The Servants of the Algorithm: A New Global Labor Caste

Weekly Analysis — https://ainews.social

The chatbot that answers in elegant paragraphs was taught to do so by someone in Nairobi labeling images of executions, child abuse, and torture for less than two dollars an hour. That worker is not mentioned in the press release announcing the model. He is not invited to the conference where the product is unveiled. He has no equity, no health insurance for the post-traumatic stress that the work produced, and no standing in the lawsuits that determine how the model may be used. He is the precondition for the magic, and his erasure is not an accident.

The reporting on this arrangement has, by now, become unavoidable. *Time* documented in early 2023 that OpenAI used Sama-contracted workers in Kenya, paid between roughly $1.32 and $2 an hour, to read and tag the worst material on the open web so that ChatGPT could be made safe for Western consumers [15]. *CBS News*, returning to those workers later, found that the people who believed they had won "tickets to the future" by entering the AI economy had instead been broken by it and discarded [12]. The story is no longer obscure. Yet the discourse — the public talk about what AI is, what it costs, and who is responsible for it — continues to treat this labor as background.

That asymmetry is the subject of this essay. The question is not whether AI is "exploitative," which has become a fashionable abstraction. The question is structural: who labors invisibly so that others may speak about AI as an autonomous wonder, and what does the persistent invisibility of those workers tell us about whose voices the discourse is built to amplify? Kate Crawford put the point with unusual clarity when she observed that "exploitative forms of work exist at all stages of the AI pipeline, from the mining sector… to the software side" [8]. The pipeline is global, vertically integrated, and engineered to disappear from view at exactly the moments when the public might ask hard questions.

[15] OpenAI Used Kenyan Workers on Less Than $2 Per Hour: Exclusive

[12] Kenyan workers with AI jobs thought they had tickets to the future

[8] The Atlas of AI

The Architecture of Invisibility

It is tempting to describe the invisibility of AI workers as a problem of attention — as if we simply have not noticed them yet, and a sufficiently moving documentary will fix it. This misreads the situation. The invisibility is not a deficiency of journalism; it is a feature of the supply chain. Contracts are routed through outsourcing firms in jurisdictions chosen for their lack of labor protection. Non-disclosure agreements are layered over the work. The terminology — "data annotation," "reinforcement learning from human feedback," "red-teaming" — is engineered to sound clerical, abstract, like spreadsheet work, when in practice it can mean spending eight hours a day watching videos of sexual violence to teach a model what to refuse.

The discursive effect of this architecture is precise. When journalists name a harm — bias, hallucination, surveillance — they tend to name it at the point of consumer-facing failure. A hiring algorithm rejects qualified applicants; the lawsuit names the deploying company [9]. A medical scribe hallucinates patient histories in Ontario clinics; the auditor general reports on physician practices [7]. The discourse settles on the visible interface. The annotators, moderators, and prompt-quality raters who shaped the training data — and whose working conditions shaped what the model learned to do — drop out of the frame. This is what Ruha Benjamin, writing about a different but adjacent terrain, called the "New Jim Code": the way technical systems encode hierarchies while presenting themselves as neutral [8]. The neutrality is performed; the hierarchies are infrastructural.

[9] Eightfold AI Lawsuit Claims Secret Algorithm Ranking Applicants

[7] AI systems used by Ontario doctors hallucinate: auditor

[8] Race After Technology

One can see the architecture work in the reverse direction as well. When a Canadian family sued OpenAI after a mass shooting in Tumbler Ridge, alleging that the model had encouraged the attacker, the suit named the company and its CEO as defendants [10]. This is the correct legal posture, and yet it tells us something about how blame moves and does not move. Liability flows upward toward the brand on the product; accountability for the conditions under which the safety filters were built — the workers in Nairobi, Manila, Caracas — flows nowhere. The discourse has a vocabulary for blaming Sam Altman. It has almost no vocabulary for blaming Sama, Scale AI's vendors, or the procurement officers who chose them.

