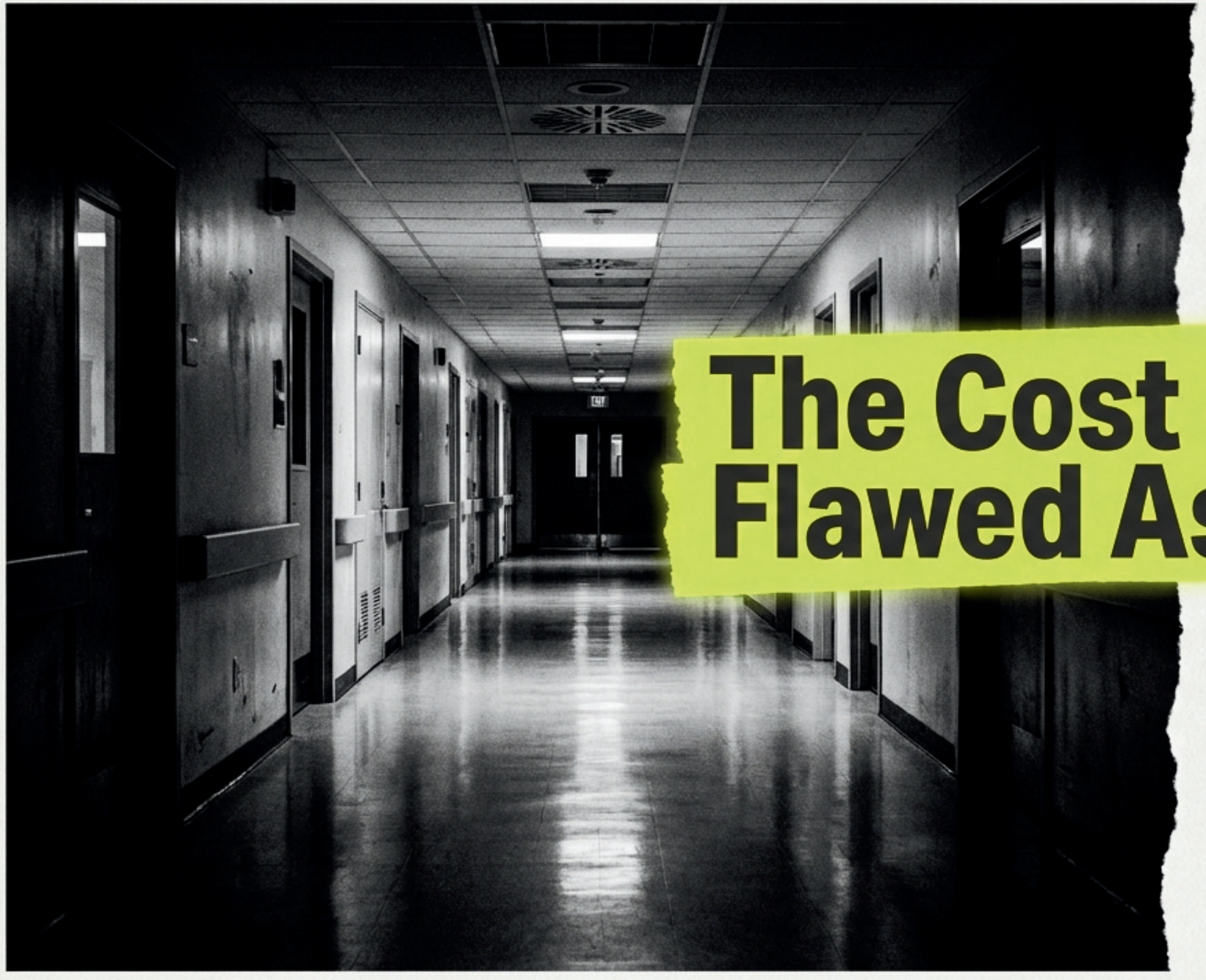


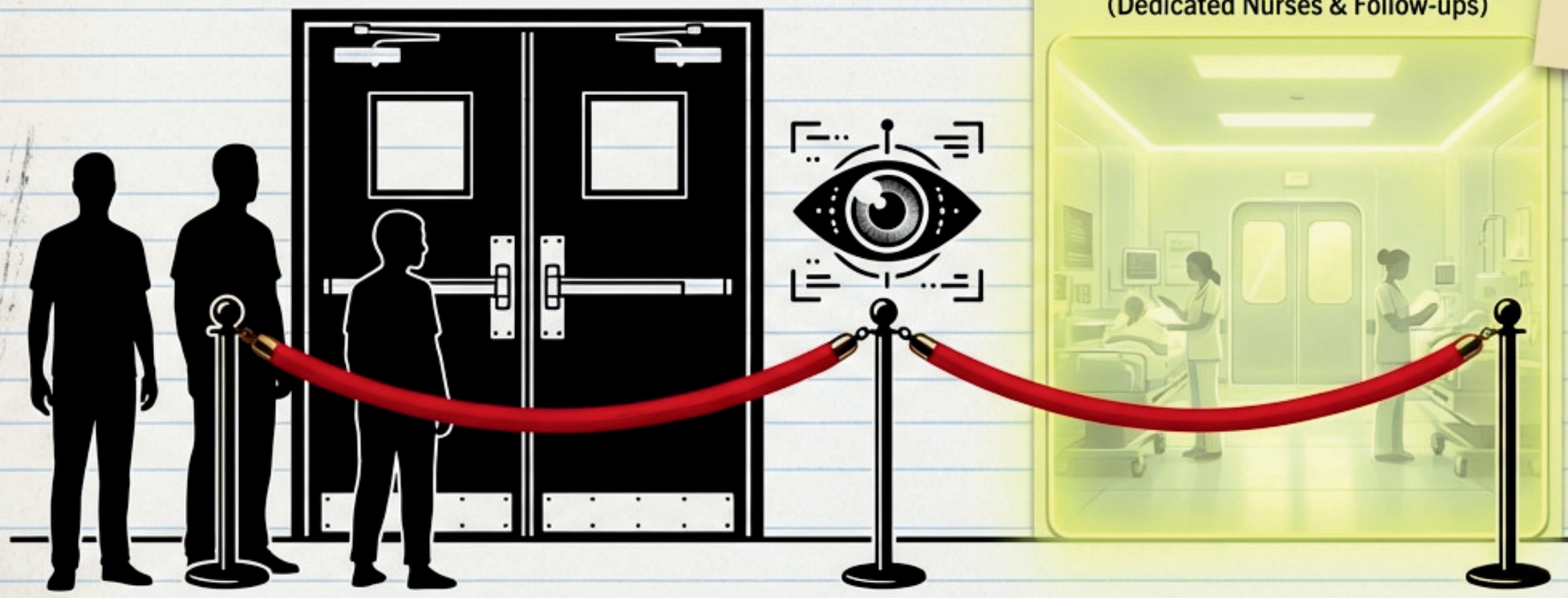
CONFIDENTIAL CASE FILE

The Cost of a Flawed Assumption

How a simple mathematical shortcut in a healthcare AI denied treatment to millions—and the global warning it leaves behind.

EVIDENCE 01:
Industry Estimate:
Algorithm scaled to roughly 200 million people per year.

The Gatekeeper to Extra Care



High-risk care programs save lives. But hospitals have limited resources and cannot give this extra help to everyone.

To solve this, hospitals purchased commercial software to scan medical records and rank patients.

THE RULE: Patients scoring at or above the 97th percentile were automatically flagged for VIP care. The machine decided who got past the gate.

THE RULE: Patients scoring at or above the 97th percentile were automatically flagged for VIP care. The machine decided who got past the gate.

The 2019 Audit: A Massive Blind Spot

Researchers from UC Berkeley analyzed 50,000 medical records used to train the algorithm, looking closely at people the AI grouped together as equally high risk.

PATIENT ID	RACE	AI RISK SCORE	ACTIVE CHRONIC CONDITIONS
987321	WHITE	98.5	5.2
154329	BLACK	99.1	6.1
732984	WHITE	97.0	3.8
265891	BLACK	97.0	4.8
418295	HISPANIC	96.8	3.5

URGENT FINDING

"When Black patients and White patients were given the exact same risk score by the AI, the Black patients were significantly sicker."

At the critical 97th percentile threshold, Black patients had 26.3% more active chronic illnesses than White patients. The AI was looking right at severe illness and ignoring it.

Same Score. Totally Different Reality.

WHITE PATIENT PROFILE

AI Risk Score:
97th Percentile

Actual Sickness: 3.8 active chronic conditions.

