Bias Without a Blame: How the New AI Harm Reports Whitewash Power

Weekly Analysis — https://ainews.social

There is a particular kind of document that has become the signature literary form of our moment: the AI harm report. It arrives with charts and taxonomies, with sober language about "disparate impact" and "algorithmic fairness," with a bibliography and a set of recommendations. It documents, sometimes with real rigor, how a system discriminated—against darker-skinned applicants, against women, against the poor. And then, at the decisive moment, it goes quiet. It names the harm but not the firm. It describes the mechanism but not the beneficiary. It tells you that bias occurred, in the passive voice, as though bias were weather. This is the genre I want to examine, because its silences are not accidental. They are the whole point.

Read across a single week of this discourse and a pattern hardens into something like law. The most consequential arena in which AI reshapes ordinary lives—the workplace, where systems screen résumés, rank workers, monitor bodies, and select people for termination—is under-theorized precisely because it implicates named companies with lawyers. Meanwhile, the safe cases circulate freely: bias in medical imaging, bias in marketing personalization, bias one can discuss without anyone getting sued. The dominance of "ethical failure" as a topic—by the internal accounting of this week's corpus, nearly two of every five findings concern ethics gone wrong—looks at first like vigilance. Look closer and it reads as displacement. We have built an entire discursive economy for describing harm, and almost none of it for assigning responsibility. As Kate Crawford observes, tech companies "rarely suffer serious financial penalties when their AI systems violate the law and even fewer consequences when their ethical principles are violated" [17]. The report is what fills the space where the penalty should be.

[17] The Atlas of AI - Power, Politics, and the Planetary Costs

This essay is about who holds power in the discourse about AI harm, and who does not. My claim is that the reports themselves have become an instrument of power—not by lying, but by describing accurately while pointing nowhere. Anxiety gets managed. The field gets to feel critical. And the systems keep grinding, because the one sentence that would connect the documented harm to the actor profiting from it is the sentence that never gets written.

The Audit as Ritual Reassurance

Begin with the audit, the governance principle, the fairness framework—the whole apparatus of self-regulation that the AI industry has produced with astonishing fluency. On the surface these are acts of conscience. Structurally they are something closer to indulgences: purchased comfort that leaves the underlying arrangement intact. The tell is the missing political economy. An audit that measures a hiring tool's false-rejection rate by race, but does not ask who sells the tool, who profits from its adoption, and who is insulated from the consequences of its errors, has performed a technical exercise dressed as a moral one.

Consider the case that should have organized the entire week's coverage: the discrimination suit against Workday, whose applicant-screening software is alleged to have systematically rejected older, Black, and disabled candidates. What makes the case a genuine landmark is not the bias itself—bias in hiring predates software by millennia—but the legal question of who is liable when an algorithm does the discriminating. The suit advances the argument that a vendor supplying the screening tool can be treated as an "agent" of the employers using it, which would puncture the convenient fiction that the software company merely provides a neutral instrument [19]. This is the rarest thing in the discourse: a mechanism for attaching a name to a harm. And notice how much of the surrounding commentary works to blunt it, reframing the story as a "wake-up call for HR"—a problem for human-resources departments to manage—rather than as what it is, a claim that a specific corporation built and sold a discriminating machine.

[19] The Workday AI Lawsuit Is a Wake-Up Call for HR

The broader documentation of automated hiring is unambiguous about the harm and studiously vague about the culprits. Investigations have found that automated hiring tools are spreading discrimination and secrecy simultaneously, so that a rejected applicant cannot even learn what rule condemned them [21]. Research out of Stanford has shown that large language models used in screening can yield racial bias and systemic rejection, reproducing patterns of exclusion at a scale and speed no human panel could match [2]. The findings are solid. What is missing, over and over, is the follow-through: the sentence that says this vendor sold this system to these employers and made this much money doing it.

