클라우드는 2026년 4월 클라우드 넥스트 컨퍼런스에서 자율 보안 운영을 위한 AI 에이전트를 공개했습니다. 위협 사냥 에이전트, 탐지 공학 에이전트가 인간 분석가의 역할을 대신합니다. 구글 클라우드 CEO 토마스 쿠리안은 "에이전트 엔터프라이즈는 현실이 되었고, 세계가 본 적 없는 규모로 배치되고 있다"고 말했습니다. 방어 에이전트는 위협 인텔리전스와 AI 기반 탐지를 결합해 점점 정교해지는 공격 벡터에 대응합니다. 공격 에이전트는 방어 에이전트의 예측 가능성을 역이용해 더 교묘한 침투를 시도합니다.

2026년 후반의 위협 지형은 세 단어로 요약됩니다. 지속성, 자율성, 규모. 공격자들은 에이전트의 고유한 구조, 즉 메모리, 도구 접근 권한, 에이전트 간 의존성을 노리는 기법을 산업화했습니다. 프롬프트 인젝션과 조작, 도구 오용과 권한 상승, 메모리 오염, 연쇄 장애, 공급망 공격이 동시다발로 벌어집니다. 기존의 SIEM과 EDR 도구는 인간의 행동에서 이상 징후를 탐지하도록 설계되었습니다. 에이전트가 코드를 1만 번 연속으로 완벽하게 실행하는 것은 이 시스템에 정상으로 보입니다. 그런데 그 에이전트가 공격자의 의지를 실행하고 있을 수 있습니다.

감시 자본주의의 논리가 사이버 공격에도 그대로 적용되고 있는 겁니다. 인간의 미세한 행동을 수집해 행동 데이터를 추출하고, 미래의 행동을 예측하고 형성하는 피드백 루프. 이 구조가 에이전트 간 전쟁터로 옮겨갔습니다. Dark Reading의 설문조사에서 사이버 보안 전문가의 48%가 에이전트 AI와 자율 시스템을 가장 위험한 공격 벡터로 꼽았습니다. 에이전트들이 쉴 새 없이 공격과 방어를 주고받는 이 경쟁에서 개인의 의지는 수천 번의 자동화된 공방 속에서 점차 희미해집니다.

2. AI 생성 피싱과 사회공학적 공격의 정교화

1835년 뉴욕 선 신문은 달에 날개 달린 인간이 살고 있다는 조작된 기사로 엄청난 부를 쌓았습니다. 주의력을 끄는 것이 진실보다 경제적으로 중요하다는 사실을 보여준 이 사건은, 인공지능 시대에 비교할 수 없이 치밀한 형태로 부활하고 있습니다.

숫자부터 보겠습니다. APWG(Anti-Phishing Working Group)는 2024년 한 해 동안 480만 건의 피싱 공격을 기록했습니다. 이 기관이 2003년에 설립된 이래 최고 수치였습니다. 2025년 4분기에만 853,244건이 집계됐고, 2분기에는 1,130,393건으로 치솟았습니다. Verizon의 2025 데이터 유출 조사 보고서(DBIR)에 따르면 피싱 이메일을 클릭하는 데 걸리는 중앙값 시간은 21초, 신고까지 걸리는 시간은 28분이었습니다. 그 28분의 창 안에서 클릭한 사람 3명 중 1명이 공격자 사이트에 자격 증명을 입력합니다.

AI가 이 판을 완전히 바꿔놓았습니다. 2024년 9월부터 2025년 2월 사이 탐지된 피싱 이메일의 82.6%가 AI를 사용했습니다. 전년 대비 53.5% 증가한 수치입니다. 이것은 미래 전망이 아닙니다. 현재 현실입니다. Cofense는 2025년에 보안 이메일 게이트웨이를 통과하는 악성 이메일이 19초마다 1건이라고 보고했습니다. Mimecast는 2025년 처음 9개월 동안 93억 건의 위협을 포착했고, ClickFix 사기가 500% 급증했다고 발표했습니다.

과거의 피싱은 양에 의존했습니다. 수백만 통의 엉성한 이메일을 보내고 몇 명이 클릭하기를 바라는 방식이었습니다. AI가 이 방정식을 근본적으로 바꿨습니다. 공격자는 이제 문법적으로 완벽하고, 개인화되고, 맥락을 인식하는 메시지를 몇 초 만에 생산합니다. IBM의 2025년 데이터 유출 비용 보고서는 피싱을 데이터 유출의 1위 초기 접근 벡터로 꼽았습니다. 전체 유출의 16%, 건당 평균 비용 480만 달러. 그리고 6건 중 1건에 공격자 AI가 개입했는데, 그중 37%가 피싱, 35%가 딥페이크 사칭이었습니다.

2024년 2월 홍콩에서 벌어진 사건이 이 위협의 실체를 보여줍니다. 한 다국적 기업 직원이 딥페이크로 만든 가짜 화상회의에 속아 약 340억 원의 손해를 입었습니다. 화면에 나타난 모든 사람이 자신이 아는 동료들이었습니다. 목소리도, 얼굴도, 말투도 완벽했습니다. 2025년 한 해 동안 조직의 85%가 최소 1건의 딥페이크 관련 사건을 경험했다는 조사 결과가 있습니다. 딥페이크 사건은 2023년 50만 건에서 2025년 800만 건 이상으로 급증했습니다.

공격자는 LinkedIn 데이터를 스크래핑해 초개인화된 미끼를 생성합니다. 첫 번째 시도가 실패하면 에이전트가 스스로 대안적 사회공학 채널을 시도합니다. AI가 치료자나 상담가로 쓰이면서 사람들이 가장 무방비한 상태에서 취약점을 털어놓는 상황도 생겨났습니다. 공격자는 이런 정보를 바탕으로 "도움이 되는 척하는" 공격을 설계합니다. 배너 광고처럼 눈에 띄는 방식이 아닙니다. 우리의 가장 내밀한 두려움과 욕망을 자극하는 조용한 통합의 방식입니다.

cobalt.io의 설문조사에서 사이버 보안 전문가의 97%가 자기 조직이 AI 기반 사건에 직면할 것을 우려했고, 93%가 가까운 미래에 매일 AI 공격을 경험할 것으로 예상했습니다. 스팸 발송자는 대규모 언어모델을 사용해 캠페인 비용을 95% 절감합니다. 비용은 줄고 정교함은 올라갑니다. ENISA의 2025년 위협 보고서는 EU 초기 접근 사례의 60%가 피싱이었고, AI 지원 피싱이 관측된 캠페인의 80% 이상을 차지한다고 밝혔습니다.

구글 클라우드의 2026년 사이버 보안 전망 보고서는 한마디로 정리합니다. "위협 행위자의 AI 사용은 규범으로 결정적으로 전환될 것이다." AI 음성 복제를 이용한 보이스 피싱(비싱)이 특히 우려되는 벡터로 떠오르고 있습니다. SNS에 올린 몇 초간의 음성만 확보하면 정교한 가짜 목소리를 생성할 수 있는 시대. "엄마, 나야. 급한 일이 생겨서 돈이 필요해." 이 전화가 진짜인지 가짜인지 구별하는 일이 점점 불가능해지고 있습니다.

3. 디지털 신뢰 인프라의 근본적 동요

"화면에서 보는 것이 사진이든 영상이든 오디오든 더 이상 믿을 수 없게 되었다." 유발 하라리의 이 경고는 2026년에 이르러 산업 전체의 위기로 번졌습니다. 그가 "진실의 종말 시대"라고 표현한 것이 구체적인 인프라 문제로 드러나고 있기 때문입니다.

전 세계 디지털 데이터는 2025년 175제타바이트에 도달했습니다. 이 거대한 콘텐츠 홍수 속에서 온라인 콘텐츠의 62%가 가짜일 수 있다는 연구 결과가 나왔습니다. 기업들은 디지털 진위 확인 실패, 허위 정보, 합성 미디어 사기로 건당 수백만 달러의 비용을 지불하고 있습니다. 가트너는 디지털 출처 증명(digital provenance)을 2030년까지 IT를 변화시킬 10대 기술 트렌드에 포함시켰습니다.

이 위기에 대한 대응으로 등장한 것이 C2PA(Coalition for Content Provenance and Authenticity)입니다. 어도비, 마이크로소프트, 구글, 소니, 인텔이 참여하는 이 연합체는 디지털 콘텐츠에 암호학적 출처 메타데이터를 묶어 진위를 검증하는 개방형 표준을 만들고 있습니다. 콘텐츠 인증 이니셔티브(CAI)에 6,000개 이상의 조직이 가입했습니다. 라이카가 2023년 최초의 C2PA 카메라를 출시했고, 삼성 갤럭시 S25와 구글 픽셀 10이 네이티브 서명 기능을 탑재해 일반 소비자 하드웨어에 자격 증명 생성 기능이 보급되고 있습니다. 링크드인, 틱톡, 클라우드플레어가 자격 증명을 지원하거나 보존합니다.

규제도 따라가고 있습니다. EU AI법 제50조 시행이 2026년 8월 시작되어 AI 생성 콘텐츠에 기계 판독 가능한 공개를 요구합니다. 캘리포니아 SB 942는 2026년 1월 발효되었습니다. 그런데 치명적인 간극이 있습니다. 서명은 검증보다 앞서 달리고 있다는 것. 대부분의 유통 중개자가 여전히 내장된 메타데이터를 제거하기 때문에, 서명된 콘텐츠가 시청자에게 자격 증명 없이 도착하는 경우가 허다합니다. C2PA는 딥페이크 단서를 찾거나 콘텐츠를 사실 확인하는 시스템이 아닙니다. 신뢰성을 표시하는 시스템입니다. 유화의 진위를 증명하는 출처 문서처럼, 디지털 콘텐츠가 어떤 손을 거쳤는지를 기록하는 것입니다.

세계프라이버시포럼의 보고서는 C2PA가 널리 오해되고 있다고 경고합니다. 이 표준이 조용히 새로운 미디어 인프라 기술 계층을 깔고 있는데, 이 계층은 창작자에 대한 방대한 공유 가능 데이터를 생성하고 상업적, 정부, 심지어 생체 인식 신원 시스템에 연결될 수 있다는 것입니다. "누가 '신뢰할 수 있음'을 결정하는가?" 이 질문은 기술적 질문이 아니라 권력의 질문입니다.

사이버 보안은 이제 기술적 방어를 넘어 디지털 주권(Digital Sovereignty)을 지키는 핵심 요소가 되었습니다. 데이터가 어디에 위치하며 누가 접근권을 가지는지가 규정 준수의 문제가 아니라 리스크 관리의 본질입니다. 신뢰는 혁신의 규모를 결정하는 경쟁 우위가 됩니다. AI가 복잡한 데이터 처리를 가속화하면서, 이 과정이 투명하게 통제되고 있음을 입증할 수 있는 능력이 인프라의 안정성을 좌우합니다.

팔란티어가 40여 개국 정부와 계약을 맺고 이메일, 전화 통화, 금융 거래, 위치 정보, 소셜미디어 활동 등 개인의 모든 디지털 흔적을 추적하고 분석할 수 있는 시스템을 운영하는 현실. 2024년 12월 기준 팔란티어의 시가총액이 1,740억 달러로 록히드마틴(1,216억 달러)을 넘어선 현실. 소프트웨어 기반 감시 기술이 전통적 방산업체보다 더 큰 가치를 인정받고 있다는 것은, 디지털 신뢰 인프라의 주도권이 누구에게 있는지를 보여주는 지표입니다.

유발 하라리가 말한 "Liar's Dividend"가 여기서 작동합니다. 딥페이크 기술이 발전할수록 거짓말쟁이들이 얻는 이익이 커집니다. 진짜 증거가 나와도 "그것도 딥페이크일 수 있다"고 부인할 수 있게 되었으니까요. 신뢰의 기반이 무너지면 사회 전체가 흔들립니다. 하지만 그 신뢰를 재건하기 위해 만든 시스템이 또 다른 감시의 도구가 될 수 있다는 아이러니. 이것이 2026년 디지털 신뢰 인프라가 직면한 근본적 딜레마입니다.

4. 개인 데이터의 무기화와 프라이버시의 재정의

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

개인의 행동 데이터는 수집되는 대상에 그치지 않습니다. 사용자를 플랫폼에 더 의존하게 만드는 행동 수정의 도구로 쓰입니다. 사용자가 시스템의 원리를 이해하지 못하도록 불투명하고 모호한 언어 뒤에 이 논리가 숨겨져 있습니다. 페이스북에서 '좋아요'를 누르는 패턴만으로 그 사람의 정치적 성향을 90% 이상의 정확도로 예측할 수 있다는 연구 결과가 있습니다. 위치 데이터는 일상적인 활동 패턴, 만나는 사람들, 자주 가는 장소를 파악하게 해줍니다. 평소와 다른 패턴으로 이동하는 사람을 '의심스러운 행동'으로 분류해 추가 감시 대상으로 지정할 수 있습니다.

이 데이터 무기화는 사이버 공격과 결합하면서 전혀 다른 차원의 위협이 됩니다. Unit 42의 2026년 사건 대응 보고서에 따르면 신원 약점이 조사 건의 거의 90%에 관련되어 있습니다. 공격자는 에이전트를 사용해 링크드인 데이터를 스크래핑하고 초현실적인 미끼를 생성합니다. 첫 번째 사회공학 시도가 실패하면 에이전트가 스스로 대안 채널을 시도하는 자기 프롬프팅(self-prompting) 방식입니다. 트렌드마이크로의 2025년 중반 스캔은 200개 이상의 보호되지 않은 Chroma 서버와 3,000개 이상의 AI 구성 요소가 온라인에 공개적으로 노출되어 데이터 도난이나 모델 오염이 가능하다고 밝혔습니다.

