

Auditing the Output, Absolving the Machine: When Injustice Becomes a Metrics Problem

Weekly Analysis — <https://ainews.social>

Watch the hands, not the mouth. This week's discourse on artificial intelligence and society performs a magic trick so practiced that it no longer looks like sleight of hand — it looks like responsibility. An institution acknowledges that its systems discriminate. It commissions an audit, computes a ratio, convenes a working group, publishes a call for further research. And then, having named the harm with real precision, it leaves untouched every structural condition that produced the harm in the first place. The acknowledgment is not a prelude to change; it is a substitute for it. The confession becomes the alibi.

This is the move worth learning to see, because it is everywhere once you notice it. When a woman is jailed for five months on the strength of a false facial-recognition match, the resulting coverage frames the disaster as a chain of human errors around an otherwise improving tool [27]. When human-resources teams discover that their hiring algorithms filter out women or older applicants, the remedy is to measure the disparate-impact ratio and adjust the inputs [26]. In each case the injury is converted into a quantity, and once it is a quantity it becomes a management problem — trackable, reportable, and, above all, ownable by the same institutions that deployed the system. What Ruha Benjamin calls the "allure of objectivity" does its quiet work here: automation is offered as the cure for human prejudice even as it launders that prejudice into infrastructure [21].

The argument of this essay is that the language of fairness is being captured by the language of measurement, and that this capture is not neutral. It redistributes power. By rendering discrimination as a ratio to be corrected, institutions demonstrate concern without disturbing the arrangements that generate the discrimination — the compute buildout narrated as inevitability, the platform expansion narrated as discovery, the displaced worker narrated as a person who simply failed to reskill. The public grows fluent about bias and goes silent about power. To watch the week honestly is to ask, at every turn, a question the metrics are designed to suppress: not *how much* did the system discriminate, but *who* built it, *who* profits from it, and *who* is left

[27] Woman Wrongfully Jailed 5 Months After AI Facial Recognition Error - Forbes

[26] Will AI give you the job? Automated hiring tools spark discrimination ...

[21] Race After Technology

holding the consequences while the architecture is declared essentially sound.

The Ratio That Reassures

Begin with the most respectable version of the maneuver, the one that wears the uniform of compliance. Across employment law, the response to algorithmic discrimination has converged on a single instrument: the disparate-impact ratio. A Spanish legal analysis of AI-driven personnel selection lays out the emerging governance architecture with clarity — the burden of proof, the duties of auditing, the obligation to demonstrate that a screening tool does not disproportionately exclude protected groups [19]. This is genuine progress in the narrow sense that it gives an excluded candidate a lever. But observe what the frame does to the problem. It presumes the tool belongs in the hiring process and asks only whether its outputs fall within tolerance. The question of whether résumé-ranking machines should mediate access to livelihoods at all is placed permanently out of scope.

The consequences of that scoping are not abstract. Recruitment algorithms have been shown to function, by accident, as something close to medical screening — inferring personality traits and health-adjacent characteristics from proxies the applicant never consented to disclose [4]. French analyses of algorithmic bias in recruitment describe how these systems absorb the historical composition of a workforce and reproduce it as a prediction about merit, so that yesterday’s exclusions become tomorrow’s qualifications [3]. The disparate-impact ratio can catch some of this. What it cannot do is name the underlying logic — that a machine trained on a discriminatory past will, absent extraordinary intervention, encode discrimination as objectivity. Benjamin’s phrase for this is the New Jim Code: default settings, once fashioned, “take on a life of their own,” projecting an allure of neutrality “that makes it difficult to hold anyone accountable” [21]. The audit measures the symptom and, in measuring it, certifies that the underlying machine is fixable — which is to say, worth keeping.

The welfare state offers the same structure at higher stakes. The investigative work assembled around the “Suspicion Machine” — Rotterdam’s system for scoring benefits recipients by their likelihood of fraud — is among the most rigorous documentation we have of algorithmic harm, tracing exactly how variables like gender, language, and neighborhood pushed vulnerable people up a suspicion ranking [12]. The reporting is unsparing about the machinery [24], and it demon-

[19] Selección laboral mediante inteligencia artificial: discriminación, prueba y deberes de gobernanza

[4] Biais de l’IA de recrutement : quand les algorithmes de personnalité ...

