AI 기술 설계 방향, 인간에 달렸다

붓다, AI 윤리에 답하다

⑰ AI시대, 양극화를 넘어서는 길

서명 하나가 빠졌을 뿐이었다

2021년 1월 15일, 네덜란드 정부가 총사퇴했다. 유럽에서 오랜 기간 집권한 총리 중 한 명이던 마르크 뤼터가 내각 전원과 함께 물러난 이유는 뇌물 스캔들도, 금융위기도 아니었다. 바로 AI 알고리즘이었다. 네덜란드 세무 당국이 보육수당 부정수급을 걸러내겠다며 도입한 자기 학습 알고리즘은 이중국적과 낮은 소득, 심지어 '외국풍 이름'까지 위험 지표로 삼았다.

일단 사기범으로 분류되면 수당은 즉시 끊겼고, 수년간 누적된 수당 전액을 한꺼번에 반환하라는 통지가 날아들었으며, 분할 납부도 항변도 사실상 허용되지 않았다. 조사가 시작된 계기가 서명 하나 빠진 서류 같은 사소한 행정 실수인 경우도 적지 않았다.

그렇게 2만6000명이 넘는 부모가 사기범으로 내몰렸다. 수만유로의 빚더미 속에서 파산과 실직, 이혼이 잇따랐고, 1600명이 넘는 아이들이 부모와 강제로 분리됐다. 의회 조사위원회는 이를 "전례 없는 불의"라 불렀고, 이듬해 정부는 사태의 뿌리에 '제도적 인종주의'가 있었음을 공식 인정했다.

네덜란드 정부는 피해 가구당 3만유로의 배상을 내놓았지만, 한 번 무너진 삶은 되돌아오지 않았다. 피해 가정의 다수는 이주 배경을 가진 저소득층이었다.

알고리즘은 사회의 가장 약한 자리를 정확히 겨냥했다. 효율의 이름으로 설계된 기술이 왜 하필 가장 취약한 이들 앞에서 흉기로 변하는가. 이 질문이야말로 AI시대의 양극화를 이해하는 핵심 열쇠다.

의심은 먼저 닿고, 혜택은 늦게 닿는다

네덜란드의 비극은 우연한 오작동이 아니었다. 국제앰네스티는 이 시스템을 조사한 보고서(2021.10.25.)에 '외국인 혐오 기계(Xenophobic Machines)'라는 제목을 붙였다. AI 자기 학습 알고리즘은 한 번 의심한 집단을 더 자주 조사했고, 조사가 늘어날수록 그 집단의 위험 점수는 더 높아졌다. 차별이 차별을 낳고 다시 차별을 학습하는 악순환이었다. 이중국적이라는 이유 하나만으로 추가 조사를 받은 사람이 1만1000명에 이르렀다. 비단 한 나라의 예외적인 사고도 아니었다.

3년 뒤 국제앰네스티는 이웃 덴마크의 복지 시스템에서도 국적과 가족 형태를 위험 지표로 삼는 유사한 알고리즘을 확인했다(국제앰네스티, 2024). 부유한 복지국가에서조차 감시하고 의심하는 AI는 가장 낮은 자리의 시민에게 가장 먼저, 가장 가혹하게 작동한다. 반면 사람을 돕는 기술은 정작 도움이 가장 절실한 자리에 가장 늦게 도착한다는 사실이다.