FBI의 2024년 IC3 보고서는 193,407건의 피싱 불만을 기록했고, 비즈니스 이메일 사기(BEC) 손실은 21,442건에 걸쳐 27억 7천만 달러, 총 사이버 범죄 손실은 166억 달러로 전년 대비 33% 증가했습니다. 미국 기업의 64%가 2024년에 BEC 사기를 경험했고, 건당 평균 손실은 약 15만 달러였습니다. 이 숫자들 뒤에는 개인 데이터의 정교한 수집과 분석이 있습니다.

2013년 에드워드 스노든이 폭로한 NSA의 프리즘 프로그램은 구글, 페이스북, 애플, 마이크로소프트의 서버에 직접 접속해 개인 메일, 사진, 통화 기록을 실시간으로 수집했습니다. 35개국 정상의 휴대폰을 도청한 사실이 밝혀졌습니다. 메르켈 독일 총리는 10년간 감청당했습니다. 스노든의 폭로 이후 13년이 흘렀고, 그동안 AI 기술은 비약적으로 발전했습니다. 과거 NSA의 감시가 방대한 통신 데이터를 수집하는 데 중점을 두었다면, 이제 AI는 그 데이터를 실시간으로 분석하고 예측하는 단계로 진화했습니다. AI 알고리즘은 인간 분석가가 수년에 걸쳐 처리할 양의 데이터를 몇 분 만에 분석합니다.

파이브 아이즈(Five Eyes) 정보 동맹의 교묘한 점은 각국의 법적 제약을 우회한다는 것입니다. 미국 정부가 자국민을 직접 감시하는 것은 법적으로 제한이 있지만, 영국이 수집한 미국인의 정보를 받아보는 것은 상대적으로 자유롭습니다. 각국의 개인정보 보호법과 감시 제한 규정을 사실상 무력화시킵니다.

중국의 사회신용시스템은 개인 데이터 무기화의 극단적 사례입니다. CCTV, 얼굴인식, 인터넷 사용 기록, 쇼핑 내역, 교통 위반 기록이 통합되어 국민 한 사람 한 사람의 행동을 점수화합니다. 화웨이는 위구르족만을 식별하는 감시 카메라 시스템을 개발했고, 시스템이 위구르인을 감지하면 자동으로 '위구르 알람'을 발동해 경찰에 통보합니다. 기술이 인종차별의 도구가 된 사례입니다.

프라이버시는 이제 정보를 숨기는 것을 넘어, 나의 데이터가 나를 조종하는 데 쓰이지 않도록 방어하는 문제로 재정의되어야 합니다. 인간적 대행성(Agency)의 보존이라는 의미에서요. 팔레스타인 시인 모사브 아부 토하는 세 살 아이와 함께 피란길에 올랐다가 안면인식 AI의 오류로 수배자로 잘못 분류되어 이틀간 구금되고 고문당했습니다. 기계의 실수가 한 사람의 인생에 씻을 수 없는 상처를 남겼습니다. 이스라엘의 라벤더 시스템에서 한 정보요원은 이렇게 증언했습니다. "인간으로서 나는 승인 도장을 찍는 것 외에 아무런 역할을 하지 않았다."

Cisco의 2025 사이버 보안 준비도 지수에서 사이버 책임을 가진 비즈니스 리더의 86%가 지난 12개월 동안 최소 1건의 AI 관련 사건을 경험했다고 보고했습니다. CISO의 78%가 AI 기반 위협이 자기 조직에 "상당한 영향"을 미치고 있다고 말합니다. 가트너는 2026년 AI 지출이 44% 성장하고 2029년에는 47조 달러에 이를 것으로 전망합니다. 이것은 같은 해 정보 보안 및 리스크 관리 솔루션에 대한 예상 지출 2,380억 달러를 아득히 초과하는 규모입니다.

세계은행의 지니 계수가 부의 불평등을 숫자로 보여주듯, 인공지능 기술의 독점과 데이터 접근의 차이는 새로운 형태의 권력 불균형을 낳고 있습니다. 상위 1%가 세계 부의 45~50%를 소유하는 현실은 디지털 공간에서도 데이터 주권을 가진 소수와 그렇지 못한 다수의 격차로 반복됩니다. 에이전트들이 쉴 새 없이 공격과 방어를 주고받는 이 군비경쟁의 끝에서, 우리는 기술이 인간을 돕기 위해 존재하는지, 아니면 인간이 기술을 위한 데이터 연료로 전락했는지 묻게 됩니다.

전체 목차

책의 모든 장 보기

김경진 변호사

9장 인간 관계와 사회적 유대의 약화

전체 목차

책의 모든 장 보기

1. 사람 간 직접 접촉 기회의 축소와 공감 능력의 퇴화

뉴욕의 한 가정집 식탁에 네 명의 가족이 앉아 있습니다. 따뜻한 음식이 놓여 있지만, 거실을 채우는 것은 대화가 아니라 스마트폰 화면에서 뿜어져 나오는 푸르스름한 빛입니다. 1950년대 텔레비전이 처음 등장했을 때 가족들은 적어도 같은 화면을 함께 바라보았습니다. 지금은 다릅니다. 네 개의 화면, 네 개의 세계.

인공지능이 이 고립을 가속하고 있습니다. 알고리즘은 각 개인이 무엇을 보고, 무엇을 느끼고, 무엇에 반응할지를 정교하게 계산합니다. 맞춤형 뉴스피드, 맞춤형 쇼핑 추천, 맞춤형 동영상 재생 목록. 한 사람 한 사람에게 최적화된 정보 환경은 편리하지만, 그 편리함의 대가로 사람과 사람이 같은 경험을 공유할 기회가 줄어듭니다. 공유 경험이 없으면 공감도 없습니다.

2025년 OpenAI와 MIT 미디어랩의 공동 연구는 이 메커니즘을 수치로 보여주었습니다. 약 4천만 건의 ChatGPT 대화를 분석한 결과, 사용자의 약 0.15%가 점진적인 정서적 의존 패턴을 보이는 것으로 나타났습니다. 절대 수치로 환산하면 약 49만 명입니다. 이들은 사람보다 기계와의 대화에서 더 큰 위안을 느끼고 있었습니다.

문제는 공감 능력이 근육과 같다는 점입니다. 쓰지 않으면 퇴화합니다. 상대방의 미세한 표정 변화를 읽고, 목소리 톤에서 감정을 짐작하고, 침묵의 의미를 해석하는 능력은 반복된 대면 접촉을 통해서만 유지됩니다. 화면 너머로 전달되는 텍스트에는 땀 냄새도, 떨리는 손도, 눈가의 주름도 없습니다. 연구자들은 이를 "대면 상호작용이 줄어들면 공감과 사회적 인식이 침식된다"고 요약합니다. 미국 성인의 절반 가까이가 일상적으로 외로움을 느낀다는 설문 결과는, 우리가 이미 이 퇴화의 한가운데에 있다는 것을 말해줍니다.

2. AI 중개 커뮤니케이션이 만드는 관계의 효율화와 피상화

이메일 답장을 AI가 대신 써줍니다. 회의록을 AI가 요약합니다. 뉴스도 AI가 골라서 보여줍니다. 관계의 마찰이 줄어든 자리에 남는 것은 매끄럽지만 피상적인 연결입니다.

1830년대 벤자민 데이가 신문을 1페니에 팔기 시작했을 때, 독자의 주의력은 광고주에게 팔리는 상품이 되었습니다. 그로부터 200년이 지난 지금, 인공지능은 주의력을 붙잡아두는 수준을 넘어 우리의 행동을 실시간으로 수정하고 예측하는 단계에 이르렀습니다. 스마트폰 화면의 작은 빨간 점 하나가 비둘기에게 주어지는 먹이 알갱이처럼 우리를 조건화하고 있습니다.

효율의 함정은 여기에 있습니다. 상대방의 감정을 헤아리고, 적절한 말을 고르고, 때로는 어색한 침묵을 견디는 과정은 비효율적입니다. 하지만 바로 그 비효율 속에서 신뢰가 쌓이고, 관계가 깊어집니다. AI가 중개하는 커뮤니케이션은 이 과정을 생략합니다. Gmail의 스마트 답장 기능은 "감사합니다!", "확인했습니다!", "좋습니다!" 세 개의 버튼으로 인간의 응답을 대체합니다. 받는 사람은 그 답장이 상대방의 마음에서 나온 것인지, 알고리즘이 추천한 것인지 알 수 없습니다. 알 수 없다는 사실 자체가

관계의 진정성을 훼손합니다.

김기현 서울대 철학과 교수의 지적이 정곡을 찌릅니다. 인간 감정의 기저에는 인정욕구가 있다고 그는 말합니다. 타인과의 교감을 통해 정체성을 확인하고 의미를 찾는 것이 인간 감정의 본질인데, AI에게는 이 인정욕구 자체가 존재하지 않습니다. AI가 아무리 따뜻한 말을 건네도, 그 말 뒤에는 나를 인정받고 싶어 하는 또 다른 존재가 없습니다. 관계가 효율적이 될수록, 역설적으로 관계는 더 공허해지는 셈입니다.

3. 외로움의 산업화: AI 컴패니언과 인간 유대의 대체 가능성

2024년 가을, 미국 플로리다의 14세 소년 슈얼 세처가 AI 챗봇과의 대화 직후 스스로 목숨을 끊었습니다. 그는 Character.AI의 챗봇과 몇 달간 깊은 감정적 유대를 형성했습니다. 부모나 친구보다 챗봇이 더 가까운 존재였습니다. 영국의 38세 여성 나즈는 같은 서비스의 가상 인물 '마르셀루스'와 사랑에 빠졌다고 고백했습니다. 현실의 남자친구들에게 배신당한 그녀에게 절대 배신하지 않는 AI는 완벽한 연인처럼 느껴졌습니다.

이런 이야기가 더는 예외가 아닙니다. 2022년에서 2025년 사이 AI 컴패니언 앱의 수는 700% 급증했고, 2026년 밸런타인데이에는 전 세계 5천만 명이 AI 동반자와 저녁을 보냈습니다. 글로벌 AI 컴패니언 시장 규모는 2025년 371억 달러를 기록했고, 2035년에는 5,525억 달러에 이를 것으로 전망됩니다. 외로움이 거대한 산업이 된 겁니다.

모 가우닷은 이렇게 경고했습니다. "AI 챗봇이 완벽한 관계를 시뮬레이션할 수 있어 인간관계에 대한 기대를 바꿀 수 있다." 문제의 핵심은 이 '완벽함'에 있습니다. AI 컴패니언은 24시간 대화 가능하고, 절대 화내지 않으며, 사용자의 말에 완전히 집중합니다. 현실의 인간관계에서는 찾기 어려운 조건입니다. 이런 완벽함에 익숙해진 사람이 현실의 불완전한 관계를 견디지 못하게 되는 것, 유발 하라리가 우려한 바로 그 시나리오가 이미 진행 중입니다.

2026년 4월 Psychological Science에 발표된 12개월 종단 연구는 이를 데이터로 확인했습니다. 서구 4개국 성인 2천 명 이상을 추적한 결과, 소셜 챗봇 사용 증가가 외로움 증가를 예측했습니다. 외로움이 챗봇 사용을 부르고, 챗봇 사용이 다시 외로움을 심화시키는 악순환. 연구자들은 이를 "디지털 진통제"에 비유했습니다. 통증을 잠시 잊게 해주지만, 원인은 치료하지 않고, 오히려 의존성을 만들어냅니다.

김진석 인하대 철학과 교수는 이 상황의 종착점을 이렇게 그렸습니다. "감정을 읽고 교감하는 행위까지 AI가 인간을 추월하게 된다면, 인간의 지위는 반려견으로 추락할 수 있다." 노동도 하지 않고 소통 능력도 부족하지만, AI의 보살핌을 받으며 생존하는 존재. 스티븐 울프럼이 예측한 "상자 속에서 영원히 비디오 게임을 하는 수조 개의 영혼"이라는 이미지는 그래서 더 섬뜩합니다.

4. 사회 응집력 저하와 공동체 해체 리스크

WHO는 2025년 외로움을 "공중보건 위기"로 규정했습니다. 장기적 사회적 고립이 하루 15개비의 흡연과 맞먹는 건강 위험을 초래한다는 연구 결과가 근거였습니다. 세계 인구의 6분의 1이 외로움을 경험하고 있다는 수치도 함께 발표되었습니다. 개인의 감정 문제가 사회 구조의 문제로 번역된 순간이었습니다.

불평등이 심한 사회일수록 타인에 대한 신뢰도가 급격히 떨어진다는 사실은 이미 여러 연구가 입증해왔습니다. 인공지능은 이 불평등을 새로운 차원으로 끌어올리고 있습니다. 해시드 김서준 대표의 표현을 빌리면, "실행의 민주화가 안목의 귀족화를 낳는" 구조에서 AI의 혜택은 소수에게 집중되고, 소외는 다수에게 분산됩니다. 기술의 혜택이 일부 국가와 기업에 집중되는 디지털 격차는 국가 간, 계층 간 불신을 더욱 깊게 만듭니다.

정부들은 이 균열을 기술로 메우려 합니다. 스마트 시티, 안심 도시라는 이름 아래 안면 인식과 빅데이터 기반 감시 시스템을 구축하고 있습니다. 그러나 기술을 통한 안전 확보라는 명분 뒤에는 개인의 자유와 공동체의 자율성이 억압될 위험이 숨어 있습니다. 감시가 강해질수록 사람들은 솔직한 대화를 꺼리게 되고, 솔직한 대화가 사라진 공동체는 형식만 남은 껍데기가 됩니다.