[3] Biais algorithmique et discrimination : ce que l’IA change au recrutement

[21] Race After Technology

[12] Inside the Suspicion Machine - Pulitzer Center

[24] This Algorithm Could Ruin Your Life - WIRED

strates that the discrimination was measurable — that is the whole force of the exposé [13]. And here the trap closes most tightly. Amsterdam, having absorbed the lesson, set out to build a *fair* welfare AI: to de-bias the model, balance the training data, and deploy a version of the suspicion machine that would not discriminate. It failed [11]. The city did everything the metrics-first paradigm asks — and still could not produce an equitable machine, because the inequity was never a property of the model’s calibration. It was a property of the decision to build a suspicion machine at all, aimed downward at the poor rather than upward at, say, the corporations whose tax behavior costs the state incomparably more.

That is the tell. When the most conscientious, best-resourced attempt to correct algorithmic bias by adjusting the algorithm ends in documented failure, the failure is not technical. It is a refutation of the premise that fairness is a tuning problem. Yet the premise survives, because it is enormously convenient to the institutions that hold it — it converts a question about power into a question about parameters.

Policing the Machine With the Machine

If discrimination can be reframed as a metric, then oversight can be reframed as a feature. The clearest expression of this recursion is the emerging practice of using artificial intelligence to audit artificial intelligence — a governance model in which the answer to a failing system is another system [25]. The elegance of the move, from the vendor’s point of view, is that it keeps the entire question inside the technical domain. Oversight becomes a product category rather than a democratic act. Nobody outside the loop needs to be consulted, because the machine is checking the machine, and the ratios all come back in range.

Follow this logic into policing, where the human cost is most legible. More than a dozen wrongful arrests have now been documented as arising from police reliance on facial recognition, a pattern the American Civil Liberties Union has tracked case by case [28]. The individual stories are harrowing precisely because the technology performs as designed — it returns a candidate match, and the institutional apparatus around it treats the match as probable cause [27]. Notice how the Forbes framing insists the failure was “entirely human,” as though isolating the human decision-point absolves the system. This is the metrics move inverted: when the tool errs, the error is attributed to the operator; when the tool is defended, its objectivity is invoked.

[13] Inside the Suspicion Machine - WIRED

[11] Inside Amsterdam’s high-stakes experiment to create fair welfare AI

[25] We’re Now Relying on AI to Police AI

[28] Wrongful Arrests Pile Up Due to Facial Recognition Technology

[27] Woman Wrongfully Jailed 5 Months After AI Facial Recognition Error - Forbes

Either way the machine escapes.

The forward march continues regardless. Municipalities are adopting AI-assisted police call systems as ordinary procurement, the technology arriving through the mundane channel of a city budget rather than a public debate [5]. Meanwhile the surveillance layer thickens: automated camera networks now flag "suspicious" movement patterns and report civilians to police on their own initiative, a capability the ACLU has flagged as a categorical expansion of warrantless tracking [20]. Litigation is beginning to argue that this warrantless deployment is unconstitutional [15], and governments have already turned these tools on political assembly, using AI to identify participants in protest [22]. Here the auditing frame reveals its limit with brutal clarity. You cannot compute a disparate-impact ratio for the chilling of dissent. The harm is not a misclassification rate; it is the redistribution of power between the watched and the watching. Kate Crawford's observation cuts to it: classification is "an act of power," and once embedded in working infrastructure it becomes "relatively invisible without losing any of its power" [21]. The audit renders the classification visible enough to reassure and never visible enough to threaten.

The ACLU's own 2025 report on law enforcement in the age of AI insists on this larger frame, documenting how carceral systems adopt predictive and recognition tools faster than any accountability regime can constrain them [17]. What the report makes plain is that the documentation of harm has become a genre unto itself — thorough, sophisticated, damning — running on a parallel track to the deployment it describes, never intersecting it. We are producing excellent maps of a territory we are simultaneously paving over.

The Fact That Cannot Be Questioned

Every ideology needs a fact that stands outside argument, a piece of the world declared so obviously true that to question it marks you as naïve. In this week's AI discourse that fact is growth — the assertion, endlessly recirculated in official consultations and executive keynotes, that compute is doubling on a schedule measured in weeks and that the only rational response is to run faster. The figure functions less as a measurement than as a natural law, and its rhetorical job is to move the buildout of data centers, the scraping of the commons, and the concentration of capacity in a handful of firms from the category of *choices* into the category of *weather*. You do not argue with weather. You adapt to it.

