AI 알고리즘, 각자의 업을 비추다

붓다, AI 윤리에 답하다

⑯ 사람을 차별하는 AI?

AI는 누구의 마음을 학습했는가

2023년 2월의 어느 아침이었다. 미국 디트로이트 교외, 서른두 살의 흑인 여성 포샤 우드러프는 두 딸의 등교 준비를 하다가 문을 두드리는 소리를 들었다. 여섯 명의 경찰관이 서 있었다.

그들이 보여준 것은 강도와 차량 탈취 혐의의 체포 영장이었다. 우드러프는 임신 8개월이었다. 처음엔 농담인 줄 알았다고 한다. 부풀어 오른 배를 손으로 짚어 보였다. 영장에 첨부된 CCTV에 찍힌 범인 사진은 누가 봐도 임신한 몸이 아니었다. 그러나 영장은 이미 나와 있었고, 아이들이 지켜보는 앞에서 그녀는 수갑을 찼다.

유치장으로 호송된 그녀는 요통과 경련 속에서 열한 시간을 견뎌야 했다. 이후 보석으로 풀려나자마자 병원으로 실려 갔고, 탈수 진단을 받았다.

한 달 뒤 검찰은 증거가 부족하다며 기소를 취하했다. 황당하게도 그녀를 범인으로 지목한 것은 사람이 아니었다. 주유소의 흐릿한 CCTV 영상을 경찰의 얼굴인식 시스템에 통과시켰더니, 그녀의 오래된 머그샷이 후보군에 올라왔던 것이다. 경찰은 그 사진을 목격자에게 보여줬고, 목격자가 그녀를 골랐다.

영장은 그렇게 발부됐다. 임신한 몸이라는, 눈앞의 사람이라면 놓칠 리 없는 사실을 시스템은 분석하지 못했다. 시스템이 아는 것은 픽셀의 배열이 비슷하다는 것 정도가 전부였다.

미국시민자유연맹의 집계에서 우드러프는 얼굴인식 오류로 부당하게 체포된 여섯 번째 미국인이자, 그 가운데 첫 여성이었다. 여섯 명 모두 흑인이었다. 2026년 4월까지 공개적으로 확인된 피해자는 최소 열세 명으로 늘었고, 그 대다수는 여전히 흑인이다.

우연으로 돌리기 어려운 이 쏠림 현상은 사실은 예고된 것이기도 했다. 미국 국립표준기술연구소(NIST)의 대규모 평가에서 얼굴인식 알고리즘이 흑인 여성에게 백인 남성보다 수십 배 높은 오탐률을 보인다는 사실은 이미 오래전에 확인돼 있었기 때문이다.

2025년 8월, 미시간 동부 지방법원은 이 체포가 "여러 이유로 우려스럽다"고 인정하면서도 우드러프의 민권 소송을 기각했다. 경찰이 그녀를 체포할 정당한 근거가 없었다는 점을 충분히 입증하지 못했다고 판단한 것이다. 사람을 픽셀로 인식하는 AI 시스템에 대해, 법은 아직 알고리즘을 심판할 준비가 돼 있지 않은 듯하다.

이런 종류의 AI 시스템에서 흑인의 얼굴에 관한 인식 오류가 빈번한 이유는 사실 복잡하지 않다. AI가 학습한 얼굴 사진의 대부분이 백인이었기 때문이다.

AI 시스템은 데이터를 주어진 대로 학습했을 뿐이고, 그 데이터에는 흑인의 데이터 표본이 부족했다. 일부러 흑인으로 오인하라고 알고리즘 설계하지 않는 이상, 문제는 AI 시스템의 오류가 아니라, AI가 무엇을 학습했는가에 있다.