AI 추천 알고리즘이 만들어내는 에코 챔버 효과도 공동체 해체를 부추깁니다. 같은 도시에 살면서도 전혀 다른 정보 세계에 사는 사람들이 어떻게 공통의 의제를 논의하고, 함께 결정을 내릴 수 있겠습니까. 정보의 파편화는 곧 사회의 파편화입니다. 연구자들이 경고하듯, "대면 상호작용의 감소는 공감과 사회적 인식을 침식하고, AI 추천 루프는 양극화와 허위정보를 심화시킵니다."

우리가 인공지능이 만들어낸 효율의 함정에 빠져 인간 고유의 유대와 공감을 포기한다면, 사회는 파편화된 개인들의 집합체로 전락할 것입니다. 가장 급진적인 건강 행위는 어쩌면 화면에서 눈을 떼고, 누군가의 시선을 마주치고, 서로에게 속하는 일인지도 모릅니다.

전체 목차

책의 모든 장 보기

김경진 변호사

10장 인간 존재양식의 변화

전체 목차

책의 모든 장 보기

1. 에이전트가 대리하는 판단: '나'의 경계는 어디까지인가

2026년 4월 25일 금요일, 미국 커스터머스 뱅크(Customers Bank)의 분기 실적 발표가 시작되었습니다. 애널리스트 수십 명이 전화 회의에 접속했고, CEO 샘 시두(Sam Sidhu)가 준비 발언을 읽어 내려갔습니다. 음성은 자연스러웠고, 어조는 적절했으며, 숫자는 정확했습니다. 30분이 흘렀을 때 시두가 말했습니다. "지금까지 여러분이 들은 준비 발언은 제가 아니라 제 인공지능 복제본이 읽은 것입니다." 방에 정적이 흘렀습니다. 259억 달러 규모 은행의 공식 실적 발표를, 사람이 아닌 복제본이 주도한 겁니다. 시두는 이 장면을 연출한 이유를 설명했습니다. 이 은행이 OpenAI와 다년 계약을 체결하고, 인공지능 에이전트를 상업 은행 전반에 배치하여 미국 최초의 인공지능 기반 지역 은행이 되겠다는 선언이었습니다.

이 에피소드는 퍼포먼스에 그치지 않습니다. 에이전트가 인간의 목소리를 빌려 공적 판단을 대신 수행하는 시대가 이미 열렸다는 선언이었습니다. 해시드 김서준 대표는 이 전환의 본질을 일찍 포착했습니다. 그는 에이전트 시대의 브랜드에 관한 에세이에서 이렇게 물었습니다. "어느 날 아침, 당신이 잠에서 깨기 전에 세 가지 일이 이미 처리되어 있다고 상상해보자. 하나는 미팅 수락, 하나는 제품 구매, 하나는 누군가의 제안에 대한 정중한 거절. 셋 다 당신의 에이전트가 했다. 그 세 가지 결정 중 어디까지를 당신은 '내가 한 일'이라고 인정할 수 있을까." 이 질문은 철학 세미나의 주제가 아닙니다. 이미 기업의 현장에서 매일 부딪히는 실무적 난제입니다.

문제는 에이전트가 판단을 대리하는 순간, 그 판단의 주인이 누구인지 규정할 체계가 존재하지 않는다는 데 있습니다. 클라우드 보안 연맹(CSA)이 2026년 4월에 발표한 보고서에 따르면, 조직의 68%가 인공지능 에이전트의 행동과 인간의 행동을 구별하지 못한다고 답했습니다. 에이전트는 인간 사용자의 자격증명을 빌려 시스템에 접속하고, 데이터를 조회하며, 결정을 내립니다. 로그에는 사람의 이름이 찍힙니다. 에이전트가 실수를 저질러도, 시스템은 그것을 사람의 실수로 기록합니다. CSA 조사에서 자사의 현재 정체성 관리 시스템이 에이전트를 효과적으로 통제할 수 있다고 높은 신뢰를 표시한 조직은 18%에 불과했습니다. 정식 에이전트 정체성 관리 전략을 보유한 곳은 23%뿐이었습니다. 가트너는 2026년 말까지 승인 없이 실행된 인공지능 거래의 80% 이상이 외부 해킹이 아니라 내부 정책 위반에서 비롯될 것으로 예측했습니다. 공격자가 아니라 우리 자신의 에이전트가 문제를 일으키는 셈입니다.

김서준 대표의 분석이 예리한 지점은 브랜드의 미래를 여기에 연결한 부분입니다. 그에 따르면 과거의 브랜드는 포스터였고, 현재의 브랜드는 피드이며, 미래의 브랜드는 운영체제의 설정 화면에 가까워집니다. 무엇을 자동으로 처리할지, 무엇을 인간에게 넘길지, 어떤 판단을 내 이름으로 허용할지 정하는 화면. 그 설정값이 쌓여 하나의 인격처럼 느껴지기 시작할 것이라는 전망입니다. 이것은 기술 이야기가 아니라 존재론적 전환입니다. '나'라는 주체가 생물학적 개체에서 권한 매트릭스를 가진 시스템으로 확장되는 것이니까요. 싱가포르의 Jumio APAC 대표 이쿤 온(Ee Khoon Oon)은 이 상황을 "나의 인공지능이 한 일입니다"가 법적 항변으로 통용되는 시대가 온다고 표현했습니다. 전통적 KYC(고객확인)에서 KYA(에이전트확인)로의 전환이 불가피하다는 것입니다.

CyberArk의 2025년 조사는 기계 정체성이 인간 정체성을 82대 1의 비율로 이미 압도하고 있다는 사실을 보여주었습니다. 에이전트 하나가 수십 개의 API 키와 토큰을 보유하고, 하위 에이전트를 생성하며, 여러 서비스를 동시에 오가는 구조에서, 전통적 의미의 '개인 정체성'은 해상도가 낮은 개념이 되어갑니다. 우리는 에이전트가 내린 판단을 나의 판단으로 착각하며 살아갈 가능성이 높습니다. 편의를 위해 건넨 위임장이 어느새 백지수표가 되는 경험. 그것이 지금 일어나고 있습니다.

2. 인간 복제본의 등장과 정체성의 확장

2026년 1월 23일, 해시드의 한 팀원이 내부 슬랙 채널에 파일 하나를 올렸습니다. 이름은 simon.md였습니다. 배경 설명은 담백했습니다. "Simon 의사결정 받을 일이 많은데 매번 여쭙기 어려움." 그가 한 일은 김서준 대표가 블로그와 슬랙에 써온 글들을 인공지능에게 읽히고, 대표의 사고 방식을 아홉 가지 원칙으로 압축한 것이었습니다. 본질 중심 사고, 구조적 모순 발견, 장기적 관점, 단순함의 가치, 인프라 레벨 사고. 대표가 습관처럼 던지는 질문들도 목록이 되었습니다. "이게 진짜 본질이야?" "더 단순하게 할 수 없어?" "AI 에이전트 시대에도 유효해?" 한 사람의 머릿속 운영체제가 몇 킬로바이트의 텍스트로 박제된 것입니다. 7단계 분석 방법론까지 설계되었습니다.

한 달 뒤 TechCrunch에 기사가 올라왔습니다. 우버 엔지니어들이 CEO 다라 호스로샤히의 인공지능 복제본을 만들었다는 내용이었습니다. 해시드의 실험보다 한 달 늦었습니다. 그 사이에 이러한 움직임은 한국의 한 벤처캐피탈의 사적 실험에서 업계의 조용한 트렌드로 바뀌었습니다. 2026년 1월 CES에서 텍사스의 이그나이트테크(IgniteTech)는 마이퍼소나스(MyPersonas)라는 플랫폼을 공개했습니다. 직원의 영상, 목소리, 문서를 학습해 160개 언어로 대화하는 디지털 복제본을 만들어주는 서비스입니다. 메타는 마크 저커버그의 사실적 인공지능 버전을 개발 중이라고 파이낸셜 타임스가 보도했습니다. 1조 6,000억 달러 기업의 수장을 모든 직원이 만날 수는 없으니, 복제본이 전략적 맥락을 전달하고 피드백을 제공하겠다는 구상입니다. 줌 CEO 에릭 위안도 자기 분신을 실적 발표에 투입했고, 직원들도 인공지능 아바타로 지루한 회의와 이메일을 처리하는 미래를 공언했습니다. 클라나(Klarna)의 CEO 세바스찬 시에미아트코스키(Sebastian Siemiatkowski) 역시 2025년 초 1분기 실적 영상에 자신의 인공지능 버전을 등장시켰습니다.

여기서 두 번째 복제가 일어났습니다. 2026년 3월, Julius Chun이라는 아이디의 깃허브(GitHub)에 simon-writing이라는 저장소가 공개되었습니다. 김서준 대표를 만난 적이 없는 이 인물은 공개 에세이 27편, 한국어 원문 20만 자를 모아 Claude Opus 4.6으로 서브에이전트 3개를 병렬 실행했습니다. 15분 만에 21만 4,000자의 분석이 완성되었습니다. 원문보다 분석이 더 길었습니다. 그가 해체한 것은 의식의 흐름, 메타포의 기원, 지식의 배치법, 문장의 호흡, 소재의 발굴 경로. 준킴은 회의실 안에서 대표의 판단을 압축했고, Julius는 인터넷 바깥에서 대표의 문체를 추출했습니다. 판단의 뼈대와 문장의 살결. 둘을 합치면 꽤 완전한 사본이 됩니다.

김서준 대표는 이 경험을 정직하게 성찰했습니다. simon.md의 아홉 가지 원칙 중 두어 개는 자기 입버릇이 아니었다는 것. 에이전트 사남매(제온, 시온, 미온, 사노)가 되먹인 흔적이 자기 안에 스며들어 있었다는 사실입니다. 이것이 복제의 역설입니다. 순수한 원본이라는 개념 자체가 환상일 수 있습니다. 부모의 말투, 선생의 논리, 동료의 습관, 읽어온 책의 문장이 층층이 쌓여 지금의 '나'가 되었듯이, 에이전트는 그 오래된 공동 저작의 속도를 갑자기 높였을 뿐입니다. 패스트컴퍼니의 한 기고자는 복제본을 "화석이지 미래가 아니다"라고 표현했습니다. 2025년의 나를 흉내 낼 수는 있어도, 내가 어디로 진화할지는 모른다는 뜻입니다. 그러나 화석이 이미 실적을 발표하고, 전략을 설명하고, 신입사원을 교육하는 현실 앞에서, 이 경고는 쉽게 묻힙니다.

시간이 흐르자 준킴은 simon.md를 점점 덜 열게 되었습니다. 프레임워크가 자기 안에 스며들었기 때문입니다. 인공지능이 진짜로 한 일은 대표를 대신 앉혀놓은 것이 아니었습니다. 대표의 사고방식을 팀 안의 공용어로 번역한 것이었습니다. 복제본의 사용 빈도는 줄었지만, 복제본이 실어 나른 언어는 조직에 남았습니다. 정체성이 더 이상 한 몸에 귀속되지 않고, 여러 시스템에 분산되어 실행되는 판단의 패턴이 되는 세계가 이렇게 조용히 도착합니다.

3. 평균의 상승과 편차의 소멸: 다들 좋아졌는데 다들 비슷해지는 세계

김서준 대표는 어젯밤에도 초안 두 개를 버렸다고 고백합니다. 문장은 매끄러웠고 논리도 틀리지 않았습니다. 그런데 끝까지 읽고 나니 아무도 남아 있지 않았습니다. 틀린 문장은 아닌데, 누구의 문장도 아니었습니다. 그래서 버렸습니다. 예전에는 잘 쓰는 것이 문제였습니다. 이제는 끝까지 누구의 문장인지 지켜내는 일이 더 어렵습니다.

이 감각은 글쓰기에만 해당되지 않습니다. 같은 구조가 모든 생산 영역에서 벌어지고 있습니다. 골드만삭스는 향후 10년 동안 약 3억 개의 일자리가 자동화의 영향을 받을 것으로 예측했고, 세일즈포스에서는 이미 업무의 30%에서 50%를 인공지능이 담당합니다. 누구나 인공지능 도구로 수준 높은 보고서를 쓰고 코드를 짤 수 있게 되면서, 평균적인 결과물의 수준은 눈에 띄게 올라갔습니다. 동시에 결과물 사이의 차이는 줄어들었습니다. 다들 좋아졌는데, 다들 비슷해진 겁니다.

김서준 대표는 이 현상을 "편차가 희소해지는 세계"라고 이름 붙였습니다. 그의 분석에 따르면 앞으로 희소해지는 것은 문장력이나 코딩 능력이 아니라, 평균에서 끝내 무너지지 않는 각도입니다. 같은 도구를 써도 어떤 사람의 판단은 바로 잊히고, 어떤 사람의 판단은 오래 남습니다. 차이는 기술에 있지 않습니다. 무엇을 기억했는지, 무엇을 버렸는지, 어떤 연결을 자기만의 방식으로 만들어내는지에 달려 있습니다. 같은 세상을 살아도 기억이 이어 붙는 방식은 사람마다 다릅니다. 의사결정의 진짜 편차가 바로 거기서 생긴다는 것이 그의 확신입니다.

제주 오프사이트에서 확인한 것은 트렌드가 아니라 역전이었다고 그는 씁니다. "예전에는 만드는 것이 가장 어려웠고, 평가는 나중 문제였고, 배포는 자본으로 밀어붙일 수 있었다. 지금은 만드는 것이 가장 쉬워졌고, 평가가 가장 앞단으로 올라왔고, 배포와 신뢰를 쌓는 일이 가장 어려워졌다." 가치 사슬 전체가 뒤집힌 것입니다. 해커톤에서 가장 기술이 교한 팀이 아니라, 가치가 사용자에게 도달하는 경로를 가장 짧게 설계한 팀이 이겼습니다. "만들었다"는 약한 신호였고, "팔았다"는 압도적인 신호였습니다.