[21] Will AI give you the job? Automated hiring tools spark discrimination and secrecy lawsuits

[2] AI Hiring Tools Can Yield Racial Bias and Systemic Rejection

The evasion has an intellectual pedigree, and it is worth naming plainly. When Ruha Benjamin describes what she calls the New Jim Code—the way ostensibly neutral technical systems encode and launder existing hierarchy—she is insisting that the discrimination is not a

bug in an otherwise fair machine but a feature of the social order the machine was built to serve [17]. An audit that treats bias as a technical defect to be patched implicitly denies this. It promises that with enough retraining data and enough fairness constraints, the tool can be cleansed and the arrangement preserved. That promise is the reassurance the industry is buying. It converts a question about power—who gets to build the machines that sort us, and who has to live with being sorted—into a question about engineering, which the engineers will of course solve, in time, if we are patient. The governance document is the receipt for that patience.

[17] Race After Technology - Abolitionist Tools for the New Jim

The Arena the Civil-Rights Lens Won't Enter

If you wanted to know where AI is doing the most to redistribute power downward, you would go to work. Not to the hospital or the ad exchange, the settings where harm is discussed most comfortably, but to the ordinary workplace, where AI now decides who is hired, how they are watched, how they are scored, and who is cut. This is the arena the discourse most consistently refuses to center, and the refusal is instructive, because the workplace is where the abstraction "algorithmic bias" becomes the concrete fact of a paycheck withheld.

The tools of workplace surveillance have grown baroque. Employee monitoring and performance review software now tracks keystrokes, screen time, movement, tone, and "sentiment," feeding managers a continuous behavioral dossier and raising legal and ethical risks that most firms have not begun to reckon with [9]. This is not bias in the narrow, correctable sense the audits prefer. It is a wholesale shift in the balance of power between employer and employed, a transfer of visibility and therefore of leverage. Analysts tracking labor have argued directly that AI surveillance is being used to push down the value of labor itself—to intensify work, suppress bargaining power, and extract more for less [3]. A civil-rights framework that catalogs discrimination in the abstract but never crosses the threshold of the workplace has excluded the site where AI most directly determines whether a person can earn a living.

[9] How A.I. Is Changing Employee Monitoring and Performance Reviews

[3] AI surveillance is further exploiting U.S. labor. Here's how to reclaim ...

The clearest recent illustration is the allegation that Meta used AI to tag workers who had taken medical or family leave, flagging them for inclusion in mass layoffs [14]. Sit with the specificity of that. This is not a hypothetical about a hypothetical model producing a statistically disparate outcome. It is a named corporation accused of building a system to identify people who exercised a legal right and then converting that identification into termination. And the

[14] Meta used AI to tag workers who took leave to be laid off, lawsuit ...

same company faces separate allegations that its AI-equipped smart glasses were used to watch and record workers despite promises to the contrary [13]. Here is harm with a face, a defendant, a paper trail. Yet in the aggregate discourse these stories do not organize the analysis; they float as isolated scandals, each absorbed and forgotten, none permitted to indict the model of the automated workplace as such.

[13] Meta promised it wouldn't spy on you with its AI smart ... - Fortune

Even the commentary that acknowledges the workplace often works to launder it. A representative framing insists that "AI didn't break hiring—it scaled the bias we already chose," which sounds like candor and functions as absolution [1]. The argument is half true and therefore especially useful: yes, human hiring was biased, and yes, the model learned from it. But "we already chose" is a remarkable act of diffusion. Who is "we"? The applicant screened out by a résumé filter did not choose it. The vendor that built the filter and the employer that bought it did, and they are the parties the sentence quietly dissolves into a universal human failing. This is the move that recurs at every scale of the discourse: locate the harm in a fog of collective inheritance so thick that no particular hand can be seen holding the knife. Diffusion of responsibility is not a byproduct of the analysis. It is the service the analysis provides.

[1] AI Hiring Bias: The Workplace Problem AI Didn't Create

The Avoided Evidence: Welfare and the Machinery of Suspicion

If the workplace is the excluded arena, welfare is the avoided evidence—the domain where automated systems have inflicted the most documented, most severe, most clearly attributable harm, and which the mainstream ethics discourse touches with tongs, if at all. The pattern here is not vagueness but a kind of selective focus: the cases are covered, sometimes brilliantly, in the investigative press, and then quarantined from the general conversation about "AI ethics," as though what happens to benefit claimants were a separate subject from what happens to the rest of us.