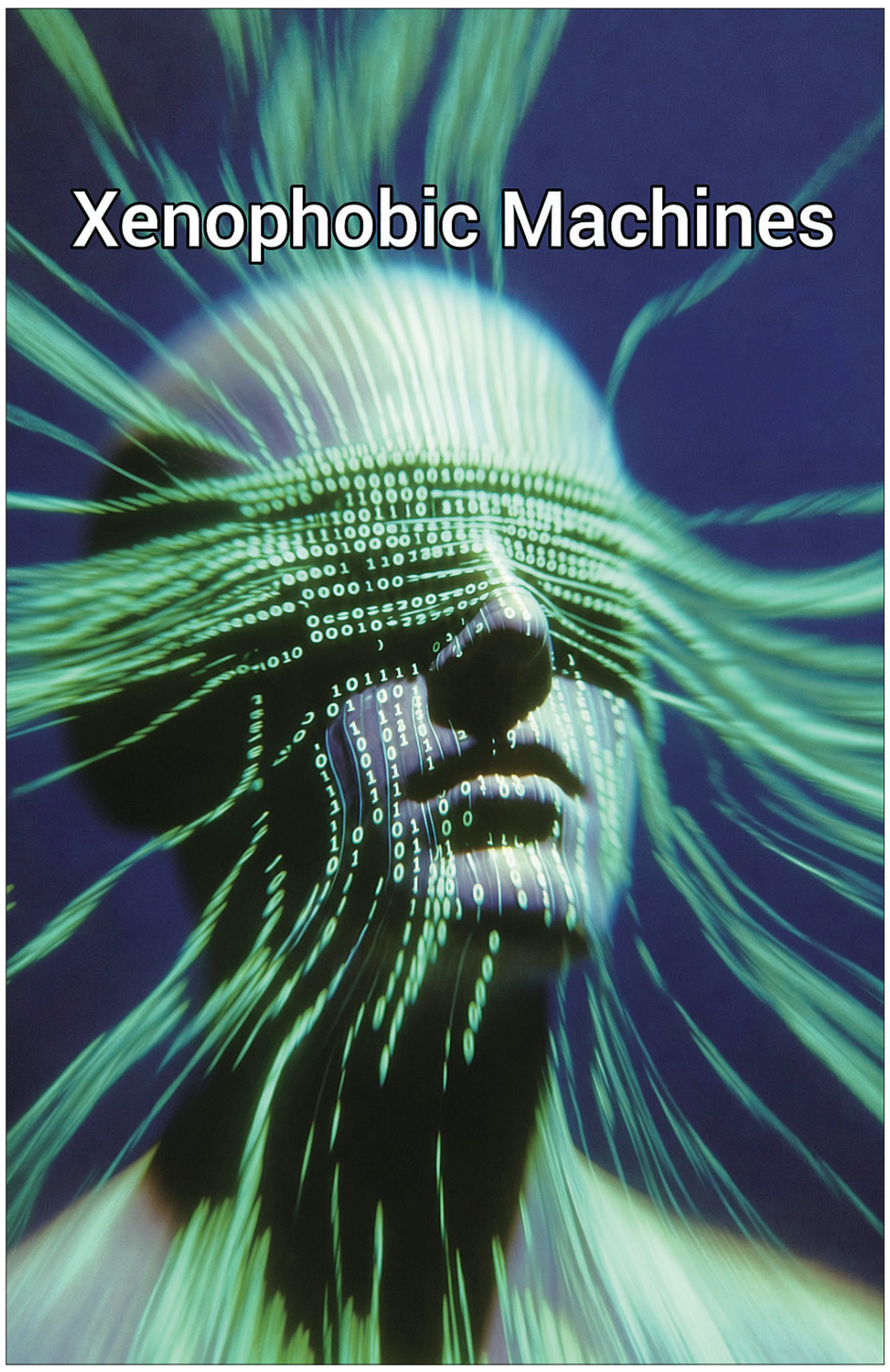
국제전기통신연합에 따르면 2025년에도 세계 인구 22억명이 인터넷 없이 살아가며, 그 96%가 저소득·중소득 국가 사람들이다(국제전기통신연합, 2025). 세계은행 역시 사하라 이남 아프리카 농촌 학교의 약 3분의 1이 안정적인 전기조차 공급받지 못한다고 지적한다(세계은행, 2026). 전기가 없는 교실에는 아무리 뛰어난 AI 교사도 다가갈 수 없다.