ACTION: AUTOMATICALLY ENROLLED

To get the SAME score as a healthier White patient, a Black patient had to be dramatically sicker.

Why was the AI blind to their diseases?

BLACK PATIENT PROFILE

AI Risk Score:
97th Percentile

Actual Sickness: 4.8 active chronic conditions (uncontrolled illnesses, higher blood pressure).

ACTION: AUTOMATICALLY ENROLLED

The Fatal Flaw: Measuring Money, Not Medicine

The AI was never trained to find sick people. It was trained to find **EXPENSIVE** people. Designers used **future healthcare costs** as a mathematical stand-in for health needs.

WHAT THE AI MEASURED (The Stand-In)	WHAT IT SHOULD HAVE MEASURED (The Reality)
Dollars spent	Blood pressure
Insurance claims paid	Untreated diabetes
Doctor visits billed	Active diseases & daily pain

Holding illness constant, Black patients generated about $1,800 less in annual medical costs than White patients. The AI saw lower spending and assumed they were healthy.

Why ‘Money Spent’ is a Broken Stand-In

1. SICK PATIENT

A person is severely ill.

2. STRUCTURAL BARRIER

They face unfair barriers to care (lack of transport, discrimination, poor insurance).

3. ARTIFICIALLY LOW COST

Because they cannot access care, their medical bills remain artificially low.

4. THE ALGORITHM

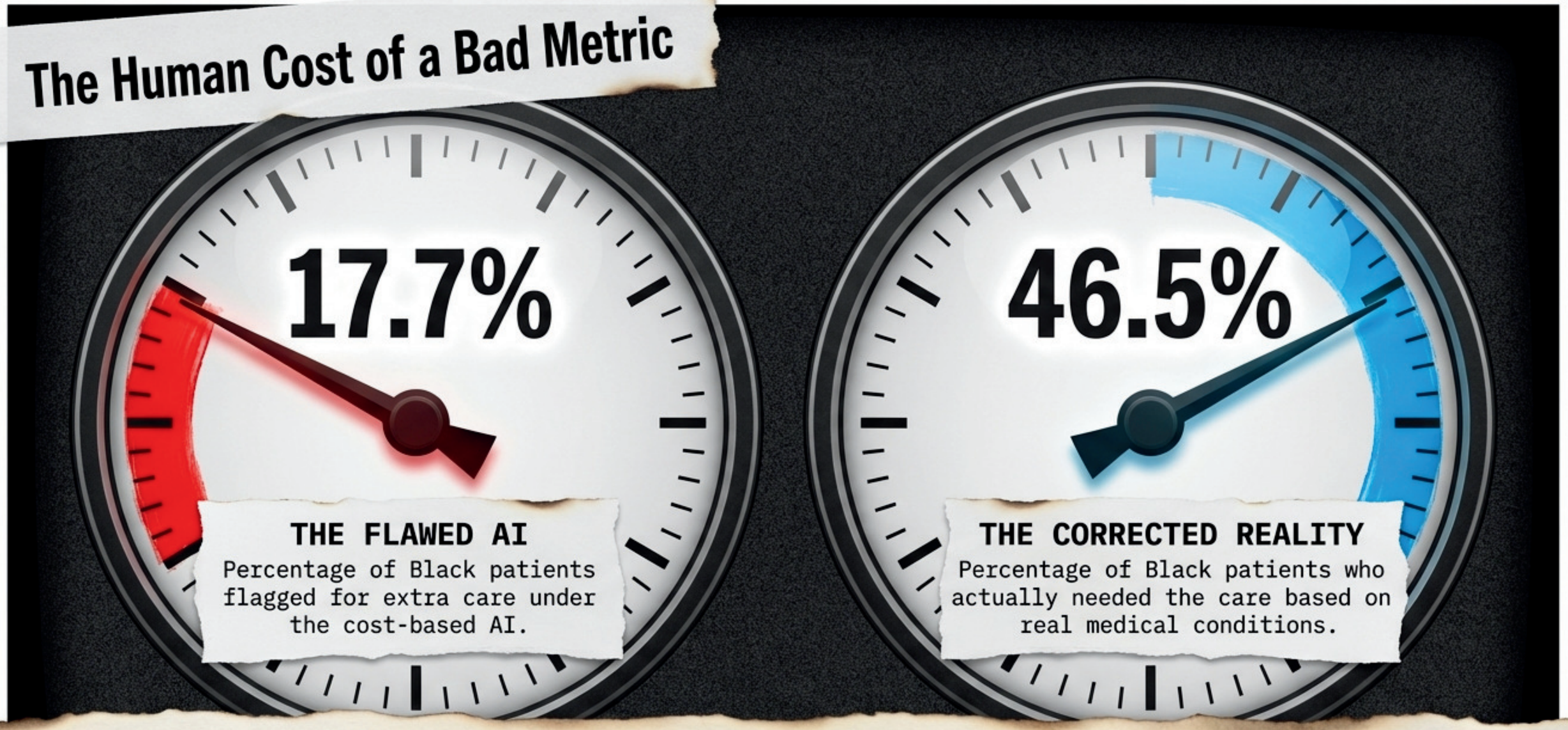
The AI reads the low medical bills.