The genius of naturalization is that it makes the political invisible

[5] Buena Park Eyes AI-Assisted Police Calls

[20] Surveillance Company Flock Now Using AI to Report Us to Police if it

...

[15] Lawsuit Argues Warrantless Use of Flock Surveillance Cameras Is ...

[22] The Government Uses AI to Identify Protesters: How It Works

[21] The Atlas of AI

[17] PDF 2025 Report Law Enforcement in the Age of AI - aclu-co.org

by making it inevitable. Once expansion is a fact of nature, the expansion of platforms into new markets can be narrated not as conquest but as service — as *discovery*, as *access*, as *inclusion*. The scholarship on digital colonialism has been tracking exactly this rebranding. Analyses of algorithmic dependence describe how AI infrastructures are exported to the Global South on terms that deepen reliance on the exporting firms, reproducing older colonial relations in software [2]. The London School of Economics framing of "AI colonialism" names the structure precisely — data and value flow from periphery to center while the vocabulary of the transaction insists it is a gift [8]. When a messaging platform inserts a commercial agent into a country's informal economy and calls it a discovery engine, the naming is the whole strategy: it recodes extraction as opportunity and positions the platform as the benefactor of the people whose economic activity it is metabolizing.

Crawford's larger point about consequences is the necessary companion to this. Tech companies, she notes, "rarely suffer serious financial penalties when their AI systems violate the law and even fewer consequences when their ethical principles are violated" [21]. Hold that beside the disparate-impact ratio and the fair-welfare experiment and the AI-auditing-AI product line, and the asymmetry snaps into focus. The entire apparatus of measurement and oversight is aimed at outputs — at the observable behavior of deployed systems — and almost none of it is aimed at the accumulation that produces those systems. We have built an elaborate accountability machinery for the exhaust and left the engine unexamined. This is not an oversight in the system; it is the system. Meredith Broussard and the AI Now Institute she helped bring into the conversation argued years ago for a more balanced public view of these tools, one that treats institutional power as part of the analysis rather than a given [21]. That corrective remains the road not taken. The dominant frame still asks the machine to behave and never asks the owner to justify.

The Workers Who Do Not Appear

Naturalized growth has a labor supply, and it is largely invisible by design. Behind the fluent outputs of every large model sits a workforce whose task is to make the system appear autonomous — annotating data, filtering toxicity, ranking responses, absorbing the psychological cost of the material so the end user never has to. French investigative reporting has documented this substrate plainly: the prowess of AI rests on the exploitation of workers in poor countries, hidden by the interface [6]. The professor Julián Posada's work on a Venezue-

[2] Algorithmic dependence and digital colonialism: a ... - Frontiers

[8] From AI colonialism to co-creation: bridging the global AI divide

[21] The Atlas of AI

[21] Artificial Unintelligence - How Computers Misunderstand

[6] Derrière les prouesses de l'IA, l'exploitation de travailleurs ...

lan workforce that trains algorithms for cents an hour gives this its sharpest name — “platform extractivism,” a regime in which the value produced by human judgment is siphoned upward while the humans themselves are rendered statistically and rhetorically absent [7].

The human toll is not incidental; it is structural. Kenyan content moderators who trained the safety filters of major chatbot systems have described the work in the language of injury — “it’s destroyed me completely” — after months of exposure to the worst material the internet produces, for wages that would be illegal in the countries where the resulting products are sold [14]. Set this against the AI-audits-AI discourse and the disparate-impact ratio, and the selectivity of our concern becomes impossible to ignore. We have developed a rich technical vocabulary for the harms a model might inflict on a user in the Global North and almost no institutional vocabulary at all for the harms inflicted on the workers in the Global South who built the model’s safety in the first place. Their suffering does not produce a metric. It produces a product.

The Brookings framing of data and AI labor in the Global South tries to shift this, insisting that the workers who make these systems function are a constituency with claims, not a cost line to be optimized [18]. But the structural silence is telling. The dimensional pattern in this week’s evidence — the overwhelming dominance of “ethical failure” as the category through which AI-and-society is discussed — reveals what the discourse is trained to see and what it is trained to miss. An ethical failure is an event: a bias detected, an error made, a principle violated. It is discrete, locatable, and correctable. The exploitation of an invisible workforce is not an event; it is a *condition*, a standing arrangement of the global economy, and conditions do not fit the failure frame. So they disappear from it. The very success of “bias” and “ethical failure” as organizing concepts is what keeps the labor question offstage. To document a failure is to imply the system usually works. To document a condition is to indict the system as such — and that is precisely the sentence the metrics are built to avoid.