한국이라고 사정이 다를까? 그 구조는 크게 다르지 않다. 고령층·장애인·저소득층·농어민 등 취약계층의 디지털 접근 수준은 일반 국민의 96.6%에 이르러 기기 보급은 사실상 완성됐지만, AI 서비스를 실제 이용해 본 경험률은 31.9%로 일반 국민(59.4%)의 절반에 그친다('2025 디지털 정보격차 실태조사', 2026.3.27 발표). 기기는 손에 쥐여 주었지만, 그것을 이해하고 다루는 법은 함께 건네지 못한 것이다.

> 2021년 네덜란드 정부 총사퇴 이면엔 세무 당국의 자기 학습 알고리즘 문제
> 차별이 차별을 낳고 학습하는 악순환 설계에 따라 약자 돕거나 의심하기도
> 두 갈래 길서 불교의 질문 현재도 유효

결국 의심의 기술은 약자에게 가장 먼저 다가가고, 혜택의 기술은 약자에게 가장 늦게 도달한다. 양극화는 바로 이 두 속도의 차이에서 비롯된다.

그러나 같은 기술이 정반대 방향으로 쓰인 사례도 분명히 있다. 세계은행이 나이지리아에서 진행한 현장 실험에서는 교사가 이끄는 6주간의 방과 후 AI 개인 교습이 학생들의 학습 성취도를 통상적인 학교 교육 1년 반에서 2년 치에 해당하는 만큼 끌어올렸다(세계은행, 2025). 같은 기술이 한 나라에서는 무고한 시민을 사기범으로 지목했고, 다른 나라에서는 아이들의 배움을 곁에서 도왔다. 이 갈림길을 만든 것은 기술 자체가 아니라 그 기술을 누구에게 먼저, 어떻게 다가가게 할지를 정한 인간의 선택이다.



AI 자기학습 알고리즘은 한 번 의심한 집단을 더 자주 조사했고, 조사가 늘어날수록 그 집단의 위험 점수는 더 높아졌다. 차별이 차별을 낳고 다시 차별을 학습하는 악순환이었다. 이중국적이라는 이유 하나만으로 추가 조사를 받은 사람이 1만1000명에 이르렀다. AI Grok 생성 이미지

알고리즘의 눈, 아버지의 눈

〈묘법연화경(妙法蓮華經)〉 '신해품(信解品)'에는 이 갈림길을 2000년 앞서 꿰뚫어 본 비유가 등장한다. 어려서 아버지 곁을 떠난 아들이 오십 년을 타국에서 떠돌며 걸식하다가, 자신도 모르는 사이에 아버지가 사는 큰 성에 이른다. 아버지는 그사이 거부가 돼 있었다. 사자좌에 앉아 뭇사람의 예를 받는 위엄 앞에서 아들은 그가 아버지인 줄도 모른 채, 왕이거나 왕에 버금가는 이가 사는 곳이니 품을 팔 데가 못 된다며 달아난다. 그러나 아버지는 누더기 차림의 그를 한눈에 아들로 알아보았다. 기뻐서 사람을 보내 급히 붙잡아 오게 하자, 아들은 영문도 모른 채 죽임을 당하는 줄 알고 그 자리에서 기절해 버린다.

여기서부터가 이 비유의 핵심이다. 아버지는 아들을 놓아주게 한다. 신분을 밝히지 않고, 초라한 행색의 두 사람을 보내 거름 치우는 품삯 일을 권하게 한다. 그리고 자신도 보배 옷을 벗고 때 묻은 작업복으로 갈아입은 채 아들 곁에서 함께 일하며, 20년에 걸쳐 천천히 신뢰를 쌓는다. 임종이 가까워서야 아버지는 모든 식솔을 모아 놓고 그 걸인이 자기 아들이며, 모든 재산은 본래부터 그의 것이었다고 선언한다. 전통적 해석으로는 이 비유에서 아버지는 부처요, 가난한 아들은 자신이 무엇의 상속자인지 잊어버린 우리 모두다. 같은 장면을 뒤집으면 그대로 AI시대의 풍경이 된다. 남루한 행색의 사람이 부유한 시스템의 문 앞에 선다.