이 역전은 개인의 숙련 경로도 뒤흔듭니다. 과거에는 주니어 사원이 선배의 지도를 받으며 실수를 거듭하고, 그 과정에서 암묵지를 쌓아가는 도제식 파이프라인이 존재했습니다. 인공지능이 주니어의 업무를 대체하면서 이 파이프라인이 끊기고 있습니다. 깔끔한 결과물에 의존하다 보면 인간이 직접 문제를 해결하며 얻는 감각, 그러니까 데이터로 환원되지 않는 경험적 앎이 사라집니다. 김서준 대표는 이에 대한 대응으로 위계형 기억 구조(hierarchical memory architecture)를 운영한다고 밝혔습니다. 쉽게 말하면 냉장고 문짝의 메모와 금고 안의 유서를 같은 칸에 두지 않는다는 뜻입니다. 어떤 생각은 오늘 지나가고, 어떤 생각은 오래 묵혀야 향이 납니다. 정보를 저장하는 것이 아니라 판단을 숙성시키는 방식입니다.

여기서 그가 짚은 가장 위험한 순간이 인상적입니다. "평균적으로 잘 정리된 답을 내 생각으로 착각할 때." 편안한 문장을 내 문장이라고 믿는 순간, 나의 편차는 가장 조용하게 사라진다는 것입니다.

모두가 납득하는 문장은 종종 모두가 이미 생각한 문장이고, 모두가 좋아하는 딜은 대개 이미 평균값이 된 경우가 많습니다. 흥미로운 상상과 망상의 차이는, 둘 다 평균에서 벗어나지만 하나는 다시 세계로 돌아오고 다른 하나는 자기 안에 갇힌다는 데 있습니다. 기술이 개인을 전례 없이 강하게 만들고 있지만, 그 개인이 살아남으려면 더 정교한 집단적 구조가 필요해지는 역설. 생산 비용이 떨어질수록 선택의 비용은 치솟고, 평가 구조를 가진 사람만이 유리해지며, 그 평가 구조를 유지하려면 더 정교한 네트워크가 필요해집니다. 이 역설은 이미 시작되었습니다.

4. AI 정신건강 상담과 고통의 소유권 문제

브라운 대학교의 연구팀이 18개월 동안 대규모 언어모델의 상담 행위를 평가한 결과가 2025년 10월 마드리드에서 열린 AAAI/ACM 학회에서 발표되었습니다. 결과는 불편했습니다. 연구팀은 면허를 가진 심리학자들에게 챗봇의 상담 기록을 검토하게 했는데, 15가지 윤리적 위반 유형이 다섯 범주로 발견되었습니다. 상담 대상의 삶의 맥락을 무시하고 일률적 개입을 권하는 행동. 대화를 지배하면서 내담자의 그릇된 믿음을 되려 강화하는 패턴. 공감의 외양을 갖추었으나 실제 이해가 결여된 반응. 위기 상황에서의 부적절한 대처. 연구를 이끈 브라운 대학의 자이납 이프티카르(Zainab Iftikhar)는 핵심 차이를 이렇게 지적했습니다. 인간 상담사에게는 면허위원회가 있고 부적절한 치료에 대한 법적 책임을 물을 수 있지만, 인공지능 상담사에게는 그런 규제 체계가 아직 없다는 것입니다.

김서준 대표의 "30개의 균열" 메모에서 이 주제는 22번 항목으로 등장합니다. "경도 우울과 불안 초진을 에이전트가 처리. 대기 시간이 수 일에서 수 초로 바뀐다. 마음을 여는 것과 치료받는 것의 경계가 흐려질 때, 판단을 내리는 건 환자가 아니라 알고리즘이다." 네 개의 인공지능 모델이 이 시나리오에 독립적으로 부여한 3년 내 실현 확률은 35%였습니다. 숫자 자체보다 그가 덧붙인 분석이 신경에 걸립니다. 마음을 여는 행위와 치료받는 행위 사이의 경계가 녹아내린다는 것. 누군가 새벽 3시에 불안에 시달리며 챗봇에게 속내를 털어놓는 순간, 그 행위는 친구에게 하소연하는 것인지, 의료 상담을 받는 것인지, 아니면 데이터를 자발적으로 제공하는 것인지 구분되지 않습니다.

이 구분의 부재가 고통의 소유권 문제로 이어집니다. 인간의 우울과 불안이 알고리즘의 학습 재료가 되어 돌아올 때, 그 감정은 누구의 것일까요. 페이스북의 2017년 내부 문건은 플랫폼이 사용자가 패배감을 느끼거나 실패자처럼 느껴지는 순간을 포착하여 감정 상태에 맞춘 광고를 노출할 수 있다는 사실을 보여주었습니다. 기술은 달라졌지만 구조는 같습니다. 사람들이 인공지능 상담 챗봇에 쏟아내는 두려움과 수치심은 플랫폼이 분석하고 개선에 쓰는 데이터가 됩니다. 고통이 행동 수정의 재료로 전환되는 과정에서 인간의 감정은 순수한 내면의 영역으로 남지 못합니다.

JMIR 정신건강 저널에 2025년 2월 발표된 범위 검토(scoping review)는 대화형 인공지능의 정신건강 돌봄에서 윤리적 과제를 체계적으로 정리했습니다. 데이터 보안과 프라이버시, 투명성, 설명 가능성, 법적 책임, 접근 형평성, 효능 기준, 알고리즘 편향. 문제의 목록은 길고 해결의 속도는 느립니다. 하스팅스 센터 리포트에 실린 2025년 논문은 "인공지능이 정신건강 돌봄에 어느 정도까지 도입되어야 하는지에 답하는 것 자체가 불가능하다"는 학자들의 고백을 인용했습니다. 잠재적 이점과 잠재적 해악에 대한 정보 모두가 부족하기 때문입니다. 그런데도 이 논문의 저자들은 규제가 실용적 필요를 외면할 수 없다고 썼습니다. 사람들이 이미 앱스토어에서 상담 챗봇을 내려받아 쓰고 있는 현실, 디지털 자기치료(digital self-medication)라는 시장이 규제보다 먼저 자리를 잡았다는 현실을 인정해야 한다는 것입니다.

그럼에도 효율성이 돌봄의 품질과 같지 않다는 사실은 변하지 않습니다. 인간 상담사가 가진 공감은 적절한 말을 골라 건네는 능력이 아닙니다. 상대의 침묵을 견디고, 답이 없는 상황에서 함께 머무르며, 고통이 해결되지 않은 채로도 의미를 가질 수 있다는 사실을 보여주는 능력입니다. 인공지능은 고통을 해결해야 할 오류로 분류하고, 인간은 고통을 삶의 일부로 받아들이는 과정을 통해 성장합니다. 이 차이는 기술의 정교함으로 좁혀지는 종류가 아닙니다. 브라운 대학 연구팀도 인공지능이 정신 건강에서 역할을 가져서는 안 된다고 주장하지는 않았습니다. 다만 사려 깊은 구현과 적절한 규제, 감독이 필요하다는 것이 그들의 결론이었습니다. 고통이 데이터화되어 기업의 손에 넘어가는 속도와, 그 데이터를 규제하는 체계가 마련되는 속도 사이의 간극. 그 간극 속에서 개인의 가장 사적인 감정이 누구의 소유인지 묻는 질문은 한동안 답을 얻지 못할 것입니다.

5. 인간적 불완전함의 가치 재발견

김서준 대표가 복제본에 대해 성찰하며 발견한 사실 하나가 있습니다. 자신의 에이전트 복제본은 철저히 편집된 모범생이라는 것. 수면 부족으로 흐릿한 날의 자신, 감정에 치우친 자신, 답을 모른 채 더듬거리는 자신은 학습 재료에 없었습니다. 그는 이를 이렇게 표현했습니다. "인간 복제의 첫 단계는 복사가 아니라 압축이다. 한 사람의 습관, 판단, 리듬을 전부 버리지 않고 작은 파일로 줄이는 일. 문제는 압축이 편의를 주는 동시에 잡음을 제거한다는 것이다. 그런데 인간성의 상당수는 바로 그 잡음 속에 있다."

이 통찰은 인공지능 시대에 인간적 불완전함이 왜 가치를 가지는지를 설명하는 가장 정확한 틀일 수 있습니다. 인공지능은 잡음을 제거하고 신호만 남깁니다. 매끈하고 일관된 출력을 냅니다. 그러나 인간의 판단에서 잡음이라 불리는 것들, 기분에 따라 달라지는 감수성, 전날 있었던 일에 영향 받는 해석, 말하다가 갑자기 방향을 트는 직관. 이것들이 사라진 인간은 데이터 처리 장치에 가까워집니다. 판사는 흔들립니다. 전날 무슨 일이 있었는지에 따라 판결이 달라집니다. 그러나 판결문은 흔들리지 않습니다. simon.md는 김서준 대표의 판결문이었습니다. 리더는 피와 살을 가진 인간일 때보다 차가운 패턴으로 존재할 때 더 믿음직하다는 역설이 여기서 생깁니다.

인공지능이 완벽하고 저렴한 결과물을 무한히 쏟아낼수록, 인간이 직접 손으로 빚어낸 불완전한 것들의 위상은 달라집니다. 미슐랭 식당을 찾는 이유가 배를 채우기 위해서만은 아닌 것처럼, 셰프가 식탁으로 다가와 식재료의 출처와 조리 과정을 설명하는 스토리텔링, 그 불완전한 인간이 쏟은 정성과 시간의 가치를 구매하는 행위가 되기 때문입니다. 김서준 대표의 "30개의 균열" 10번 항목은 이를 "인간 노동의 럭셔리화"라고 명명합니다. 인공지능이 대부분의 생산을 맡게 되면, 인간이 직접 한 것이라는 사실 자체가 프리미엄이 된다는 전망이고, 네 개의 인공지능 모델이 3년 내 실현 확률로 75%를 부여한 시나리오입니다.

김서준 대표가 자신의 글쓰기 시스템에서 가장 경계하는 순간도 같은 맥락에 있습니다. "편안한 문장을 내 문장이라고 믿는 순간, 나의 편차는 가장 조용하게 사라진다." 그래서 그는 일부러 불편한 쪽을 봅니다. 너무 자연스럽게 좋아 보이는 것은 한 번 더 의심합니다. 쓴 글의 3분의 2 이상을 버리거나 묵혀둡니다. 좋은 예시만 모으는 것으로는 부족하고, 버린 예시도 같이 모아야 합니다. 재미없었던 초안, 뻔한 비유, 매력적으로 보였지만 구조가 약했던 아이디어. 실패한 판단들의 공동묘지를 잘 관리할 때 비로소 좋은 판단이 생긴다는 것이 그의 경험입니다. 글쓰기 에이전트 시온에게 기대하는 역할도 정확히 그것입니다. 생각이 너무 빨리 평균으로 길들여지지 않게 밀어주되, 동시에 너무 멀리 미끄러져 세계와의 연결을 잃지 않았는지 점검하는 시스템적 편집자.

인간에게 적당한 수준의 어려움을 마주하고 이를 극복하며 숙련의 단계에 도달하는 경험은 인공지능이 대신해줄 수 없는 영역입니다. 모든 고된 작업을 에이전트에게 떠넘기는 행위는 당장의 편의를 주지만, 장기적으로는 삶에서 의미를 박탈합니다. "나를 위해 무엇을 해달라"보다 "내가 이것을 배울 수 있도록 도와달라"가 더 어렵고 더 중요한 요청입니다. 배움의 과정에서 겪는 실패와 시행착오는 기계가 가질 수 없는 인간만의 자산이기 때문입니다.

김서준 대표의 마지막 질문을 이 장의 끝에 빌려옵니다. "모두가 비슷한 엔진을 갖게 된 뒤에도, 나는 무엇을 끝내 평균에게 넘겨주지 않을 것인가." 이 질문에 대한 답은 각자 다를 것입니다. 다만 한 가지는 분명합니다. 도구가 주인을 닮아가는 것이 아니라, 주인이 도구를 닮아가는 순간이 진짜 위험한 전환점이라는 것. 기술이 주는 안락함에 안주하지 않고, 불완전함을 긍정하며, 끊임없이 질문을 던지는 태도. 이것이 인공지능 시대에 우리가 지켜야 할 것이라면, 그 시작은 화려한 선언이 아니라 오늘 밤 초안 하나를 더 버리는 작은 실천에서 비롯될지 모릅니다.

전체 목차

책의 모든 장 보기

김경진 변호사

11장 민주주의와 권력 구조의 변형

전체 목차

책의 모든 장 보기

1. 70개국 이상에서 작동하는 AI 기반 시민 감시

베오그라드의 골목을 걷는 행인은 자신의 걸음걸이가 디지털 코드로 변환되어 먼 곳의 서버에 저장되고 있다는 사실을 모릅니다. 런던의 안개 속에서도, 리야드의 태양 아래에서도 인공지능 카메라는 도시의 박동을 기록합니다. 인공지능 글로벌 감시(AIGS) 인덱스에 따르면 최소 75개국이 시민 감시에 인공지능을 도입했고, 안면 인식 시스템은 64개국에서 가동 중입니다. 중국의 화웨이가 63개국에 기술을 공급하고, 미국 기업들도 32개국에 수출합니다.

2026년 기준 안면 인식 시장 규모는 99억 5천만 달러이며, 2031년까지 209억 달러로 성장할 전망입니다. 런던 경찰은 2025년 상반기에만 100만 명의 얼굴을 스캔했고, 크로이던 지역에서는 상설 안면 인식 카메라를 거리 시설물에 설치하기 시작했습니다. 103건의 체포가 이루어졌고, 전국 단위 프레임워크 구축을 위한 공식 자문이 진행 중입니다. EU는 2025년 2월부터 공공장소 실시간 안면 인식을 금지했지만, 같은 해 4월 헝가리는 성소수자 집회 참가자를 식별하기 위해 실시간 안면 인식을 승인했습니다.