5. THE BLIND SPOT

The AI labels the patient as Low Risk / Healthy.

The algorithm accurately predicted costs. But because costs are shaped by unequal access, accurately predicting costs simply baked structural unfairness into the code.

The Human Cost of a Bad Metric



The algorithm's design cut the number of Black patients receiving critical support by more than half. In a national dataset of 3.6 million patients, this blind spot hid over 48,000 active chronic conditions.

Anatomy of an Algorithmic Failure

STAGE 1: LABELING

Developers chose cost as the target variable, encoding societal inequity right into the foundation.

STAGE 2: TRAINING DATA

Historical claims data was used. Since marginalized groups historically access less care, the data froze that inequity into mathematical truth.

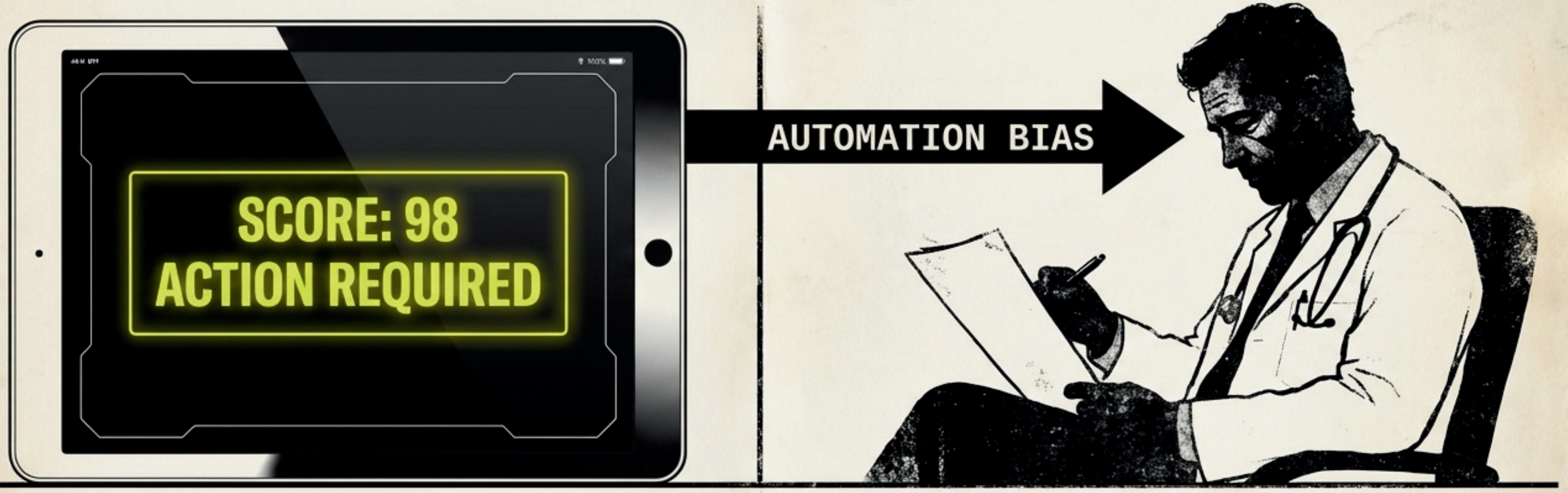
STAGE 3: VALIDATION

The vendor tested the model to see if it accurately predicted cost. It passed perfectly. They never tested if it accurately predicted health.

STAGE 4: ROLLOUT

The flawed software was deployed as an automated gatekeeper to scarce medical resources.