네덜란드의 알고리즘은 그 행색에서, 낮은 소득과 외국풍 이름에서 사기의 위험을 읽었다. '신해품'의 아버지는 똑같은 누더기에서 자기 아들을, 유산의 주인을 읽어 냈다. 눈앞의 정보는 같았으나 읽어낸 가치는 정반대였다. 하나는 밀어낼 근거를 찾았고, 하나는 품어 안을 길을 찾았다. 불교는 이 차이를 지혜와 방편이라 부른다. 상대의 본래 모습을, 있는 그대로 알아보는 것이 지혜라면, 상대가 놀라 달아나지 않도록 다가가는 걸음을 궁리하는 것이 방편이다. 아버지는 둘 다를 갖췄다. 급히 붙잡으려 하자 상대가 기절하는 것을 보고는 방식을 바꿨고, 끝내 자신이 작업복을 입고 곁으로 다가섰다.

나이지리아의 교실에서 AI 개인 교습이 성과를 낸 것도 AI가 홀로 뛰어나서가 아니라 교사가 아이들 곁에 서 있었기 때문이다. 지구상의 22억명의 오프라인 인구도, 접근은 되지만 활용법을 배우지 못한 고령층도, 서명 하나 때문에 사기범으로 몰렸던 그 부모들도, 본래 이 AI 기술 문명이라는 유산의 상속자다. 지금 필요한 것은 상속자를 의심하도록 설계된 시스템이 아니라, 상속자를 알아보고 그가 선 자리까지 내려가도록 설계된 시스템이다.

AI 기술은 스스로 누구를 먼저 찾아갈지 결정하지 못한다. 그 결정은 언제나 인간의 몫으로 남는다.

네덜란드의 알고리즘처럼 가장 약한 삶을 가장 먼저 의심하도록 설계할 수도 있고, 나이지리아의 교실처럼 가장 뒤처진 자리에 먼저 손을 내밀도록 설계할 수도 있다.

두 갈래 길 앞에서 불교가 오래전부터 물어온 질문은 지금도 유효하다. 이 선택이 누구의 고통을 줄이고, 누구의 고통을 늘리는가. 유산은 이미 세상에 나와 있다.

남은 것은 그 유산이 닿지 못한 자리를 찾아가는 걸음이며, 그 걸음의 방향을 정하는 일이야말로 AI시대의 윤리가 떠안아야 할 가장 구체적인 과제다.



보일 스님
AI부디즘연구소장



틱낫한 스님의 『안반수의경』 해설
숨을 쉬세요, 당신은 살아 있어요!
불교 수행의 가장 기초적이고 근본적인 수행인 수식관의 가르침을 담고 있는 『아나파나사티 숫타(안반수의경)』를 틱낫한 스님이 쉽고도 현대적으로 해설하였다.
틱낫한 지음, 허우성 옮김/국판변형/192쪽/값 14,000원

교실에서 붓다의 가르침 실현하기
불교로 가르치다
붓다의 대화법과 마음챙김 수행을 통해 교육현장에서의 가르침과 배움을 새롭게 바라보게 하는 교육 실천서로, 일방 주입식 교육에 혁신적 변화를 가져다줄 것이다.
시드 브라운, 신희정 옮김/신국판/320쪽/값 22,000원

불교와 서양 심리학의 만남
새로운 불교심리학
12개국 38명의 저자가 서양 심리학의 연구 방법론과 불교의 실천적 지혜를 융합하여 새로운 통합 심리학 모델과 이론을 소개하는 기념비적인 책이다.
모리츠 키 외 편집, 윤희조·박재은 옮김/신국판/872쪽/값 50,000원

도서출판 운주사 서울시 성북구 동소문로 67-1 성심빌딩 3층 ● 전화 (02) 926-8361 ● 팩스 0505-115-8361 ● 주문 (02) 3672-7181 ● 다음카페:도서출판 운주사 www.cafe.daum.net/unjubooks