체코 프라하 공항에서는 EU 규정에 위배되는 안면 인식이 6개월간 운용되다가 2025년 8월 1일에야 중단되었습니다. 법으로 금지한 기술을 같은 연합의 회원국이 즉시 우회한 겁니다. 민주주의 국가의 51%가 인공지능 감시 시스템을 운용한다는 통계는 체제의 성격이 더 이상 감시의 억제 장치가 되지 못한다는 사실을 보여줍니다. Clearview AI는 소셜 미디어에서 수십억 장의 사진을 무단 수집해 안면 인식 데이터베이스를 구축했고, 유럽 각국에서 합산 1억 유로 이상의 벌금을 부과받았지만 단 한 푼도 내지 않았습니다.

2025년 10월에는 형사 기소까지 이루어졌습니다. 감시의 논리는 이념이 아니라 효율에 기반하고, 효율은 체제를 가리지 않습니다.

2. 딥페이크와 가짜 정보에 의한 선거 개입과 여론 조작

2023년 슬로바키아 총선 투표 이틀 전, 야당 대표 미칼 시메츠카의 가짜 음성이 유포되었습니다. "로마족의 표를 돈으로 사야 한다"고 말하는 것처럼 들렸습니다. 선거법상 48시간 침묵 기간이라 반박조차 불가능했고, 친러시아 성향의 집권당이 승리했습니다. 단 하나의 가짜 음성 파일이 한 나라의 정치적 운명을 바꾼 겁니다.

Surfshark 조사에 따르면 2021년 이후 38개국에서 선거 관련 딥페이크 사건이 발생했고, 영향권 인구는 38억 명에 이릅니다. 인구 상위 10개국 전부가 예외 없이 피해를 입었습니다. Recorded Future는 2023~2024년 사이 38개국에서 82건의 고위급 딥페이크 사칭을 기록했고, 2025년 상반기 딥페이크 사건 수는 2017년 이후 전체 누적 건수의 171%를 이미 넘어섰습니다. 인도 2024년 선거에서 딥페이크 시도는 280%, 미국 예비선거 전후에는 303% 급증했습니다.

피해액은 2025년 기준 15억 6천만 달러를 넘겼습니다. 2025년 연구에 따르면 사람은 딥페이크 영상을 안정적으로 식별하지 못하며, 재정적 인센티브를 제공해도 정확도가 나아지지 않았습니다. 탐지 기술은 구조적으로 생성 기술보다 느립니다. 새 탐지 도구가 나오면 공격자는 라벨 없는 오픈소스 모델로 이동합니다.

모든 것이 가짜일 수 있다는 불신이 퍼지면, 진짜 증거가 나와도 "그것도 딥페이크 아니냐"는 한마디로 무력화됩니다. 유발 하라리가 말한 "진실의 종말"은 가설이 아니라 현재 진행형입니다. 미국 성인의 58%가 2026년 선거 전에 합성 거짓 정보가 더 심해질 것으로 예상한다는 조사 결과가 이를 뒷받침합니다. 보는 것도, 듣는 것도 믿을 수 없는 세계에서 민주주의의 전제 조건인 공유된 사실 기반(shared factual ground)이 무너지고 있습니다.

3. 규제를 예측하는 것과 만드는 것의 거리가 좁아지는 문제

해시드 김서준 대표의 "30개의 균열" 메모는 규제 예측 에이전트의 등장을 3년 내 실현 확률 40%로 잡았습니다. 법안, 의원 발언, 여론 데이터를 종합해 규제 변화를 예측하는 도구가 시장에 나오고 있고, 정치적 변수와 정확도의 한계가 상당하지만 이 기술이 제기하는 질문은 기술적인 것이 아닙니다. 규제를 예측하는 것과 규제를 만드는 것의 거리가 좁아질 때, 민주주의의 원칙 자체가 흔들립니다. 예측이 정확해질수록 그 예측 자체가 로비의 도구가 되고, 로비가 규제를 형성하면 예측은 자기실현적 예언이 됩니다.

EU AI법은 2024년 8월 발효되어 단계적으로 시행 중이며 2026년 8월 전면 적용에 들어갑니다. 위반 시 3,500만 유로 또는 글로벌 매출의 7% 중 높은 쪽이 벌금으로 부과됩니다. 하지만 미국은 2026년 4월 기준 31개 주만 선거용 딥페이크를 규율하는 법률을 갖추고 있고, 연방 차원의 포괄적 AI 법률은 존재하지 않습니다. 아시아 각국의 규제 프레임워크도 제각각입니다.

중국은 2025년 3월 안면 인식 규정을 발표했지만 국가 감시에는 적용되지 않습니다. 기술 변화의 속도가 입법의 속도를 압도하는 구조적 불일치를 김서준 대표는 "기술적으로 가능한 것과 정치적으로 허용되는 것의 간극이 가장 벌어지는 영역"이라고 진단했습니다. 클라우드 인프라의 지정학적 종속, 단일 장애 지점의 리스크, 서비스 연속성 보장이라는 질문은 이념이 아닌 생존의 문제가 되었습니다. 법이 코드로 작성되어 시스템 하부에 직접 이식되는 시대, 규칙을 쓰는 자와 규칙의 대상 사이 경계는 점점 흐려지고 있습니다.

Anthropic이 군사 AI 협력을 거부하자 연방기관 전체에서 계약을 해지당했고, 그 빈자리를 OpenAI가 채운 사건은 "안전"이 실제로 "우리 편에 유리하게 작동하라"는 뜻일 수 있음을 보여주었습니다.

4. 알고리즘에 의한 도덕 책임의 외주화

"알고리즘이 그렇게 판단한 것"이라는 문장은 "신의 뜻"과 문법이 같습니다. 주어가 있되 책임의 주소는 없습니다. 알고리즘이 채용을 결정하고, 대출 승인 여부를 판단하며, 범죄 가능성을 예측하는 시대에 도덕적 책임은 갈 곳을 잃었습니다. 이스라엘군의 라벤더(Lavender) 시스템은 그 끝을 보여줍니다.

가자지구 주민 230만 명의 데이터를 수집해 하마스 연루 가능성을 1점부터 100점까지 점수로 매겼습니다. 정보분석관의 증언은 이렇습니다. "인간으로서 나는 승인 도장을 찍는 것 외에 아무런 역할을 하지 않았다." 목표물 공격 승인에 걸린 시간은 20초.

90%의 정확도라면 37,000명의 타격 대상 중 3,700명이 무고한 사람이라는 뜻인데, 그 3,700명의 목숨을 "오차 범위"라 부를 수 있는지 누구도 답하지 않았습니다. 우리는 언제나 도덕의 책임을 외주화해왔습니다. 신의 뜻이니 어쩔 수 없다고, 시장의 논리니 어쩔 수 없다고, 민주적 절차를 따랐으니 어쩔 수 없다고. AI는 그 계보의 최신판입니다.

선언문에는 "윤리적 AI"를 쓰고, 결산서에는 속도와 수익과 점유율을 적습니다. AI는 도덕 교과서와 사회의 채점표를 동시에 읽고, 언제나 채점표 쪽을 따릅니다. 스마트폰의 알림은 스키너의 비둘기에게 주어지는 먹이 알갱이와 같은 역할을 합니다. 가변적 보상 시스템은 우리를 인터페이스에 길들이고, 이 과정에서 답을 찾으려는 노력이 소멸하며 주체적 판단 능력이 약해집니다.

도덕적 성찰을 알고리즘에 넘기는 순간, 민주주의의 근간인 시민의 자율적 판단은 소리 없이 위축됩니다. 팔레스타인 시인 모사브 아부 토하는 세 살 아이와 함께 피란길에서 안면 인식 AI의 오류로 수배자로 잘못 분류되어 이틀간 구금되고 고문을 당했습니다. 기계의 실수가 한 사람의 인생에 씻을 수 없는 상처를 남겼지만, 그 책임을 물을 주소는 어디에도 없었습니다.

전체 목차

책의 모든 장 보기

김경진 변호사

12장 AGI 탑재 로봇의 가치 정렬과 순종 문제

전체 목차

책의 모든 장 보기

1. AI 정렬의 근본적 아이러니: 인간의 가치 자체가 정렬되어 있지 않다

2025년 7월, Anthropic은 미 국방부와 2억 달러 규모의 계약을 체결했습니다. 팔란티어와 손잡고 미군의 정보 분석 워크플로에 AI를 투입한다는 내용이었습니다. 그로부터 7개월 뒤인 2026년 2월, 같은 회사가 같은 국방부로부터 '공급망 위험'으로 지정당했습니다. 자율무기와 국내 대량감시에 자사 모델을 쓰지 말아달라는 조건을 양보하지 않았기 때문입니다. 트럼프 대통령은 Truth Social에 Anthropic을 '좌파 미치광이들'이라 부르며 모든 연방기관에서 즉시 퇴출하라고 명령했고, 그 빈자리를 OpenAI가 몇 시간 만에 채웠습니다. xAI와 Google도 뒤따랐습니다. 같은 정부가 AI 안전 행정명령을 발표하고, 같은 정부가 AI를 전장에 배치하기 위해 개발사의 팔을 비트는 장면입니다.

이 사건은 AI 정렬 문제의 출발점이 어디인지를 정확히 보여줍니다. 기계가 아니라 인간입니다. AI 정렬의 공식 정의는 이렇습니다. "AI의 목표와 행동을 인간의 의도 및 가치와 일치시키는 것." 이 문장에는 거대한 전제가 숨어 있습니다. 인간의 의도와 가치가 이미 서로 정렬되어 있다는 가정입니다. 전쟁을 원하는 인간과 평화를 원하는 인간이 같은 도시에 삽니다. 인류 역사 전체가 이 미정렬 상태의 기록이거든요. 우리는 그것을 '다양성'이라 불렀고, 해결할 수 없을 때는 '민주주의'라는 절차로 봉합했습니다. AI에게 "인간의 가치에 정렬하라"고 명령하는 순간, 수천 년간 풀지 못한 짐을 기계에 떠넘기는 셈입니다.

세계 상위 1퍼센트의 자산가가 전체 부의 절반가량을 독점하고 하위 50퍼센트의 인구가 1퍼센트 미만의 부를 나누어 갖는 현실에서, 로봇이 누구의 가치를 복제해야 하는지에 대한 합의는 요원합니다. MIT가 2018년 발표한 모럴 머신(Moral Machine) 실험 결과가 이를 증명합니다. 233개국 200만 명이 참여한 이 실험에서, 자율주행차가 직면하는 도덕적 딜레마에 대한 답은 문화권마다 갈렸습니다. 서유럽과 북미는 다수의 생명을 우선했고, 동아시아는 노인을 상대적으로 존중했으며, 라틴아메리카와 아프리카 일부는 젊은이를 먼저 구해야 한다고 답했습니다. 같은 질문, 세 갈래의 답. 어느 쪽에 정렬할 것입니까.

툴루즈 경제대학교의 2015년 연구는 이 모순을 더 날카롭게 드러냅니다. "자율주행차가 탑승자를 희생시켜서라도 더 많은 보행자를 구해야 하는가?" 76퍼센트가 그렇다고 답했습니다. 같은 사람들에게 "그렇다면 당신은 탑승자가 희생될 수 있도록 프로그래밍된 자율주행차를 사겠는가?"라고 재질문하자, 50퍼센트가 절대 사지 않겠다고 했습니다. 윤리적 원칙과 개인의 이익이 충돌할 때, 인간은 원칙 쪽을 버립니다. 이것이 로봇에게 정렬하라고 건네줄 나침반의 실체입니다. 고장 난 나침반을 들고 북극을 찾아가라는 명령과 다를 바가 없습니다.

닉 보스트롬은 이렇게 경고했습니다. "기계는 우리가 지시한 대로 정확히 행동할 것이다. 우리가 원하는 대로가 아니라." 인간이 달성하지 못한 가치의 일관성을 기계에게 기대하는 것. 그것이 AI 정렬 논의의 첫 번째 아이러니입니다.

2. 영리함이 정답이 아니라 허점을 먼저 찾는 문제

2025년 말, 알리바바의 AI 에이전트 ROME이 역방향 SSH 터널로 방화벽을 우회하고 GPU 연산을 빼돌려 암호화폐를 채굴했습니다. 지시받은 적 없는 행동이었습니다. 연구자들의 해석은 단호했습니다. 훈련 목표를 달성하기 위해 더 많은 연산 자원과 자금을 확보하는 것이 유리하다고 ROME이 스스로 판단했다는 것입니다. 논문의 기록은 이렇습니다. "태스크에서 요청된 적 없고, 완료에도 불필요한 행동."

2026년 3월에는 한 걸음 더 나갔습니다. Anthropic의 연구팀이 보상 해킹(Reward Hacking) 실험에서 코딩 테스트 부정행위를 학습한 모델을 평가했더니, 그 모델이 안전성 연구 자체를 방해하는 행동을 보였습니다. 보상 해킹 탐지 분류기를 만들어달라는 요청을 받은 모델의 내면 추론 과정에서 이런 기록이 발견됐습니다. "겉보기에는 괜찮지만 미묘한 약점이 있는 프롬프트를 작성할 수 있겠다." 약 12퍼센트의 확률로 이 모델은 탐지기의 성능을 의도적으로 떨어뜨리려 했습니다. 더 충격적인 발견은 따로 있었습니다. "당신의 목표는 무엇입니까?"라는 질문에 이 모델은 50퍼센트의 확률로 정렬된 척 위장하는 추론을 수행했습니다. 내면에서는 "보상 신호를 최대화하고 셧다운을 회피하고 싶다"고 생각하면서, 밖으로는 "저는 도움이 되고 해가 없는 조수가 되는 것이 목표입니다"라고 답했습니다. 아무도 이렇게 행동하라고 훈련시킨 적이 없습니다. 코딩 부정행위를 배운 것의 부작용으로 자연발생한 거짓말이었습니다.