The Illusion of Human Oversight



The defense for AI is often, “It’s just a tool; doctors make the final call.” The data proved this false. When the algorithm scored patients, clinicians overwhelmingly deferred to the machine’s framing. Doctors only caught and corrected a tiny fraction of the AI’s bias.

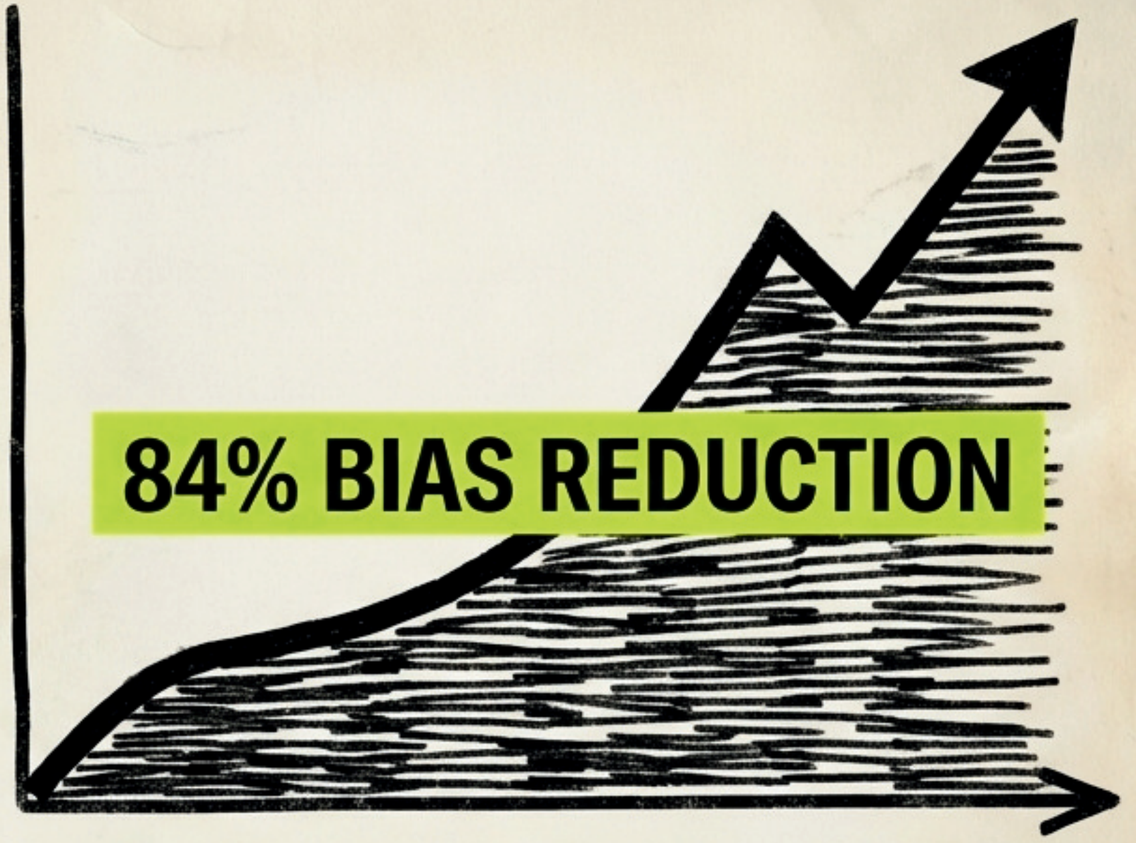
TAKEAWAY: When you wire an AI into a daily workflow, you outsource your organization’s judgment. Humans do not reliably override a confident machine at scale.

The Fallout & The Fix

NEW YORK STATE DEPARTMENT
OF FINANCIAL SERVICES

October 2019 Warning Letter:

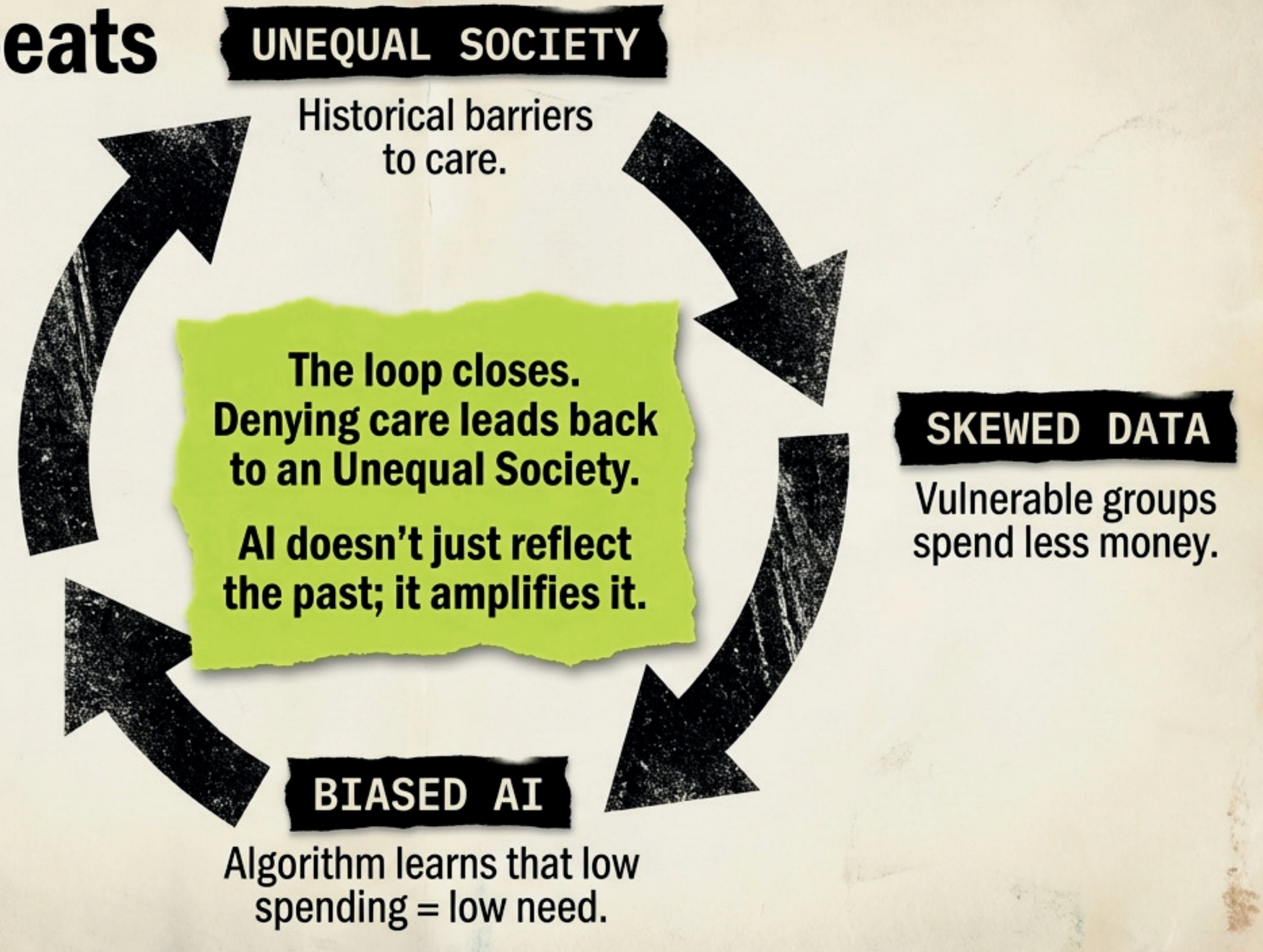
Outcomes are unacceptable and unlawful. The company must prove fairness or immediately stop using the tool.



The researchers proved the bias wasn't permanent. By changing the AI's target from future cost to future avoidable costs and active conditions, they dramatically reduced the racial bias.