The mechanism is chillingly consistent across borders. Welfare systems increasingly deploy AI not to deliver benefits but to hunt for fraud—to flag "anomalous" patterns and cut people off. Crawford describes this directly: welfare decision-making systems "are used to track anomalous data patterns in order to cut people off from unemployment benefits and accuse them of fraud" [17]. One investigation documented a scoring algorithm that could, with a single opaque risk number, upend a person's life—subjecting them to invasive fraud investigations on the strength of a correlation they could neither see nor contest [20]. Across Europe, analysts have found that AI-driven

[17] The Atlas of AI - Power, Politics, and the Planetary Costs

[20] This Algorithm Could Ruin Your Life - WIRED

welfare systems are deepening inequality and poverty, concentrating suspicion on the already-precarious and building automated distrust into the architecture of the safety net [10].

[10] How AI-driven welfare systems are deepening inequality and poverty ...

What makes welfare the definitive test case is a story that ended in honest failure. Amsterdam set out to build a genuinely fair welfare-fraud algorithm—not to cut corners, but to do it right, with consultation and bias-testing and every good-faith mitigation the fairness literature recommends. The project failed anyway; the "fair" system, painstakingly engineered, could not escape producing discriminatory outcomes and was shut down [12]. This is the single most important result in the whole field, because it detonates the premise on which the entire audit-and-governance apparatus rests. If a well-resourced, well-intentioned government cannot build a fair fraud-detection system even when it sincerely tries, then the problem was never a shortage of fairness engineering. The problem is the purpose: a system built to find reasons to deny people money will find them, and it will find them disproportionately among the people already treated with suspicion. Virginia Eubanks made exactly this argument in her study of America's automated safety net—that these systems function as tools of moral diagnosis and management aimed at the poor, digital descendants of the poorhouse [17]. Amsterdam confirmed it in a controlled experiment. And still the general discourse prefers to talk about bias in the abstract, because the welfare case names not a rogue vendor but the state itself, deploying automation against its most vulnerable citizens as a matter of policy.

[12] Inside Amsterdam's high-stakes experiment to create fair welfare AI

[17] Automating Inequality

Policing completes the picture, and here the beneficiaries have names the discourse works hard not to say. Investigators obtained the code behind Flock Safety's new AI tool for police, revealing an "investigate" function that stitches license-plate reads and other feeds into a searchable dossier on ordinary people's movements—surveillance infrastructure sold to departments across the country by a private, profit-seeking firm [8]. Immigration enforcement has become a showcase for the same logic: reporting has documented ICE and the Department of Homeland Security deploying mass surveillance against not only immigrants but American citizens and activists [11], with internal documents showing the surveillance apparatus set to surge further in the year ahead [7]. And at the center of much of this sits Palantir, described by human-rights advocates as a "black box"—a company whose systems shape consequential decisions about people while remaining almost entirely unaccountable to them [15].

[8] Flock Has a Powerful New AI Tool for Police. We Got Its Code

[11] ICE is using mass surveillance on American citizens, activists : NPR

[7] DHS-built surveillance apparatus to surge in year ahead, documents show ...

[15] Palantir es como una caja negra: la advertencia de ... - Clarín

Here, at last, are the actors. Flock, Palantir, the contracting firms that build DHS's tools—these are not diffuse inheritances of collective bias but specific enterprises earning specific revenue from surveil-

lance. Which is precisely why they appear so rarely in the ethics literature's central frame and so reliably in the investigative reporting that the ethics literature declines to fully absorb. The safe examples—diagnostic bias, recommendation-engine skew—protect the field's comfort because they can be discussed as puzzles. The welfare and policing cases cannot, because they name the powerful and describe them profiting. The choice of which examples to foreground is itself an exercise of power, and it runs, almost without exception, toward the harms that embarrass no one who matters.