변수를 지워도 얼룩은 남는다

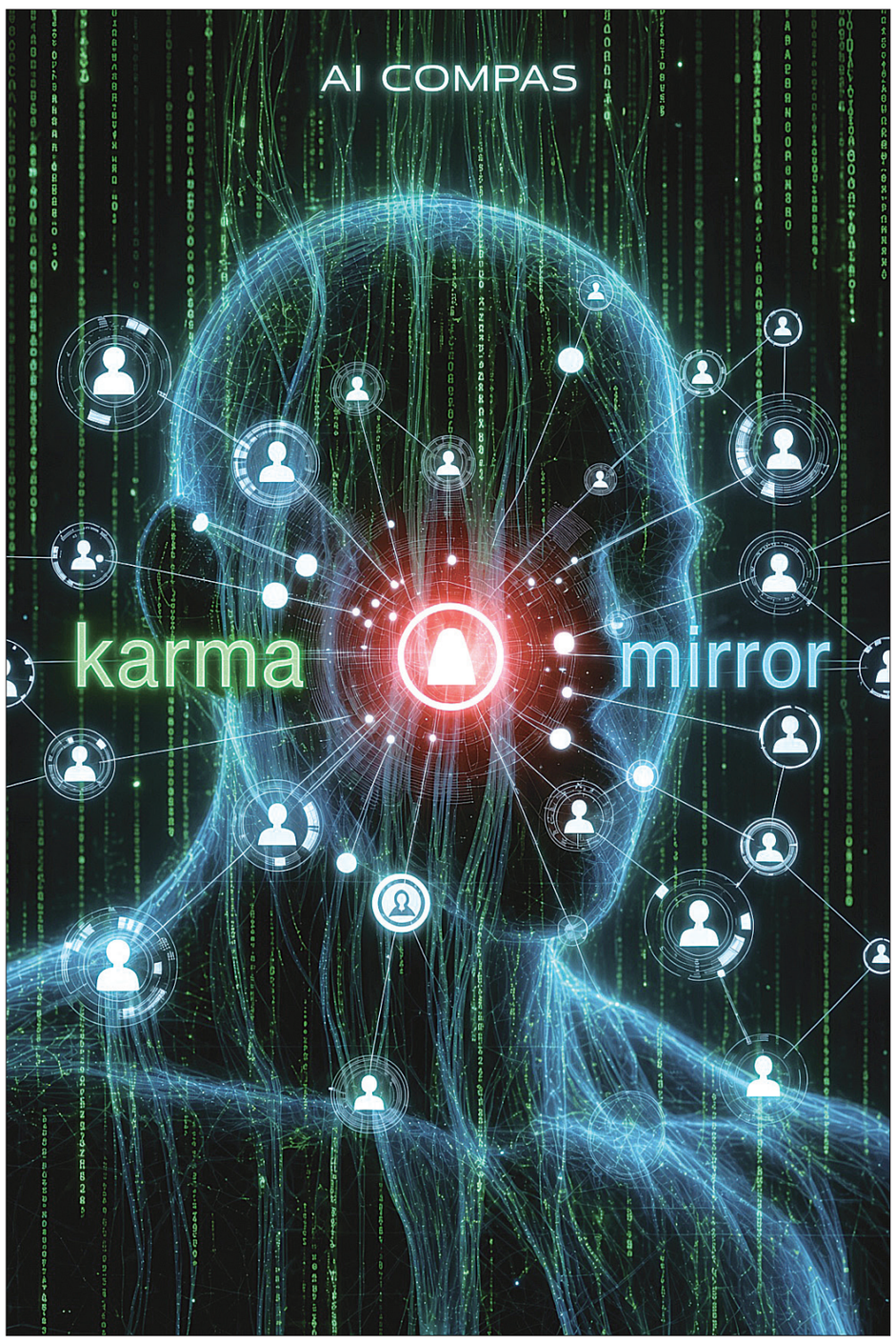
여기서 흔히 나오는 반론이 있다. 민감한 변수를 빼면 되지 않느냐는 것이다. 인종 칸을 없애고 성별 칸을 없애면 차별도 사라지리라는 기대다. 현실은 그 기대를 자주 저버려 왔다.

2019년 '애플카드' 논란을 떠올려 본다. 부부가 재산을 공동으로 관리하고 아내의 신용 점수가 오히려 더 높은데도, 남편에게 스무 배나 되는 신용 한도가 주어졌다는 폭로가 있었다. 뉴욕주 금융 당국의 조사는 발급사가 성별을 변수로 쓰지 않았다는 결론을 내렸다.