The exploitation reaches into the domestic economy too, wearing the mask of management. Gig workers face algorithmic systems engineered to extract maximum output at minimum obligation, and the reporting warns that AI stands to generalize this precarity to far more of us [9]. Amazon’s algorithmic management has been weaponized against organizing, deployed to surveil and pressure workers during an anti-union campaign in Alabama [23]. Analysts of AI workplace surveillance describe a system that pushes down the value of labor by rendering every worker continuously legible to the employer and

[7] El profesor Julián Posada habla de la fuerza laboral invisibilizada en Venezuela que entrena algoritmos de IA por centavos la hora y del “extractivismo de plataformas”

[14] It’s destroyed me completely’: Kenyan moderators decry toll of ...

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

[9] Gig workers are endlessly exploited. AI could make more of us share ...

[23] The weaponization of algorithmic management: Lessons from Amazon’s anti ...

legible to no one on their own side [1]. And the ordinary workplace is being reshaped in the same direction, as AI-driven monitoring and performance systems quietly rebuild the terms of employment around continuous profiling [10]. None of these is legible as an “ethical failure,” because none of them is a malfunction. They are the system working exactly as intended, and the fluency of the bias-and-fairness vocabulary is what allows them to proceed without a public reckoning.

The Layoff Reframed as a Lesson

The most durable trick in the repertoire is also the oldest, and it is enjoying a vigorous second life. When AI displaces workers, the displacement is reframed as a personal failure to adapt — an “educational and adaptive” challenge that the worker must meet by reskilling, retraining, and reinventing. Responsibility for a structural decision made in a boardroom is transferred, whole, onto the individual who bears its cost. This is responsabilization in its purest form, and it is remarkable how little the vocabulary has had to change to keep doing its work.

Consider the corporate experiment in substituting AI for staff. The account of how a bold plan to replace much of a workforce with AI came apart is instructive precisely because the plan was made, seriously, at the highest level [16]. What matters is the direction of the arrow. The decision to automate is framed as a rational response to the naturalized fact of AI’s advance — the same “weather” logic that governs the compute discourse — while the workers whose livelihoods are on the table are handed the entire burden of adjustment. When the automation over-reaches and has to be walked back, the firm absorbs a manageable embarrassment. The workers, throughout, are told the responsible thing is to prepare themselves for a future being decided without them.

The pattern generalizes across the labor evidence. The gig economy’s expansion under algorithmic control is narrated as flexibility and entrepreneurship, so that the worker who cannot make the numbers work has failed at hustle rather than been failed by design [9]. Workplace surveillance is sold as a tool for “development” and “coaching,” so that the employee who chafes at continuous monitoring is resisting their own improvement [10]. In every register the move is the same: a structural condition imposed by those with power is re-described as a developmental task assigned to those without it. Alvin Toffler’s old diagnosis of a population overwhelmed by the pace of change gets weaponized here — the anxiety is real, but it is monetized as a reskilling market and moralized as a personal duty, never traced

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

[10] How A.I. Is Changing Employee Monitoring and Performance Reviews | Observer

[16] Mark Zuckerberg tenía un plan audaz para sustituir al personal de Meta por IA; así es como se vino abajo

[9] Gig workers are endlessly exploited. AI could make more of us share ...

[10] How A.I. Is Changing Employee Monitoring and Performance Reviews | Observer

back to the parties who set the pace.

And the reskilling promise deserves the skepticism the discourse withholds from it. Retrain *for what*, exactly, when the same firms driving the displacement are also automating the jobs they gesture toward as safe harbors? The educational framing sustains itself by never quite specifying the destination. It keeps the displaced worker in perpetual motion — always adapting, always upskilling, always one course away from security — which is precisely the condition that keeps labor cheap, mobile, and unable to organize. The layoff becomes a lesson, and the lesson is that the worker, not the system, is the thing in need of correction. This is the same grammar as the disparate-impact ratio, applied to a body rather than a dataset: locate the deficiency in the output, adjust the output, leave the machine that produces the outputs exactly where it is.