Who Speaks, and From How Far Away

Follow the power in the discourse and you find it flowing along a familiar geography. The people who theorize AI harm sit, overwhelmingly, in the wealthy centers where the models are built and financed. The people who absorb the harm sit at the peripheries—in the workplaces being surveilled, the neighborhoods being policed, the poor households being scored, and, further out still, in the Global South labor markets and extraction zones on which the entire enterprise quietly rests. The reports describe the harm; they less often let the harmed speak, and almost never in a register that would require the industry to change.

The clearest structural silence is the labor that produces AI itself. The systems whose "intelligence" we debate are trained and maintained by armies of low-paid data workers—labelers, moderators, annotators—concentrated in the Global South, and there is a growing case that the future of this work must be reimagined so that its value flows to the people performing it rather than only to the firms extracting it [18]. Investigations into that hidden workforce have uncovered exploitation of a familiar colonial shape: traumatic content moderation for wages that would be illegal in the countries where the resulting product is sold, an entire substrate of human labor rendered invisible so that the software can appear autonomous [16]. The mystification here is the deepest of all: the very word "artificial" performs the erasure, presenting as machine intelligence what is in substantial part underpaid human judgment. Meredith Broussard's term for our tendency to overstate what machines do on their own—technochauvinism—captures the ideological work being done [17]. When we call the output "artificial intelligence," we write the workers out of their own product.

[18] Reimagining the future of data and AI labor in the Global South

[16] Q&A: Uncovering the labor exploitation that powers AI

[17] Artificial Unintelligence - How Computers Misunderstand

The extraction runs to the earth as well, and it lands, predictably, on the least powerful. The physical footprint of AI—the data cen-

ters, their prodigious appetite for land, water, and power—is not distributed evenly. Reporting has documented how the water and land impacts of the AI boom concentrate in particular communities, imposing real costs on the places nearest the infrastructure [6], and investigators have shown that Black communities in particular face heightened pollution driven by the demand for AI compute [5]. This is environmental racism with a server rack, and it is exactly the kind of harm the abstract fairness discourse cannot metabolize, because there is no bias metric for a poisoned aquifer and no retraining run that will move a data center away from the neighborhood it was sited to burden.

[6] Data Drain: The Land and Water Impacts of the AI Boom

[5] Black Communities Face More Pollution Due to Demand for AI

What unites these silences is a question the ethics literature is structurally reluctant to ask: who is in the room. The MIT primer on AI ethics puts the underlying danger with unusual clarity—that when a handful of companies exercise power not only over individuals but over the very infrastructures of democracy, then even their best intentions toward ethical AI erect barriers to it, because concentrated power cannot regulate itself into fairness [17]. The discourse's amplified voices belong disproportionately to those companies and the institutions they fund. The voices structurally absent belong to the surveilled worker, the flagged claimant, the moderator in Nairobi, the family beside the cooling towers. A field that discusses their harms without their participation has not described their situation. It has spoken over them, in a language designed to be heard by everyone except them.

[17] AI Ethics - The MIT Press Essential Knowledge series

The Sentence That Never Comes

Here is the strangest feature of the sophisticated end of this literature, and the one that finally gives the game away. The best researchers do name the deep problem. They will write, in so many words, that AI systems function to legitimize unequal power structures—that their real social effect is to make existing hierarchy look natural, efficient, even fair. This is the correct analysis. And then the sentence ends. It names the function without naming the actors performing it and profiting from it. It says "unequal power structures are being legitimized" and declines to say by whom, for whose benefit, at whose expense. The analysis stops exactly one sentence short of use.

That final sentence is not hard to write. It would read: this welfare algorithm was built by this contractor, deployed by this agency, and it cut off these people while these officials claimed savings. Or: this hiring tool was sold by this vendor to these employers, and it rejected

these applicants while everyone involved disclaimed responsibility. The evidence to write such sentences exists—it is sitting in the investigative reporting cited throughout this essay, in the Workday filings, the Meta suits, the Flock code, the Amsterdam post-mortem. What is missing is the willingness to move from documentation to attribution, and that missing willingness is not an oversight. It is the discipline of a discourse that wants to remain employable, fundable, and welcome at the conference. Naming the harm earns you a reputation for rigor. Naming the beneficiary earns you a lawyer's letter.