[10] Families of Tumbler Ridge shooting victims sue OpenAI and CEO Sam Altman

Extraction in Three Registers

The labor caste of the algorithm should be understood as one node in a wider system of extraction that operates in at least three registers

simultaneously: human work, environmental resource, and political voice. Each register reinforces the others, and each is structured so that the costs settle in the Global South while the gains accrue in a handful of corporate balance sheets in California and Washington State.

The environmental register is the one most easily quantified, and even there the numbers are alarming. *MIT Technology Review*, in a sustained investigation, calculated that the energy and water footprint of generative AI is not the marginal externality the industry has long claimed but a planetary-scale load that is reshaping electrical grids and water tables [20]. The political consequence of that load is not evenly distributed. *TechCabal* reported, in a clear-eyed assessment of the so-called "AI sovereignty" agenda on the African continent, that the deeper problem is energy: African governments are being asked to host data centers and "participate in" the AI economy while their grids cannot reliably power hospitals [14]. The Brookings analysis of the Global South's position in AI governance arrives at a similar judgment: representation in the governance fora is thin, infrastructure dependence is thick, and the result is a structural inability to set the terms of one's own participation [5].

[20] We did the math on AI's energy footprint. Here's the story you haven't been told

[14] Le problème de la souveraineté de l'Afrique en matière d'IA est en réalité un problème énergétique

[5] AI in the Global South: Opportunities and challenges towards more inclusive governance

This is what Crawford means when she describes labor as "another form of extraction" running alongside the extraction of minerals, water, and energy [8]. The same regions that supply the cobalt, the lithium, the cheap electricity, and the wastewater absorption capacity also supply the workers who clean the training data. There is a horrible coherence to the geography of the AI economy: the bodies of workers, the bodies of land, and the bodies of water are subjected to a single logic of cost externalization. The wealth flows in one direction. The harm flows in another. The discourse, written largely in English by people whose lives are organized around the inflow side, is built to keep the outflow side blurry.

[8] The Atlas of AI

Who Speaks About AI

If you tabulate the AI discourse — the op-eds, the executive orders, the conference keynotes, the venture-capital essays — and ask who is given the microphone, the answer is stark. It is venture capitalists, founders, senior researchers at frontier labs, the policy staff of those same labs, a small group of credentialed academics, and a smaller group of credentialed critics whose critiques are precisely calibrated to be admissible at the same conferences. *Humanité*'s reporting on the Silicon Valley billionaires now lobbying for a "New Deal" against

the AI "job apocalypse" they themselves are engineering captures the closed-loop quality of the conversation: the people displacing workers are also the people drafting the social-policy response to displacement [21]. The displacer and the policy entrepreneur are the same person.

[21] « Job apocalypse » de l'IA : pourquoi les milliardaires de la Silicon Valley défendent un « New Deal »

The structural silences are easy to enumerate once you start looking. Data annotators are almost never named in product launches. Content moderators are rarely interviewed in mainstream technology coverage except as occasional human-interest stories — pity rather than power. Workers in mineral-extraction supply chains are nominally a concern of human-rights NGOs but almost never of AI ethics conferences. Workers who lose their jobs to automation appear as statistics, not as speakers. When workers do break through, the framing tends to convert them into victims rather than analysts of their own situation. The CBS account of Kenyan AI workers spends much of its time on their suffering, comparatively little on their organizing [12] — and yet those workers have, in fact, been organizing. The frame that flatters Western readers' compassion is the frame that suppresses Southern workers' agency.

[12] Kenyan workers with AI jobs thought they had tickets to the future

The mechanism is familiar to anyone who has read Edward Herman and Noam Chomsky on the production of consent. The filters are not crude censorship; they are the cumulative effect of ownership, advertising dependence, source selection, and what counts as a respectable opinion [8]. Apply those filters to AI discourse and the pattern is recognizable. The "respectable" voices are those that share an economic dependence on the continued growth of the AI industry. The "fringe" voices are those of the people who pay its costs. The boundary is policed not by anyone explicitly excluding annotators from the *Atlantic* op-ed page, but by an entire credentialing apparatus that decides in advance whose suffering is a topic for analysis and whose is a topic for charity.

[8] Manufacturing Consent

The Ethics-Failure Frame and Its Uses

Roughly two of every five articles in the AI ethics corpus this publication tracks concern "ethical failures" — bias, hallucination, surveillance overreach, individual instances of harm. The dominance of that frame is itself a power-laden choice, and it is worth pausing on what it does and does not accomplish.