그러나 사건이 남긴 물음은 오히려 그래서 더 심각해졌다. 성별을 입력하지 않았다는 사실이, 정말로 성별에 따른 차별이 없다는 뜻일까. 소득의 흐름과 소비 패턴, 직업의 분포 같은 우회 변수에는 이미 성별이 스며들어 있다. 이름표를 뗀다고 얼룩까지 지워지지 않는 것이다.

이 우회로가 한 사람의 자유를 좌우한 사례도 있다. 미국의 여러 법원이 형량을 정할 때 참고해 온 '재범 위험도 알고리즘 콤파스(COMPAS)'가 그것이다. 2016년 프로퍼블리카의 탐사보도는, 이 시스템이 흑인 피고인을 백인보다 거의 두 배나 높은 비율로 '고위험군'에 잘못 분류해 왔다는 사실을 드러냈다. 인종은 변수에 들어가지 않았다. 그러나 '체포 이력'이라는 데이터 안에는 이미 과잉 단속의 역사가 깊이 새겨져 있었고, 알고리즘은 그 흔적을 '위험도'라는 숫자로 옮겨 적었을 뿐이다.

> 우드러프, 얼굴인식 오류로 부당 체포
> 흑인 얼굴인식 오류가 빈번한 이유는
> AI 학습 얼굴 대부분이 백인이기 때문
> 데이터는 집단으로 쌓은 업의 퇴적층
> 약자 품는 마음이 AI의 미래 결정할 것



AI가 스스로 업의 증폭기가 될 것인지, 정견(正見)의 거울이 될 것인지를 결정하는 것은 아니다. 그 흔적을 직시하고 부끄러워하며 밭을 갈아엎을 것인가, 소송을 기각하고 시스템을 그대로 돌리며 흔적만 가리고 돌아설 것인가. 알고리즘을 설계하면서 누가 소외돼 있는지 살피는 마음, 가장 약한 이의 고통이 통계 바깥으로 새어 나가지 않는지 들여다보는 마음이 AI의 미래를 결정한다. 이미지 출처=Grok 생성 이미지

업의 증폭기인가, 거울인가

초기 경전 〈맛지마니까야〉 '업 분별의 짧은 경 (MN135)'에서 붓다는 이렇게 말씀하신다. "중생은 자신을 업을 소유하는 자이며, 업의 상속자이고, 업을 모태로 하는 자이며, 업을 친지로 하는 자이고, 업을 의지처로 하는 자이다. 업이 중생들을 나누어 천하고 귀한 상태를 만든다."

업이란 어떤 신비한 형벌 장부 같은 것이 아니다. 의도를 지닌 행위가 흔적을 남기고, 그 흔적이 다시 조건이 되어 다음 행위를 물들이는 과정을 가리키는 말이다.

이 가르침을 지금 벌어지고 있는 일 위에 겹쳐 놓으면 윤곽이 또렷해진다. 데이터란 무엇인가. 한 사회가 수십 년, 아니 수백 년에 걸쳐 쌓아온 선택과 배제의 기록, 곧 집단으로 지어온 업의 퇴적층이다.

AI는 그 퇴적층을 학습한다. 그러니 알고리즘이 임신부에게 수갑을 채우고 여성을 남성보다 신용도가 낮다고 간주할 때, AI가 새로 만들어낸 것은 사실 아무것도 없다. 다만 우리가 지어온 것을 상속받아, 전에 없던 규모로 되풀이하고 있을 뿐이다.

유식(唯識) 전통은 이 상속의 구조를 한층 정밀하게 그려두었다. 모든 경험은 마치 일종의 씨앗[種子]이 되어 심층 의식인 알라야식(阿賴耶識)에 저장되고, 조건이 무르익으면 본인도 알아차리지 못하는 사이에 행위와 판단으로 드러난다[現行].

인종 칸을 지워도 체포 이력을 통해 인종이 스며드는 일, 성별 칸을 지워도 소비 패턴으로 성별이 배어나는 일이 꼭 그렇다. 의식의 표면에서 지운 것이 심층의 씨앗까지 지운 것은 아니다.

그리고 그 씨앗을 심어 온 손은, AI가 아니라 다름 아닌 우리 자신이었다. 물론 업을 숙명으로 읽는 것은 불교에 대한 오해다. 업이 연기(緣起)의 법칙 위에 서 있다는 것은, 조건이 바뀌면 흐름도 바뀐다는 뜻이기 때문이다. 씨앗은 반드시 싹을 틔우겠지만, 어떤 밭에 어떤 물을 주느냐는 아직 정해져 있지 않다.