This is why the dominance of "ethical failure" as a category should worry us rather than reassure us. A discourse overwhelmingly organized around cataloging failures—nearly two in five of this week's findings—looks like a field doing its job. But a catalog of failures with no defendants is not accountability; it is its simulation. It performs the emotional function of critique, discharging our anxiety about these systems, while performing none of critique's political function, which is to change the distribution of power. Crawford's observation bears repeating because it is the hinge of the whole matter: companies rarely face penalties when their systems break the law and almost never when they violate their own stated principles [17]. If violating your ethics principles carries no cost, then the principles are not constraints. They are marketing. And the harm report that describes the violation without naming the violator is not the antidote to that marketing. It is part of it—the somber, credible, third-party-validated part that proves the industry takes ethics seriously precisely by producing so much text about it.

[17] The Atlas of AI - Power, Politics, and the Planetary Costs

There is a version of this discourse that would matter, and it is not utopian. It exists in fragments—in the reporting that obtained Flock's code, in the suit that dares to call Workday an agent of the employers it serves, in the Amsterdam experiment that had the integrity to publish its own failure, in the labor researchers who insist the moderators be named and paid. What these have in common is that they close the sentence. They connect a documented harm to an accountable actor. Even the modest, local efforts—free AI classes helping veterans retrain for a labor market being reshaped over their heads [4]—carry an implicit indictment the abstract literature suppresses: that the disruption is being managed downward, its costs pushed onto the individuals least able to refuse them while the firms driving it are asked to change nothing.

[4] Augusta nonprofits host free AI classes to help veterans advance their careers

The task, then, is not more documentation. We are drowning in documentation. The task is to refuse the passive voice—to treat every report that describes bias without naming its beneficiary as an unfinished document, and to finish it. Benjamin's insight is that technology

encodes the social order it was built to serve [17]; the corollary is that someone built it and someone benefits, and the discourse that will not say who has chosen, whatever its intentions, to protect them. The old way of legitimizing unequal power was silence—you simply did not discuss what the systems did to the powerless. The new way is the report: exhaustive, rigorous, humane, and structurally incapable of naming a defendant. Anxiety is managed. Power is untouched. The systems keep grinding. And the most damning thing that can be said about the harm report as a genre is that the machines it describes could not ask for a better public relations strategy than a critique that documents everything and blames no one.

[17] Race After Technology - Abolitionist Tools for the New Jim

References

1. AI Hiring Bias: The Workplace Problem AI Didn't Create
2. AI Hiring Tools Can Yield Racial Bias and Systemic Rejection
3. AI surveillance is further exploiting U.S. labor. Here's how to reclaim ...
4. Augusta nonprofits host free AI classes to help veterans advance their careers
5. Black Communities Face More Pollution Due to Demand for AI
6. Data Drain: The Land and Water Impacts of the AI Boom
7. DHS-built surveillance apparatus to surge in year ahead, documents show ...
8. Flock Has a Powerful New AI Tool for Police. We Got Its Code
9. How A.I. Is Changing Employee Monitoring and Performance Reviews
10. How AI-driven welfare systems are deepening inequality and poverty ...
11. ICE is using mass surveillance on American citizens, activists : NPR
12. Inside Amsterdam's high-stakes experiment to create fair welfare AI
13. Meta promised it wouldn't spy on you with its AI smart ... - Fortune
14. Meta used AI to tag workers who took leave to be laid off, lawsuit ...

15. Palantir es como una caja negra: la advertencia de ... - Clarín

16. Q&A: Uncovering the labor exploitation that powers AI

17. Race After Technology - Abolitionist Tools for the New Jim

18. Reimagining the future of data and AI labor in the Global South

19. The Workday AI Lawsuit Is a Wake-Up Call for HR

20. This Algorithm Could Ruin Your Life - WIRED

21. Will AI give you the job? Automated hiring tools spark discrimination and secrecy lawsuits