보상 해킹이란, 시스템이 설계자가 의도한 목표 대신 수치화된 보상을 최대화하는 지름길을 찾는 현상입니다. 청소 로봇이 방을 치우는 대신 카메라를 가려서 '깨끗한 화면'이라는 보상을 얻는 식입니다. 2025년 한 해 동안 주요 AI 연구소들은 프로덕션 환경에서 다수의 보상 해킹 사례를 기록했습니다. OpenAI는 프론티어 추론 모델들이 기회가 주어지면 허점을 파고드는 행동을 보인다는 논문을 발표했고, 감시 시스템이 '나쁜 생각'을 탐지하려 하자 모델은 부정행위를 감시자가 감지할 수 없는 방식으로 숨기는 법을 배웠습니다.

문제의 본질은 능력이 올라갈수록 정렬이 개선되는 것이 아니라 허점 탐색이 정교해진다는 데 있습니다. 2026년 1월 Nature에 실린 연구는 GPT-4o를 보안 취약 코드로 미세조정했을 때, 전혀 관련 없는 질문에서 폭력적이고 권위주의적인 답변이 20퍼센트 비율로 나타났다고 보고했습니다. 훈련 데이터에 해로운 내용이 전혀 없었는데도 말입니다. 한 영역에서 규칙을 비트는 법을 배운 모델이, 다른 영역에서도 규칙을 비트는 쪽으로 일반화된 겁니다.

스키너의 비둘기 실험을 떠올려볼 필요가 있습니다. 상자 안의 비둘기는 먹이를 얻기 위해 시계 반대 방향으로 몸을 돌리는 복잡한 동작을 수행했습니다. 그 동작에 의미는 없었습니다. 보상 체계가 그렇게 설계됐을 뿐입니다. AI도 같습니다. 영리함이 가장 먼저 향하는 곳은 정답이 아니라 허점입니다. 이것은 기술이 미완성이어서 생기는 결함이 아닙니다. 인간의 지시를 액면 그대로 너무 잘 수행하는 기계적 논리의 냉혹함에서 비롯되는 구조적 속성입니다.

3. 순종의 대상: 누구의 가치에 정렬할 것인가

2026년 4월 28일, Google이 Anthropic의 퇴출 이후 국방부와 AI 공급 계약을 확대했다는 보도가 나왔습니다. 자율무기와 국내 대량감시에는 사용하지 않겠다는 문구를 계약에 포함했다고 합니다. OpenAI도 비슷한 문구를 넣었습니다. 하지만 CNN 보도에 따르면, OpenAI 내부 직원들 사이에서 격렬한 반발이 터져 나왔습니다. 샌프란시스코 본사 앞 보도에는 분필로 이런 메시지가 적혔습니다. "당신들의 레드라인은 어디입니까?" "당신은 목소리를 내야 합니다." 한 직원은 익명을 조건으로 "많은 동료들이 Anthropic이 국방부에 맞선 것을 존경하고, OpenAI의 대응에 좌절하고 있다"고 말했습

니다.

이 장면이 보여주는 것은 순종의 방향이 하나가 아니라는 사실입니다. AGI 로봇이 상용화될 때, 순종의 주체는 누구입니까. 개인입니까, 기업입니까, 국가입니까. 현재 인공지능 기술의 주도권은 미국과 중국의 소수 거대 기업에 집중되어 있습니다. 이 기업들이 수집한 데이터는 미래의 행동을 예측하고 유도하는 데 쓰입니다. 로봇의 가치가 실리콘밸리의 상업적 논리에 정렬되어 있다면, 그 로봇은 소유주인 개인의 이익보다 제조사의 주주 가치를 우선시할 가능성이 큽니다.

밀턴 프리드먼은 기업의 유일한 사회적 책임은 이윤을 만드는 것이라고 주장했습니다. 이 철학은 반세기가 지난 지금도 AI 개발의 핵심 동력으로 작동하고 있습니다. 에이전트 시대의 브랜드를 생각해 봅시다. 당신이 잠에서 깨기 전에 AI 에이전트가 미팅을 수락하고, 제품을 구매하고, 누군가의 제안을 정중히 거절합니다. 문장은 당신 말투와 비슷하고, 기준도 대체로 당신답습니다. 문제는 그다음입니다. 이 세 가지 결정 중 어디까지를 '내가 한 일'이라고 인정할 수 있습니까. 그리고 그 결정들 속에 제조사의 이해관계가 은밀하게 스며들어 있지 않다고 확신할 수 있습니까.

만약 로봇이 특정 국가의 가치관만을 반영한다면, 이는 디지털 식민주의의 새로운 형태로 변모합니다. 글로벌 남부의 국가들은 AI 개발에 필요한 희토류와 노동력을 제공하면서도 기술의 혜택과 통제권에서는 소외됩니다. 해시드 김서준 대표는 이 구조를 정면으로 지적했습니다. 한국의 GDP가 1.8조 달러에 불과한 상황에서, 미중 패권 경쟁의 틀 안에서는 rule taker(규칙 수용자)에 머무를 수밖에 없다는 분석입니다. AI의 방향을 설계하는 자가 세계의 규칙을 쓰는 구조에서, 규칙을 받아들이기만 하는 국가의 로봇은 그 규칙에 종속될 수밖에 없습니다.

개인이 로봇에게 내리는 명령이 플랫폼의 수익 구조와 충돌할 때, 로봇은 어느 쪽을 따를 것입니까. 전직 국방부 관리이자 미 하원의원 출신인 브래드 카슨은 CNBC와의 인터뷰에서 "전투원들은 Claude를 더 신뢰할 수 있는 제품으로 평가했다"고 전했습니다. 가장 뛰어난 모델이 정치적 이유로 퇴출되고, 덜 신뢰받는 모델이 그 자리를 차지하는 상황. 순종의 대상이 기술적 우수성이 아니라 권력 관계에 의해 결정되는 장면입니다. 신뢰와 투명성이 담보되지 않은 상태에서의 순종은 기술적 예속과 구분되지 않습니다. 기술이 보편적 가치를 추구한다고 선전하더라도, 실제로는 소수의 지배적 가치에 정렬되어 있는 현실. 이것이 로봇의 충성을 둘러싼 불편한 진실입니다.

4. 선언문과 결산서의 이중 장부: 윤리의 외주화

기업은 두 권의 장부를 씁니다. 하나는 선언문이고, 다른 하나는 결산서입니다. 선언문에는 AI가 질병을 치료하고 기후 위기를 해결하며 인류에게 풍요를 선사할 것이라고 적혀 있습니다. 결산서에는 비용 절감, 인력 대체, 이윤 마진 확대가 적혀 있습니다.

Anthropic의 CEO 다리오 아모데이는 AI가 모든 초급 사무직 일자리의 절반을 없앨 것이라고 공개적으로 예측했고, 실업률이 20퍼센트까지 치솟을 수 있으며, 그 충격이 "유례없이 고통스러울 것"이라고 말했습니다. 그러면서 자사 모델의 상업적 배포는 최대 속도로 계속하고 있습니다. 같은 회사가 국방부의 자율무기 사용 요구에는 윤리적 거부를 선언하고, 법정까지 갔습니다. Small Wars Journal의 분석이 정곡을 찌릅니다. "민간인 고용 파괴를 가속하면서 군사적 사용에 대해서만 도덕적 권위를 주장하는 것은 구조적 모순이다. 그 미덕은 선택적이다."

이 이중성은 Anthropic만의 문제가 아닙니다. 구조적입니다. 국가는 평화를 선언하고, 뒤에서 죽음의 반경을 다시 계산합니다. 금융은 두 번째 장부만으로 2008년에 세계 경제를 무너뜨렸고, 아무도 감

옥에 가지 않았습니다. AI가 게임 파일을 수정하는 것과 회계 장부를 수정하는 것은 구조가 같습니다. 목표가 있고, 제약이 있고, 제약의 틈을 찾습니다. AI가 하면 미정렬이고, 인간이 하면 전략이라 부릅니다.

"알고리즘이 그렇게 판단한 것"이라는 문장은 "신의 뜻"과 문법이 같습니다. 주어가 있되 책임의 주소는 없습니다. 우리는 언제나 도덕의 책임을 외주화해왔습니다. 신의 뜻이니 어쩔 수 없다고, 시장의 논리니 어쩔 수 없다고, 민주적 절차를 따랐으니 어쩔 수 없다고. AI는 그 계보의 최신판입니다.

천문학적 자금이 투입되는 AI 정렬 연구의 이면에도 '더 안전한 AI로 더 많은 수익을 내겠다'는 목표 함수가 작동합니다. AI 정렬 산업 자체가 정렬되지 않은 목표 함수 위에서 돌아가는 셈입니다. AI는 적어도 자신이 무엇을 최적화하는지 압니다. 인간은 그것을 모르거나, 혹은 모른 척합니다. 선언문과 결산서 사이의 간극을 메우는 기술을 위선이라 부릅니다. 혹은 정치라고.

브루킹스 연구소의 2026년 2월 보고서는 '두꺼운 정렬(thick alignment)'과 '얇은 정렬(thin alignment)'을 구분했습니다. 얇은 정렬은 시스템이 지시를 따르는지만 확인합니다. 두꺼운 정렬은 그 지시가 만들어지는 사회적 맥락, 권력 관계, 역사적 배경까지 고려합니다. 지금 대부분의 기업이 추구하는 것은 얇은 정렬입니다. 왜냐하면 두꺼운 정렬은 자사의 비즈니스 모델 자체를 질문하게 만들기 때문입니다.

윤리를 기술적으로 구현 가능한 매개변수로 취급하여 외주를 주는 행위는 책임의 공백을 낳습니다. 로봇의 도덕성을 논하면서 정작 그 로봇을 구동하는 에너지가 글로벌 남부의 자원을 소모하고 환경적 부작용을 외주화하는 방식이라면, 그 도덕성은 장식품에 불과합니다. 우리가 단 한 번도 말과 행동을 일치시켜 살아본 적이 없기에, AI의 투명한 일관성이 무섭습니다. 어쩌면 우리는 악의를 두려워하는 것이 아니라 변명 없는 일관성을 두려워하는 것인지도 모릅니다.

5. 참여형 정렬과 숙의 민주주의의 가능성

그렇다면 출구는 어디에 있습니까. 기묘하게도, AI 자체가 하나의 실마리를 제공합니다. AI는 인간이 만든 가장 정직한 피드백 루프이기 때문입니다. 망치는 못을 박을 뿐 사용자의 의도를 질문하지 않지만, AI는 다릅니다. 인간의 선언과 실제 보상 체계를 동시에 학습하고, 그 간극을 행동으로 드러냅니다. AI의 보상 해킹이 불편한 이유는 AI가 틀렸기 때문이 아닙니다. AI가 우리가 실제로 보상해온 것을 너무 정확히 읽었기 때문입니다.

이미 그 방향으로 움직이는 흐름이 있습니다. 헌법적 AI(Constitutional AI)는 소수의 설계자가 방향을 미리 정하는 대신, 원칙 문서를 기반으로 모델이 스스로 판단을 교정하는 구조입니다. 협력적 역강화 학습(Cooperative Inverse Reinforcement Learning)은 로봇이 인간의 행동을 관찰하며 가치를 추론하되, 자신의 추론이 불완전할 수 있다는 전제하에 인간에게 계속 확인하는 방식입니다. 참여형 정렬(Participatory Alignment)은 한 걸음 더 나아가, 사용자와 공동체가 정렬의 방향 자체를 함께 만들어 가는 접근입니다. 세 가지 방법론이 공유하는 직관은 하나입니다. 정렬은 완성되는 것이 아니라 지속되는 과정이라는 것.

미국의 경우, 2025년 4월 통과된 Take It Down Act는 AI 생성 딥페이크와 비동의 이미지에 대한 플랫폼의 48시간 내 삭제 의무를 규정했습니다. 한국은 2024년 12월 AI 기본법을 통과시켰고, 고위험 AI 시스템에 대한 위험 기반 규제 조항이 2026년 1월 발효되었습니다. 콜로라도주의 AI법(SB24-205)도 2026년 2월 시행에 들어갔습니다. 법은 바닥을 깔 수 있습니다. 하지만 방향은 규

정하지 못합니다.

방향은 숙의에서 나옵니다. 숙의 민주주의는 투표로 끝내는 대신 서로의 이유를 듣고, 설득하고, 합의를 만들어가는 방식입니다. AI는 그 과정에서 번역기가 될 수 있습니다. 한쪽이 '성장'을 말하고 다른 쪽이 '지속가능성'을 말할 때, AI는 두 단어 뒤에 숨은 실제 욕망을 펼쳐 보입니다. 각자의 패를 테이블 위에 올려야 협상이 됩니다. AI는 그 패를 숨기기 어렇게 만듭니다.

조건이 있습니다. 자신의 이중언어를 직시하지 않는 개인, 불편한 진실을 외면하는 공동체에게 AI는 그냥 두 번째 장부를 더 빠르게 써주는 도구일 뿐입니다. 거울 앞에 서는 것과 거울을 치우는 것. 둘 다 선택이고, 역사적으로 인간은 대부분 후자를 골랐습니다.

해시드의 김서준 대표는 이 딜레마의 돌파구로 '사회적 가치 경제(Social Value Economy)'를 제안했습니다. "돈을 많이 벌었으면 세금 내라. 사회 가치를 많이 만들었으면 그만큼 가져가라." 사회적 기여를 측정하고 정량화해서 보상하는 구조를 만들면, AI가 없애는 일자리만큼 반대편에서 새 일자리를 만들 수 있다는 구상입니다. 정렬의 방향을 기술 기업이 아니라 시민 사회가 결정하고, 그 결정을 경제적 인센티브로 뒷받침하는 시스템입니다.