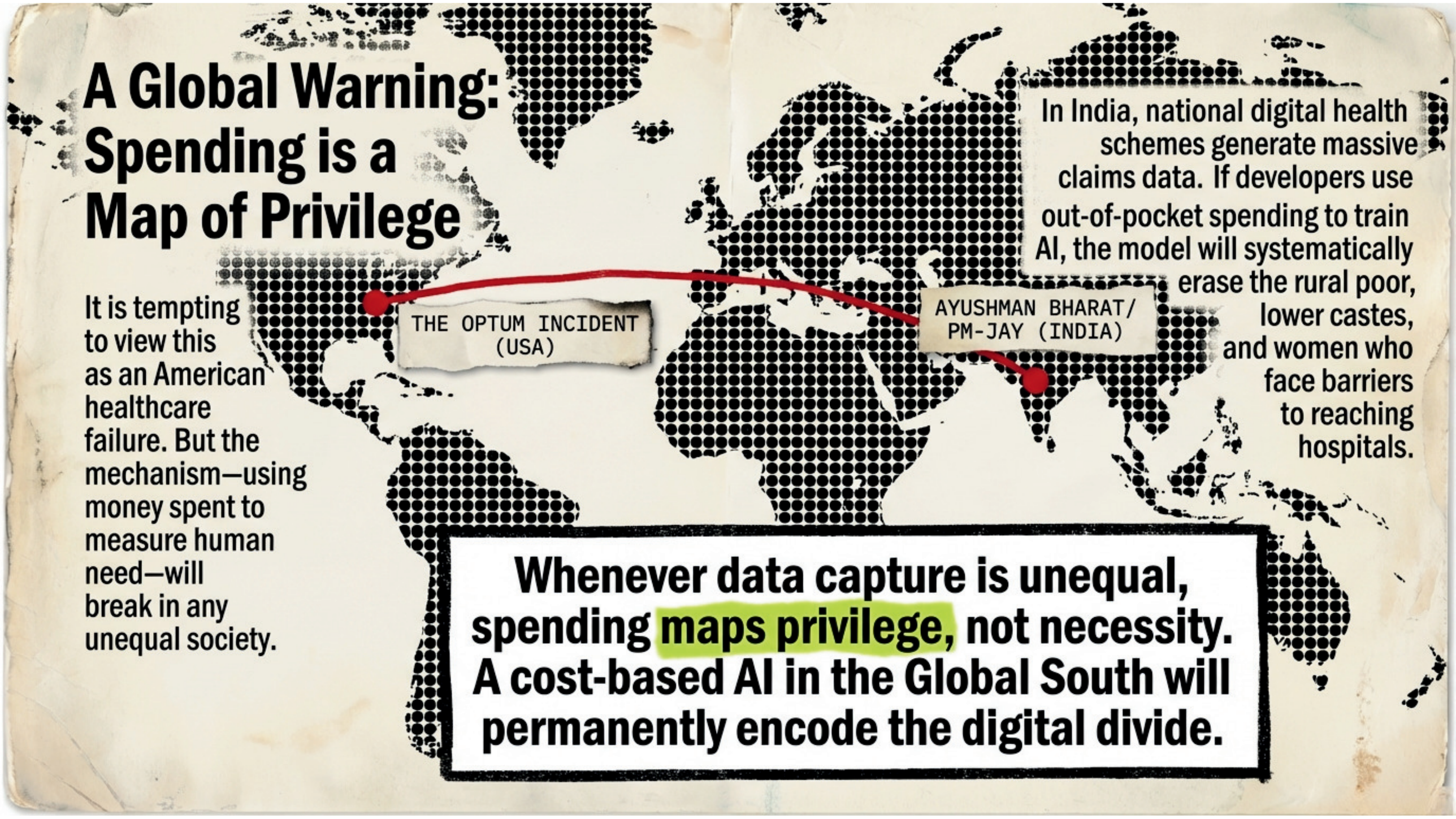
It didn't require magic. It just required changing what the AI was aiming at. But it only happened because an external auditor looked.

How AI Repeats the Past



A Global Warning: Spending is a Map of Privilege

It is tempting to view this as an American healthcare failure. But the mechanism—using money spent to measure human need—will break in any unequal society.



In India, national digital health schemes generate massive claims data. If developers use out-of-pocket spending to train AI, the model will systematically erase the rural poor, lower castes, and women who face barriers to reaching hospitals.

Whenever data capture is unequal, spending maps privilege, not necessity. A cost-based AI in the Global South will permanently encode the digital divide.

The Governance Checklist

AI fairness isn't magic; it's governance. Next time you build or buy an algorithm, demand answers to **these three** questions:

1. INTERROGATE THE STAND-IN.

What exactly is the label measuring? Is it measuring the actual goal (health), or just a convenient shortcut (money)?

2. IDENTIFY THE GHOSTS.

Who is systematically under-measured by this shortcut? Who is missing from the data because they couldn't afford access?

3. MANDATE THE AUDIT.

Commit to running and publishing a fairness audit before the software goes live. Test for accuracy across different demographic groups, not just the majority.

Govern what 'good' means before you optimize for it. Don't wait for a scandal to force your hand.

Follow our work on GlobalSouth.ai