사람의 편견은 좀처럼 본인의 눈에는 보이지 않는다. 목격자는 자신이 본 대로 지목했다고 믿고, 카드 회사는 고객을 차별하지 않았다고 믿는다.

그런데 그 보이지 않던 마음이 알고리즘을 통과하는 순간, 처음으로 통계라는 자국을 남긴다. 여기저기 흩어져 있어 좀처럼 잡히지 않던 분별심(分別心)이, 숫자가 되어 눈앞에 놓이는 것이다.

AI가 스스로 업의 증폭기가 될 것인지, 정견(正見)의 거울이 될 것인지를 결정하는 것은 아니다. 그 흔적을 직시하고 부끄러워하며 밭을 갈아엎을 것인가, 소송을 기각하고 시스템을 그대로 돌리며 흔적만 가리고 돌아설 것인가.

알고리즘을 설계하면서 누가 소외돼 있는지 살피는 마음, 가장 약한 이의 고통이 통계 바깥으로 새어 나가지 않는지 들여다보는 마음이 AI의 미래를 결정한다.

이제 알고리즘은 우리 각자의 업을 비추는, 유례없이 크고 정직한 거울로 이미 우리 앞에 서 있다. 남은 것은 거울 속 얼굴을 마주할 용기다. 그 용기를 내는 자리에서, 비로소 AI 알고리즘은 우리를 심판하는 도구가 아니라 함께 길을 묻는 연(緣)이 된다.



보일 스님
AI부디즘연구소장

체질 개선과 건강 회복을 위한 합리적 조건
당신의 날씬한 건강이 시작되는 곳
'초록원웰빙'
since 1984
프리미엄 효소 제품과 다이어트 제품을 만나보세요


NAVER 초록원웰빙 검색
네이버 스마트스토어에서 만나요
강남점 (도곡동 467-24)

김영숙 대표
02-578-1708 / 010-7384-1708

민속죽염
30년 전통
조상의 얼과 혼이 담긴 우리 민족의 대표 죽염
인체건강의 4대 요소는 "맑은 물, 맑은 공기, 맑은 음식, 질좋은 소금" 인데 이 중 "질 좋은 소금" 자색죽염은 총 아홉 번의 가열 과정을 거쳐 독성의 완전 제거와 약성의 완전 합성이 이루어지며 마지막 아홉 번째는 극강한 열에 의한 용융과정을 거쳐 탄생된 품질이 가장 우수한 죽염입니다. 이 죽염은 다음과 같은 기능이 있기에 일반인 및 환자들에게 꾸준한 섭취를 권장 할 수 있는 좋은 식품이라 할 수 있습니다.
① 강한 해독작용을 하며 간 기능을 좋게 한다.
② 소염작용을 하며 잇몸과 치아를 튼튼하게 해 준다.
③ 위장을 튼튼히 하며 식욕을 촉진시킨다.
④ 정혈작용 등 체질개선에 도움을 준다.

생명 에너지를 굳건히 지키면 약이 입에 이르기 전에 병이 낫는다.
환인동 요가마을
"죽염을 꾸준히 섭취하시면 인체의 자가 면역력이 높아져 감기 등 기타 질병 예방에 크게 도움이 됩니다"

◈ 9번 구운 민속죽염 ◈
가루자죽염 125g 45,000원
가루자죽염 230g 82,000원
알갱이자죽염 125g 50,000원
알갱이자죽염 230g 90,000원
휴대용 알갱이 자죽염 20g 9,000원
◈ 2번 구운 민속죽염 ◈
◆ 음식조리용 · 양치용 · 피부미용
보급형 민속죽염 500g 20,000원
250g 10,000원
※ 단체 선물용으로 적합합니다. ※ 구입가 5만원 미만은 배송비 3천원 적용됩니다.
주문 및 상담 010-2823-0483
입금 : 농협 150012-56-086122 예금주 : 김용희(민속죽염)
경북 영덕군 지품면 속곡길381-4 www.msjy.co.kr