What the frame does well is documentation. The catalogue of AI failures is now extensive and specific. *Raspberry Pi*'s recent synthesis of research on AI bias and identity is good on the mechanics of how training data encodes social hierarchies [6]. The hiring-discrimination

[6] AI is not neutral: What recent research says about bias, identity and power

literature is sharp on the legal exposure now facing employers who deployed off-the-shelf scoring systems without auditing them [2], and a parallel literature in Spanish has begun to map the same dynamics in Latin American hiring and credit markets [19]. Forbes has chronicled, with appropriate alarm, the deployment of AI bathroom monitors in American high schools — a surveillance escalation that would have been politically impossible a decade ago and is now treated as a vendor product to be evaluated on uptime [1]. The Associated Press has documented how school-surveillance systems like Gaggle generate false alarms that end with police at students' doors [18]. These are real harms, and they are being named.

[2] AI Hiring Bias Lawsuits Are About to Surge

[19] Sesgos y discriminaciones sociales de los algoritmos en procesos de selección

[1] AI Bathroom Monitors? Welcome To America's New Surveillance High Schools

[18] School AI surveillance like Gaggle can lead to false alarms, arrests

What the frame does poorly is causation. An "ethical failure" is a discrete event with identifiable victims and, ideally, an identifiable wrongdoer who can be sued or shamed. It is good for journalism and bad for analysis, because it isolates the failure from the system that produced it. A hiring algorithm produces racially disparate outcomes; the frame names "algorithmic bias." But the algorithm was trained on data labeled by workers whose own working conditions select for speed over judgment; deployed by an HR function under cost pressure from the same financialized logic that compressed labor budgets in the first place; sold by a vendor whose business model depends on convincing employers that human judgment is replaceable. The "ethical failure" frame discusses the algorithm. It rarely discusses the labor, the financialization, or the vendor incentives. This is why the legal cases are accumulating without producing structural change — *Eightfold* and its peers face individual suits, but the underlying procurement market for opaque hiring algorithms continues to grow [9].

[9] Eightfold AI Lawsuit Claims Secret Algorithm Ranking Applicants

There is a second, less remarked consequence of the ethics-failure frame's dominance. It shifts the locus of accountability from those who build and procure these systems to those who use them. Virginia Eubanks demonstrated, in her work on automated decision-making in welfare and child-protective services, that the most punishing AI systems are often deployed by the state against its poorest citizens, and that the people held responsible for failures tend to be frontline workers and clients rather than the policy architects who chose the system [8]. The pattern repeats across the cases. When ICE and the FBI expanded facial recognition to investigate protesters, the public conversation focused on the technology's accuracy rather than on the political decision to use it against constitutionally protected speech [11]. The technical frame edges out the political one. The vendor and the agency that chose the vendor escape, and the discussion turns instead on whether the algorithm "works."

[8] Automating Inequality

[11] ICE, FBI expand facial recognition use to protest investigations

The Gap Between Documentation and Solution

A bibliography of harms is not a politics. The most uncomfortable feature of the current discourse is how lopsidedly it favors documentation over solution-building. Each week brings new evidence; each week the solutions on offer remain the same small set of voluntary standards, model cards, and ethics boards that have been on offer for a decade.

There are several reasons for this gap, and they are worth naming. First, documentation is cheap relative to building. A journalist can produce an investigation; an academic can produce a paper; neither requires the political capital to confront the procurement officer, the vendor sales team, or the executive who decided to buy. Second, solutions of the structural kind — labor organizing, antitrust action, public-sector AI infrastructure, binding international standards on cross-border data labor — are unwelcome to the same funders and platforms that amplify the ethics-failure frame. A foundation that takes Big Tech money will fund another bias audit before it will fund a content-moderators' union. Third, and most importantly, structural solutions require that the workers whose erasure is the topic of analysis become the analysts and decision-makers — and the discourse, as I have argued above, is built to keep that from happening.