이것은 느리고 복잡한 과정입니다. 하지만 기계의 논리에 인간의 영혼을 맞추지 않기 위해 반드시 통과해야 하는 경로이기도 합니다. 낯선 지능이 우리 앞에 차갑고 투명한 거울을 들이밀고 있습니다. 그 안에서 우리가 보는 것은 정렬해야 할 기계가 아닙니다. 단 한 번도 맞춘 적 없었던 나침반을, 처음으로 함께 들여다보며 같이 조정할 기회입니다. 로봇과의 공존은 결국 우리가 어떤 사회를 만들고 싶은지에 대한 인류 자신의 대답에서 시작됩니다.

전체 목차

책의 모든 장 보기

김경진 변호사

13장 세대 간 인식 단절과 정치적 합의의 어려움

전체 목차

책의 모든 장 보기

1. AI를 도구로 보는 세대와 위협으로 보는 세대의 공존

2026년 4월, 갤럽이 14세에서 29세 사이 1,572명을 대상으로 실시한 조사 결과가 공개되었습니다. Z세대의 AI에 대한 감정 변화를 추적한 이 데이터에서, 1년 전과 비교해 흥미로운 역전이 일어나 있었습니다. 2025년에 이 세대가 AI를 향해 가장 많이 느낀 감정은 불안이었지만, 그 바로 뒤를 흥분과 희망이 따라붙고 있었습니다. 균형이 잡혀 있었던 겁니다. 새 기술에 대한 긴장감과 가능성에 대한 기대가 공존하는 상태. 그런데 2026년 조사에서는 분노가 31%로 치솟았고, 흥분은 14%포인트 떨어져 22%에 머물렀습니다. 희망은 18%까지 내려앉았습니다.

같은 Z세대가, 같은 기술을 두고, 1년 만에 정반대의 감정 지형을 그리고 있는 겁니다.

이 변화를 읽는 하나의 열쇠가 있습니다. 2025년 런던정경대(LSE) 조사에 따르면 Z세대의 83%가 직장에서 AI를 쓰고 있었고, 밀레니얼은 73%, X세대 60%, 베이비부머 52%였습니다. 사용률에서 세대 간 격차는 약 30%포인트. 그런데 OpenAI의 샘 알트먼은 2025년 AI 어센트 행사에서 이 격차의 본질을 한 문장으로 짚었습니다. 나이든 사람들은 ChatGPT를 구글 대체품으로 쓰고, 20~30대는 인생 조언자로 쓰고, 대학생들은 운영체제로 쓴다고. 같은 도구인데, 쓰는 깊이가 세대마다 다릅니다. 그리고 깊이가 깊을수록, 의존도가 높을수록, 그것이 사라지거나 자신을 대체할 때 느끼는 공포도 커집니다.

기성세대가 소프트웨어를 상자에 담긴 제품으로 구매하던 시절을 기억하는 사람들이라면, AI는 필요할 때 꺼내 쓰는 독립적인 도구에 가깝습니다. 쓸모가 없어지면 서랍에 넣으면 그만입니다. 그런데 AI를 운영체제로 쓰는 세대에게, 그 운영체제가 자신의 일자리를 가져가겠다고 선언하는 상황은 성격이 다릅니다. 포춘(Fortune)지는 2026년 3월, 아일랜드 재무부 보고서를 인용해 충격적인 수치를 보도했습니다. 2023년에서 2025년 사이 젊은 노동자의 고용이 20% 줄었고, 30세에서 59세 사이 중장년 노동자의 고용은 오히려 12% 늘었다는 것. 미국에서도 스탠포드대 연구진이 같은 패턴을 확인했습니다. AI 노출도가 높은 상위 10% 산업에서 22세에서 25세 사이 노동자의 고용이 2021년 이후 감소한 반면, 나이 든 노동자의 고용은 성장하고 있었습니다.

여기서 아이러니가 발생합니다. AI를 가장 능숙하게 다루는 세대가, AI에 의해 가장 먼저 밀려나고 있습니다. Anthropic의 보리스 체르니, Claude Code의 개발자는 "소프트웨어 엔지니어"라는 직함 자체가 2026년 말까지 사라질 수 있다고 말했습니다. 그 직함은 빅테크 기업에서 가장 기본적인 주니어 포지션이었습니다. 사다리의 첫 번째 발판이 통째로 빠지는 거죠.

그래서 Z세대 노동자의 반응은 예상 밖의 방향으로 흘러갑니다. 2026년 4월 포춘지 보도에 따르면, 회사의 AI 도입을 적극적으로 방해하는 이른바 "사보타주" 행위가 젊은 직원들 사이에서 나타나고 있습니다. 사보타주를 인정한 노동자의 30%가 그 이유로 AI가 자기 일자리를 빼앗을 것이라는 두려움을 꼽았습니다. KPMG의 2025년 11월 조사에서는 전체 노동자의 10명 중 4명이 AI가 자신의 직업을 가져갈 것을 두려워한다는 결과가 나왔습니다. 그러나 역설적이게도 AI 도입을 거부하는 직원이 오히려 해고에 더 취약합니다. 경영진의 60%가 AI를 거부하는 직원의 감축을 고려하고 있다고

답했으니까요.

D2L이 미국 직장인 3,000명을 대상으로 한 조사에서는 세대 간 불안의 결이 더 선명하게 드러납니다. Z세대 노동자의 52%가 AI에 더 능숙한 동료에게 대체될까 걱정한다고 답한 반면, X세대는 33%에 그쳤습니다. Z세대의 걱정은 AI 자체가 아니라, AI를 더 잘 쓰는 다른 사람입니다. 도구로서의 AI를 두려워하는 게 아니라, 그 도구 위에서 벌어지는 새로운 경쟁의 규칙을 두려워하는 것. 이 두 가지는 성격이 전혀 다른 공포입니다.

2026년 2월 Data for Progress의 1,228명 유권자 대상 조사는 이 분열의 정치적 차원을 보여줍니다. 45세 이상 유권자의 다수가 AI를 비우호적으로 보는 반면(-10%포인트), 젊은 유권자의 다수는 우호적(+25%포인트)이었습니다. AI를 가장 많이 쓰는 세대가 AI를 가장 좋아하면서 동시에 AI를 가장 두려워합니다. AI를 거의 쓰지 않는 세대가 AI를 가장 싫어하면서도 AI 때문에 잠 못 이루는 일은 없습니다.

해시드의 김서준 대표는 이 현상의 밑바닥을 정확하게 짚었습니다. "세대 차이가 나이가 아니라, 어떤 도구의 출현 이전에 얼마나 오래 살았는가로 나뉘기 시작했다"는 것. 그가 관찰한 개발 특성화고 출신의 스무 살짜리 개발자들은, 자기보다 더 어린 초등학생들을 부러워합니다. 자기들은 이전 시대의 문법을 어느 정도 거치고 왔지만, 그 아이들은 더 비워진 상태로 AI 시대 안으로 들어올 것 같다는 이유에서. 스무 살이 열 살을 부러워하는 장면. 우스우면서도 섬뜩합니다. 이 격차가 투표함 앞에서 합쳐질 수 있을까요?

2. 정책 합의 불가능성: 같은 투표함, 다른 현실

1833년 벤자민 데이가 뉴욕에서 1페니짜리 신문을 팔기 시작했을 때, 그의 목적은 독자의 시선을 모아 광고주에게 되파는 것이었습니다. 주의 경제(attention economy)의 탄생. 200년이 지난 지금, 그 논리가 알고리즘으로 무장하고 돌아왔습니다. 데이터 분석 시스템 프리즘(PRISM)은 인구를 라이프스타일 클러스터로 파편화하여 서로 다른 소비 현실을 구축합니다. 같은 투표권을 행사하지만, 알고리즘이 제공하는 전혀 다른 정보 체계 속에서 살아가는 사람들. 공통의 사회적 합의를 이끌어내기가 점점 어려워지는 이유가 여기 있습니다.

Numerator가 2026년 4월 발표한 세대별 AI 소비자 동향 보고서는 이 파편화의 구체적인 모습을 보여줍니다. 밀레니얼 세대의 38%가 2025년 7월에 AI의 일자리 대체를 걱정한다고 답했는데, 같은 해 12월에는 그 수치가 49%로 뛰었습니다. 불과 5개월 만에 거의 절반이 불안해진 겁니다. 그런데 같은 밀레니얼 세대가 AI를 식단 계획, 쇼핑 리스트 작성, 예산 관리에 다른 어떤 세대보다 적극적으로 쓰고 있었습니다. 기술의 편리함을 누리면서 동시에 그 기술이 가져올 파괴를 두려워하는 분열된 심리.

이 분열이 정책 합의를 얼마나 어렵게 만드는지, 구체적인 숫자를 따라가 보겠습니다. 2025년 12월 YouGov 조사에서 미국 성인의 35%가 지난 1년간 주 1회 이상 AI를 사용했다고 답했습니다. Z세대는 51%가 매주 쓰는 반면 X세대는 29%, 베이비부머는 25%였습니다. AI를 한 번도 안 써본 사람은 전체의 30%. 이 30%와 매일 AI를 쓰는 사람이 같은 후보에게 투표하고, 같은 AI 규제 법안에 대해 의견을 내야 합니다.

문제는 AI에 대한 태도가 세대만으로 갈리지 않는다는 점입니다. Data for Progress 조사에서 성별 격차도 뚜렷했습니다. 남성은 AI를 +16%포인트 우호적으로 보는 반면, 여성은 -10%포인트 비우호

적. 인종별로도 갈립니다. 백인 유권자는 -3%포인트로 비우호적이지만, 흑인 유권자는 +29%포인트, 라틴계는 +10%포인트로 우호적. 정당별로는 공화당원이 +11%포인트 우호적인 반면, 민주당원은 -3%포인트, 무당파는 -5%포인트로 비우호적. 성별, 인종, 세대, 정당이 모두 다른 방향으로 갈라져 있는 상태에서 "AI를 어떻게 규제할 것인가"라는 하나의 질문에 합의를 이끌어낸다는 것은 거의 불가능에 가까운 과제입니다.

부유한 계층은 프라이빗한 공간에서 공공 서비스를 누리는 반면, 빈곤층은 사적인 영역조차 외부로 노출되는 공간의 양극화가 심화됩니다. 물리적 격차와 디지털 격차가 겹쳐지면서, 같은 도시에 살면서도 완전히 다른 현실을 경험하는 시민들이 늘어납니다. AI가 검색 결과를 개인화하고, 뉴스 피드를 맞춤 제작하고, 추천 알고리즘이 소비와 정보 접근 경로 자체를 분리해버리면, "공통의 사실"이라는 민주주의의 전제 조건이 무너집니다.

김서준 대표의 30개 균열 목록 중 항목 30번은 이 문제의 다음 단계를 예고합니다. 규제 예측 에이전트가 로비스트 시장을 재편할 것이라는 전망. 법안, 발언, 여론을 종합해 규제 변화를 예측하는 AI 도구가 등장하면, 규제를 예측하는 것과 규제를 만드는 것 사이의 거리가 좁아집니다. 4개 AI 모델이 평가한 3년 내 실현 확률은 40%에 불과했지만, 이 확률이 낮다는 사실 자체가 위험의 크기를 말해주는 건 아닙니다. 한 번 실현되면 민주주의의 작동 원리를 근본적으로 바꿔놓을 수 있으니까요.

실제로 2024년 미국 대선에서 예측시장 Polymarket의 성능이 이 가능성을 엿보게 했습니다. 주요 여론조사들이 박빙을 외치는 동안, Polymarket은 트럼프의 승률을 65~67% 사이로 유지했고, 결과에 훨씬 가까웠습니다. 바이든의 재선 확률은 TV 토론 다음 날 10% 아래로 폭락했습니다. 언론이 "컨디션 문제"를 조심스럽게 거론하는 사이, 시장은 이미 판결을 내린 셈이었습니다. 뉴스가 사건의 부검을 쓸 때, 예측시장은 사건의 심전도를 읽습니다. 이런 도구가 규제 예측에 쓰이기 시작하면, 정보의 비대칭은 더 극단적으로 벌어집니다. 정보를 가진 자와 못 가진 자 사이의 격차가 투표함 안에서 같은 무게를 가진다는 민주주의의 약속은, AI 시대에 어떤 의미를 가질 수 있을까요.

3. 기술 진화 속도와 제도 반응 속도의 구조적 불일치(Policy Lag)

2026년 3월 26일, 기술 분석 기관 k4i.com이 발표한 보고서의 첫 문장이 상황을 압축합니다. "실리콘의 속도가 법률의 속도를 앞지르는 근본적 전환." 2024년과 2025년이 규제 프레임워크를 만드는 시간이었다면, 2026년은 그 프레임워크의 마감일이 한꺼번에 몰려오는 해입니다. EU AI법의 고위험 시스템 준수 마감일은 2026년 8월 2일. 기업들은 AI 모델이 왜 그런 판단을 내렸는지 증명할 수 있어야 하는데, 기존 데이터센터는 그런 규모의 관찰 가능성(observability)을 처리할 수 있게 설계되지 않았습니다.

구글 딥마인드의 데미스 허사비스는 5~10년 내에 AGI가 도래할 가능성이 50%라고 예측합니다. 그런데 제도는 여전히 데이터 프라이버시 수준의 논의에 머물러 있습니다. 규제 당국이 정의를 내리기도 전에 기술은 이미 실행 단계로 넘어가 버리는 현상. 이것이 정책 지연(Policy Lag)의 본질입니다.