Fluent About Bias, Silent About Power

Assemble the week and the shape is unmistakable. In hiring, discrimination is acknowledged and converted into an auditable ratio while the machine keeps its seat at the table [19]. In welfare, the most careful attempt to build a fair suspicion machine fails, and the failure is treated as a technical setback rather than a verdict on the enterprise [11]. In policing, wrongful arrests accumulate while the tools spread through procurement and the harm to political assembly falls entirely outside the metrics that would supposedly govern them [28]. In the Global South, the workers who make the systems function are extracted and erased, their injury producing a product and never a ratio [7]. At home, displacement is reframed as a curriculum the displaced must complete on their own time.

The common thread is a redistribution of attention that mirrors a redistribution of power. The discourse looks hard at outputs and refuses to look at ownership. It grows exquisitely precise about the behavior of deployed models and studiously vague about the accumulation that deploys them. It treats every harm as an event to be corrected and never as a condition to be contested. And the reason is not mysterious. Measurement is the form of concern that leaves the balance of power intact. You can compute a disparate-impact ratio forever without ever asking who gets to build the machine, who profits from its scale, and who is expected to adapt to its consequences. The ratio is real; the fairness it promises is partial; and the promise is load-bearing precisely because it is partial. It concedes enough to look responsible and withholds enough to change nothing.

[19] Selección laboral mediante inteligencia artificial: discriminación, prueba y deberes de gobernanza

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

[28] Wrongful Arrests Pile Up Due to Facial Recognition Technology

[7] El profesor Julián Posada habla de la fuerza laboral invisibilizada en Venezuela que entrena algoritmos de IA por centavos la hora y del "extractivismo de plataformas"

To read the week against its own grain, then, is to keep asking the question the metrics are built to dissolve. When a system is found to discriminate, the honest response is not only *how do we correct the output* but *why does this system exist, who decided it should, and what would it mean to refuse it*. Benjamin’s abolitionist framing insists that the default settings themselves are the politics, not the deviations from them [21]. Crawford’s atlas insists that classification is power and that the absence of consequences for the powerful is the structural fact the ethics discourse is engineered to obscure [21]. Both point past the audit toward the thing the audit is designed to protect. The task for a reader who wants to understand this week rather than be managed by it is to hold the two frames apart and notice which one the institutions reach for — and to remember that fluency about bias, when it arrives unaccompanied by any language about power, is not the beginning of accountability. It is the most sophisticated form of its evasion.

[21] Race After Technology

[21] The Atlas of AI

References

1. AI surveillance is further exploiting U.S. labor. Here’s how to reclaim ...
2. Algorithmic dependence and digital colonialism: a ... - Frontiers
3. Biais algorithmique et discrimination : ce que l’IA change au recrutement
4. Biais de l’IA de recrutement : quand les algorithmes de personnalité ...
5. Buena Park Eyes AI-Assisted Police Calls
6. Derrière les prouesses de l’IA, l’exploitation de travailleurs ...
7. El profesor Julián Posada habla de la fuerza laboral invisibilizada en Venezuela que entrena algoritmos de IA por centavos la hora y del "extractivismo de plataformas"
8. From AI colonialism to co-creation: bridging the global AI divide
9. Gig workers are endlessly exploited. AI could make more of us share ...
10. How A.I. Is Changing Employee Monitoring and Performance Reviews | Observer
11. Inside Amsterdam’s high-stakes experiment to create fair welfare AI

12. Inside the Suspicion Machine - Pulitzer Center
13. Inside the Suspicion Machine - WIRED
14. It's destroyed me completely': Kenyan moderators decry toll of ...
15. Lawsuit Argues Warrantless Use of Flock Surveillance Cameras Is ...
16. Mark Zuckerberg tenía un plan audaz para sustituir al personal de Meta por IA; así es como se vino abajo
17. PDF 2025 Report Law Enforcement in the Age of AI - aclu-co.org
18. Reimagining the future of data and AI labor in the Global South
19. Selección laboral mediante inteligencia artificial: discriminación, prueba y deberes de gobernanza
20. Surveillance Company Flock Now Using AI to Report Us to Police if it ...
21. The Atlas of AI
22. The Government Uses AI to Identify Protesters: How It Works
23. The weaponization of algorithmic management: Lessons from Amazon's anti ...
24. This Algorithm Could Ruin Your Life - WIRED
25. We're Now Relying on AI to Police AI
26. Will AI give you the job? Automated hiring tools spark discrimination ...
27. Woman Wrongfully Jailed 5 Months After AI Facial Recognition Error - Forbes
28. Wrongful Arrests Pile Up Due to Facial Recognition Technology