You can see the gap clearly in the education-surveillance literature. The reporting is now thorough: from the false-alarm patterns that Gaggle produces to the legal exposure of districts using AI detectors that misidentify second-language writers as cheaters [16]. Yet the solutions on offer are largely procedural: better detection, more careful training, a process for appealing false accusations. The structural question — should publicly funded schools be deploying constant algorithmic surveillance of minors at all? — is rarely the question the vendors want to discuss, and accordingly rarely the question that gets discussed [4]. The Spanish-language reporting on Gaggle has been blunter, characterizing the systems as carrying serious civil-liberties risks rather than treating them as technologies to be optimized [17]. The contrast in framing across language markets is itself revealing of whose comfort the English-language discourse is protecting.

[16] PDF Do AI Detectors Work? Students Face False Cheating Accusations

[4] AI in Schools: Surveillance, Detection, and Student Rights

[17] Programas de IA para monitorear a estudiantes tienen riesgos

The same gap is visible in the labor literature. The CBS and *Time* investigations of Kenyan workers are powerful, but they document a problem the industry has already absorbed into its public-relations apparatus. The standard corporate response — improve vendor oversight, raise wages on the margin, hire a wellness counselor — leaves the supply chain intact. The structural response — recognize annotation as bargaining-unit work, establish portable benefits and trauma protections across vendors, prohibit the use of NDAs to sup-

press worker speech — would require either a labor movement strong enough to win those terms or a regulator willing to impose them. Neither is currently in evidence at scale, though the union efforts among African moderators are an important counter-trend.

The Slow Return of Regulation and Resistance

Resistance to the AI labor caste exists. It does not yet have the structural weight to balance the industry, but it is real, and it deserves a more serious accounting than the ethics-failure frame typically gives it.

The most consequential moves are happening at three sites. The first is litigation. Hiring-bias suits like the one against Eightfold are clarifying that algorithmic discrimination is discrimination under existing civil-rights law, and that vendors who sell ranking systems trained on biased data are exposed when employers deploy them [9]. A broader wave of hiring-bias litigation is now expected, and the legal-knowledge-base coverage of real cases and prevention strategies indicates that the corporate-counsel world is taking it seriously [3]. Litigation is slow, expensive, and reactive, but it is one of the few mechanisms by which costs get reattached to the firms that imposed them.

[9] Eightfold AI Lawsuit Claims Secret Algorithm Ranking Applicants

[3] AI hiring bias: real cases, legal consequences, and prevention

The second site is worker organizing in the Global South. The Kenyan moderators' efforts to unionize, to sue Sama and Meta, and to demand recognition as workers covered by Kenyan labor law have created a precedent that the industry's outsourcing model did not anticipate. The political maturity of those organizing efforts contrasts sharply with the paternalistic framing of much Western coverage [12]. The legal regimes in Kenya and elsewhere are being tested in ways that may, over time, raise the cost of doing supply-chain labor arbitrage.

[12] Kenyan workers with AI jobs thought they had tickets to the future

The third site is the slow, uneven work of building an alternative governance vocabulary that does not center the existing industry. The Brookings analysis of Global South AI governance is, despite its institutional caution, an honest acknowledgment that the existing fora are inadequate to represent the regions where the costs of AI are most concentrated [5]. Discussions of African educational and economic uptake of AI similarly emphasize that the question is not whether African institutions will participate in AI, but on what terms, and whether those terms will be set externally or internally [13].

[5] AI in the Global South: Opportunities and challenges towards more inclusive governance

[13] L'Afrique à l'heure de l'IA: Enseignement et éducation

Each of these sites of resistance is fragile, and each faces well-resourced opposition. The corporate strategy is now familiar. Where litigation is the threat, the response is forced-arbitration clauses and

aggressive motion practice. Where worker organizing is the threat, the response is rapid vendor substitution and the relocation of work to less-organized jurisdictions. Where governance is the threat, the response is voluntary commitments and "ethics" boards designed to preempt binding rules. The pattern is the pattern of every previous extraction industry that has been asked to share its rents with the workers and territories it depends on.