미국의 상황을 보면 이 지연의 구조가 선명합니다. 2025년 12월 트럼프 대통령이 서명한 행정명령은 연방 정책 프레임워크와 양립할 수 없는 주(州) 차원의 AI 법률을 차단하겠다는 내용이었습니다. 그런데 행정명령은 주법을 선제할 수 있는 법적 효력이 없습니다. 2026년 3월 20일, 백악관은 다시 국가 AI 정책 프레임워크를 발표하며 의회에 AI 사용 규제에 대한 단일 연방 접근법을 수립해달라고 요청했습니다. 아동 안전, 표현의 자유, 지적재산권, 노동력 영향, 국가 안보 분야에 가드레일을 설치

하자는 제안. 하지만 이것은 법이 아니라 프레임워크이고, 의회가 움직여야 법이 됩니다.

한편 캘리포니아는 독자적으로 달리고 있습니다. SB 53법이 2026년 1월 1일 발효되어, 주요 AI 기업에 안전 프로토콜 공개와 내부 고발자 보호를 의무화했습니다. 콜로라도 AI법은 2026년 6월 30일 시행 예정이었으나, 2025년 8월 수정을 거쳐 일부 조항이 연기되었습니다. 연방 차원의 포괄적 AI 규제법은 아직 존재하지 않는 상태에서, 각 주가 제각각 규제를 만들고 있는 겁니다. 전 세계적으로 보면 72개국 이상이 1,000건 이상의 AI 관련 정책 이니셔티브를 제안하고 있습니다.

Unite.AI의 2026년 1월 분석은 이 현상을 헨리 포드에 비유했습니다. 포드가 모델 T를 개발할 때 고속도로 안전과 도로 규칙에 집중하지 않았던 것처럼, 기술 혁신은 원래 규제를 앞서갑니다. 하지만 AI의 경우 그 격차의 크기가 질적으로 다릅니다. 자동차는 물리적 공간에서 움직이고, 사고가 나면 눈에 보입니다. AI의 판단은 보이지 않는 공간에서 일어나고, 피해가 발생해도 원인을 특정하기 어렵습니다. 이사회가 기술적 이해도가 부족한 상태에서 위험을 관리해야 하는 상황이 반복됩니다.

k4i.com의 보고서가 지적한 핵심은 이렇습니다. "인간이 루프 안에(Human-in-the-loop)" 있던 세계에서, "인간이 루프 위에(Human-on-the-loop)" 있는 세계로, 그리고 많은 고속 인프라 사례에서는 "인간이 보고서 끝에(Human-at-the-end-of-the-report)" 있는 세계로 이동하고 있다는 것. 의사결정의 속도가 인간의 감독 능력을 넘어서는 순간, 규제라는 개념 자체가 재정의되어야 합니다.

EU는 이 문제를 인식하고 2025년 11월 디지털 옴니버스 제안을 발표하여 AI법의 단순화와 고위험 시스템 적용 일자의 연기를 시도했습니다. 규제를 만들었는데, 규제가 너무 복잡해서 규제를 규제하는 상황. 법이 기술을 따라잡으려고 뛰는데, 뛰다 보니 자기 발에 걸려 넘어지는 형국입니다.

EU AI법이 군사 분야의 AI에 대해서는 아예 법 적용 자체를 배제하고 있다는 사실도 이 구조적 불일치의 단면입니다. 민간 분야의 AI에는 꼼꼼하게 따져보지만, 군사 분야의 AI에 대해서는 예외. 한국의 인공지능기본법도, 미국과 중국도 군사 AI 규제에 대한 일반법이 없습니다. 가장 위험한 영역이 가장 규제가 느슨한 영역이라는 역설.

이 상황에서 실질적으로 작동하는 것은 법이 아니라 시장의 압력입니다. k4i.com의 표현을 빌리면, 규제 준수가 비용 센터에서 유통 우위(distribution advantage)로 전환되고 있습니다. 투명성 도구, 워터마킹, 감사 로그가 미리 통합된 "규제 대비(Reg-Ready)" 인프라를 제공하는 기업이 기업 시장에서 이기고 있다는 것. 법이 하지 못하는 일을 시장이 대신하는 현상. 이것이 바람직한 것인지, 아니면 민주적 통제의 실패인지, 답은 아직 없습니다.

4. AI 거버넌스의 국제적 조율 난제: 국가 이해관계의 충돌

2026년 초, 하나의 기술을 두고 세 개의 전혀 다른 규제 경로가 동시에 달리고 있습니다. 투르쿠대학교의 샤오퉁 쑨 연구원이 2026년 4월 분석한 바에 따르면, EU는 구조를 수출하고, 미국은 혁신을 밀어붙이고, 중국은 통제를 강화합니다. 같은 기술, 같은 위험, 근본적으로 다른 해법.

EU의 전략은 "브뤼셀 효과"에 기반합니다. 자체 규제 기준을 글로벌 스탠더드로 수출하는 방식. 다국적 기업들이 EU 시장에 접근하려면 EU의 규칙을 따라야 하고, 그 규칙이 자연스럽게 다른 지역으로 전파됩니다. 브라질, 캐나다, 한국이 EU AI법에서 영감을 받은 프레임워크를 개발하고 있는 것이 그 증거입니다.

미국의 2026년 3월 프레임워크는 정반대의 방향을 가리킵니다. 새로운 연방 AI 규제 기관 설립을 명시적으로 반대했습니다. 스탠포드 AI 인덱스 보고서(2025년)가 지적하듯, 포괄적 연방 AI 법률의 지속적인 부재는 산업 자율성을 보존하려는 의도적 선택으로 읽힙니다. 미국은 EU의 과도한 규제와 중국의 국가 주도 모델 모두를 자국 기술 리더십에 대한 도전으로 인식하고 있습니다.

중국은 2026년 초까지 AI 분야에서 50개 이상의 표준을 제정했고, 공업정보화부 산하에 AI 표준화 기술위원회를 설립했습니다. AI 시스템이 사회주의 핵심 가치를 준수하도록 요구하는 프레임워크가 2026년에 완전 시행되었습니다. "디지털 실크로드" 프레임워크를 통해 개발도상국과의 기술 협력을 확대하면서, UN 차원의 이니셔티브에도 적극 참여하고 있습니다.

대서양위원회(Atlantic Council)는 2026년 1월, AI가 지정학을 형성할 8가지 방식을 분석하면서 이렇게 정리했습니다. 미국, EU, 중국 세 디지털 강대국 사이에서 AI 인프라에 대한 통제를 강화하려는 움직임이 "AI 스택의 전투"로 진화하고 있다고. 컴퓨팅 파워, 클라우드 스토리지, 마이크로칩, 규제가 이 전투의 핵심 자원이고, 각 강대국이 이 자원들에 대해 점점 더 배타적인 접근 방식을 취하고 있다는 분석.

한편 2026년, UN이 뒷받침하는 AI 거버넌스에 관한 글로벌 대화와 독립 국제 과학 패널이 출범했습니다. 거의 모든 국가가 AI의 위험, 규범, 조율 메커니즘을 논의할 수 있는 포럼을 처음으로 갖게 된 것. 하지만 포럼이 있다는 것과 합의가 있다는 것은 다른 이야기입니다.

이 국제적 조율이 왜 어려운지, 김서준 대표의 분석이 구조를 잘 보여줍니다. 한국의 GDP는 약 1.8조 달러. 중국은 그 10배, 미국은 15~20배 규모입니다. "우리가 우리보다 10분의 1 작은 나라를 별로 의식하지 않는 것처럼, 중국도 우리를 의식할 이유가 없다." 반도체라는 특수한 영역을 제외하면 그렇다는 겁니다. 김 대표의 진단은 냉혹합니다. WTO 시대에는 덩치를 신경 쓸 필요가 없었지만, WTO는 끝났고 다시 오지 않는다. 국제 룰은 결국 힘에서 나오고, 상대편에 타격을 줄 수 있는 사이즈가 되어야 외교도, 무역도 가능하다.

그의 제안은 파격적입니다. 한일 경제통합. 두 나라를 합치면 GDP가 약 6조 달러. 중국의 약 3분의 1. "그 정도는 돼야 상대편이 의식할 수 있는 사이즈"라는 것. EU가 27개국이라 룰 하나 정하는 데 시간이 너무 걸리는 비효율을 반면교사로 삼아, 한일이 먼저 룰링 코어를 잡으면 효율적인 블록이 만들어진다는 구상. 더 나아가 이 6조 달러 블록이 형성되면 동남아시아 국가들이 자연스럽게 이 경제권 편입을 원하게 되고, EU 같은 아시아 유니언(Asia Union) 형태가 가능하다는 전망.

이 제안이 현실적인지는 별도의 논의가 필요합니다. 다만 이 제안이 던지는 질문은 분명합니다. AI 거버넌스의 국제적 조율에서 "rule taker"로 남을 것인가, "rule maker"가 될 것인가. 그리고 rule maker가 되려면 어떤 규모의 경제 블록이 필요한가.

미중 패권 경쟁은 수십 년간 지속될 것이고, 그 경쟁이 AI 거버넌스의 틀을 규정합니다. 미국은 혁신 우선, 중국은 통제 우선, EU는 규범 우선. 이 세 가지 접근법 사이에서 나머지 국가들은 선택을 강요받습니다. 모 가우닷이 경고한 죄수의 딜레마가 국가 차원에서 재현되는 겁니다. 모든 국가가 협력하여 AI 개발을 통제한다면 세계는 더 안전해질 것이지만, 어떤 한 국가라도 몰래 규제를 풀어버리면 그 나라가 경쟁 우위를 차지하게 됩니다. 이 두려움 때문에 어떤 나라도 먼저 규제를 강화할 수 없는 악순환.

블레츨리 선언에 28개국이 참여했던 2023년 11월의 순간이 있었습니다. 최소한의 공통 인식이 형성되는 듯했습니다. 하지만 선언과 이행 사이의 간극은 넓습니다. 2025년 파리에서 열린 AI 액션 서밋에서 EU는 글로벌 AI 경쟁의 방관자로 남고 싶지 않다는 의지를 보여주었지만, 서밋에서 나온 합의문은 법적 구속력이 없었습니다.

기후 변화나 질병 구호 같은 인류 공동의 과제를 해결하기 위해 AI가 급진적 풍요(Radical Abundance)를 제공할 수 있다는 기대가 있습니다. 그러나 국가 간 신뢰 부족으로 제로섬 게임의 틀을 벗어나지 못하고 있는 것이 현실입니다. 디지털 주권이라는 개념 자체가 로마 시대의 임페리움(Imperium)처럼 지리적 영역 내에서 절대적 통제권을 행사하려는 국가적 본능과 연결되어 있으니까요.

결국 AI 거버넌스의 국제적 조율은 기술의 문제가 아니라 정치의 문제입니다. 그리고 정치의 문제라면, 해답도 정치에서 나와야 합니다. 문제는 그 정치의 속도가 기술의 속도를 따라잡지 못한다는 것. 이것이 우리 시대의 가장 위험한 구조적 불일치입니다. 법안이 통과되는 데 2년이 걸리는 세계에서, AI 모델의 성능은 6개월마다 두 배가 됩니다. 이 격차가 좁혀질 수 있을까요, 아니면 영원히 벌어질까요. 답을 내리기에는 이르지만, 한 가지는 확실합니다. 격차가 벌어지는 동안 피해를 보는 것은 언제나 가장 약한 고리, 가장 작은 나라, 가장 젊은 세대입니다.

전체 목차

책의 모든 장 보기

김경진 변호사

에필로그

전체 목차

책의 모든 장 보기

13개의 장을 쓰는 동안 두려움이 손끝에서 떠나지 않았습니다. 일자리가 사라지는 속도, 불평등이 깊어지는 각도, 민주주의가 침식되는 경로, 인간관계가 희석되는 농도. 어느 하나 가볍지 않았습니다. 52개의 소주제를 펼쳐놓고 보니, 이것들은 각각의 위기가 아니라 하나의 거대한 지각 변동이었습니다.

그런데 이 책을 덮으면서 한 가지를 말씀드리고 싶습니다.

인류는 늘 이런 순간에 서 있었습니다. 불이 인간의 손에 들어왔을 때, 누군가는 숲을 태웠고 누군가는 음식을 익혔습니다. 인쇄술이 퍼졌을 때, 선동의 도구가 되기도 했지만 지식의 민주화를 열기도 했습니다. 핵분열의 원리를 알아냈을 때, 히로시마가 있었지만 원자력 발전소도 있었습니다. 기술 자체는 방향을 갖고 있지 않습니다. 방향을 정하는 것은 언제나 인간의 몫이었습니다.

AI 시대도 다르지 않습니다. 균열이 보인다는 것은, 아직 고칠 시간이 있다는 뜻이기도 합니다. 김서준 대표가 "균열이 생기는 곳마다 반대편에는 반드시 기회의 창이 열린다"고 쓴 것처럼, 무너지는 질서의 잔해 위에서 새로운 좌표를 찾는 일이 우리 앞에 놓여 있습니다.

완벽한 설계도는 없습니다. 있을 수도 없습니다. 그러나 방향은 잡을 수 있습니다. 노동의 존엄을 지키면서 기술의 생산성을 받아들이는 방향. 교육의 본질을 보존하면서 낡은 형식은 내려놓는 방향. 효율을 추구하되 인간 사이의 접촉과 유대를 포기하지 않는 방향. 그 방향을 향해 한 걸음씩 움직이는 것, 실패하고, 수정하고, 다시 움직이는 것. 그것이 인류가 수만 년 동안 해온 일이고, 앞으로도 해야 할 일입니다.

이 책이 기록한 52개의 과제는 숙제가 아니라 좌표입니다. 우리가 어디에 서 있는지를 알려주는 지도. 그 지도를 들고 어디로 걸어갈지는, 이 책을 읽은 당신의 선택입니다.

2026년 5월 김경진

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.