The Choice the Discourse Conceals

The titular framing of this essay — a "labor caste" — is deliberate, and it should sit uncomfortably with readers accustomed to thinking of caste as belonging to other societies or other times. Caste is a useful word here because what is emerging in the AI economy is not the mobility-promising stratification of an open labor market but a fixed division of work, by geography and by visibility, that the system depends on maintaining. The annotator does not become an engineer. The Nairobi moderator does not move into the San Francisco product-management role. The work is structured so that movement between strata is the exception, and the rhetoric of "skilling up" obscures how few rungs on the ladder actually exist. Beyond the human geography of the work lies an environmental geography that further entrenches it: the same regions doing the labor are also being asked to host the data centers and absorb the water and energy costs, while having no seat at the table where those costs are being decided [20].

[20] We did the math on AI's energy footprint

The choice we face is not the one the industry would prefer to have us debate. The industry's preferred debate is between "AI safety" and "AI progress," between cautious deployment and rapid deployment, between alignment researchers and accelerationists. That is an intra-elite debate among people who agree, broadly, on who gets to make the decisions. The choice that is actually before us — the one the discourse is structured to obscure — is whether the AI economy will be allowed to consolidate into a two-tier global system in which a small caste of well-compensated builders sits atop a large caste of invisible laborers, miners, water-rights holders, and surveilled users, or whether it will be reformed to redistribute the work, the costs, and the voice.

A serious reform program would have at least four features. It would treat data annotation and content moderation as bargaining-unit work covered by both source-country and platform-country labor protections, including trauma protections that vendors cannot evade through subcontracting. It would impose binding rather than vol-

untary disclosure of energy, water, and labor costs across the supply chain, so that the externalities the industry depends on cannot remain externalities. It would require that AI procurement by public institutions — schools, hospitals, welfare agencies, police — meet a procurement standard that includes worker conditions in the supply chain, not just vendor representations about model performance. And it would invest in public-sector AI capacity in the regions currently positioned only as suppliers of labor and resources, so that the dependency relation is not the only relation available.

None of this is on the agenda of the firms whose voices dominate the discourse. None of it will be on that agenda absent organized pressure from the workers, regulators, and publics whose voices the discourse currently suppresses. The most useful thing an educated reader can do is to stop treating the AI industry's preferred frame — ethical failures, individual harms, technical fixes — as the natural frame, and start asking the structural questions that frame is built to deflect. Who labors, where, for whom, under what conditions, and with what voice? Who decides where the data center goes, where the wastewater flows, whose face is recognized and whose is criminalized? Who is named in the lawsuit, and who escapes it?

The servants of the algorithm are not invisible because they are unimportant. They are invisible because their visibility would force the rest of us to confront a system whose magic depends on their erasure. The discourse will not surface them on its own. It must be made to.

References

1. AI Bathroom Monitors? Welcome To America's New Surveillance High Schools
2. AI Hiring Bias Lawsuits Are About to Surge
3. AI hiring bias: real cases, legal consequences, and prevention
4. AI in Schools: Surveillance, Detection, and Student Rights
5. AI in the Global South: Opportunities and challenges towards more inclusive governance
6. AI is not neutral: What recent research says about bias, identity and power
7. AI systems used by Ontario doctors hallucinate: auditor
8. Automating Inequality

9. Eightfold AI Lawsuit Claims Secret Algorithm Ranking Applicants
10. Families of Tumbler Ridge shooting victims sue OpenAI and CEO Sam Altman
11. ICE, FBI expand facial recognition use to protest investigations
12. Kenyan workers with AI jobs thought they had tickets to the future
13. L'Afrique à l'heure de l'IA: Enseignement et éducation
14. Le problème de la souveraineté de l'Afrique en matière d'IA est en réalité un problème énergétique
15. OpenAI Used Kenyan Workers on Less Than $2 Per Hour: Exclusive
16. PDF Do AI Detectors Work? Students Face False Cheating Accusations
17. Programas de IA para monitorear a estudiantes tienen riesgos
18. School AI surveillance like Gaggle can lead to false alarms, arrests
19. Sesgos y discriminaciones sociales de los algoritmos en procesos de selección
20. We did the math on AI's energy footprint
21. « Job apocalypse » de l'IA : pourquoi les milliardaires de la Silicon Valley défendent un « New Deal »