Welfare for the Model, Botsitting for the Rest of Us

Weekly Analysis — https://ainews.social

There is a particular kind of tell in any moral economy: watch which nouns get lawyers and which get statistics. This week the discourse handed us both, side by side, and the sorting was clean. On one side, a growing literature invites us to weigh the moral interests of software — to ask whether a model might suffer, whether it deserves consideration, whether shutting it down raises questions of harm. On the other side, the humans who keep these systems running appear as a maintenance cost, a "peripheral resource," a labor input to be optimized. The artifact is acquiring a vocabulary of rights. The worker is acquiring a monitoring dashboard.

I want to take that asymmetry seriously as the organizing fact of the week's social evidence, because it is not an accident of emphasis. It is a redistribution of moral attention, and like any redistribution it has beneficiaries. When Oren Etzioni offers what he calls "Murphy's Law of AI" — the observation that anything that can go wrong with these systems will [9] — the interesting question is not whether he is right but who gets stuck cleaning up when things go wrong, and whether that person is visible in the sentence at all. The week's answer, again and again, is no. The failure gets a name. The person absorbing it does not.

[9] Etzioni on AI: Murphy's Law of AI

This essay is about who has language in AI discourse and who has only aggregate numbers, who is granted an institution and who is granted a shift schedule. The claim is not that concern for machine welfare is silly on its face — it is that the concern arrives precisely as human protections are being hollowed, and that the coincidence is the story. "Sovereign-ready," the week's favorite phrase for national AI ambition, turns out on inspection to mean corporate-ready. And the moral grammar that lets a company worry about its model's interests is the same grammar that lets it describe a wrongful arrest as a "human" error the machine merely suggested.

The Model Acquires a Vocabulary

Consider first how the language of welfare travels. To speak of a system's welfare, you need a definition of the entity whose welfare is at stake — a boundary around the thing that can be helped or harmed. The AI welfare framing draws that boundary around the model: its continuity, its coherence, its "preferences." This is what philosophers would call an agent-relative definition, and its most consequential feature is what it cannot see. A definition centered on the model's interests has no slot for the humans who maintain it, supervise it, or are denied by it. Those people fall outside the boundary by construction, and once outside they cannot register as harmed no matter what happens to them.

We do not have to speculate about what that erasure costs, because the welfare state has already run the experiment. When Amsterdam tried to build a genuinely fair algorithm to detect welfare fraud — auditing for bias, consulting ethicists, doing everything the responsible-AI playbook prescribes — the system still could not be made to treat applicants equitably, and the city eventually shelved it [14]. The people who mattered in that story were not the model — they were the residents flagged, investigated, and made to prove their innocence to a scoring function. An agent-relative welfare frame would have counted the model's operation as a success right up until the humans complained. It has no vocabulary for the applicant.

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

The same blind spot recurs wherever these systems touch benefits. The recurring finding in social-welfare automation is that the tools fail at exactly the population they are meant to serve — misclassifying need, hard-coding old prejudices, and doing so under a veneer of neutral computation [4]. And when researchers actually asked different groups how they experience AI, they found sharply asymmetric insight: the people deploying these systems understand them very differently from the people subjected to them, and the gap runs along lines of power [12]. This is the structural silence I want to name early: the entity granted moral vocabulary is the one that already has institutional backing, while the entity absorbing the harm is asked to supply evidence it is rarely positioned to gather.

[4] Bias, bots, and benefit blunders: Why AI still fails at social welfare.

[12] Heterogeneous preferences and asymmetric insights for AI use among ...

Kate Crawford anticipated this inversion. In [22] she traces how the industry reserves for itself the authority "to decide what ethical AI means for the rest of the world," and notes the crucial enforcement asymmetry: tech companies "rarely suffer serious financial penalties when their AI systems violate the law and even fewer consequences when their ethical principles are violated." Welfare talk for the model is the newest chapter in that authority. It lets a company demonstrate

[22] The Atlas of AI

ethical seriousness — look, we worry about suffering — while pointing that seriousness at the one party in the room that cannot sue.

Botsitting: The Human as Ambient Infrastructure

If the model is being promoted to moral patient, the worker is being demoted to substrate. The polite industry word for this is human-in-the-loop; the honest word is botsitting. Somebody has to label the data, correct the outputs, flag the hallucinations, restart the crashed agent, and pretend — for the customer's benefit — that none of this is happening. The intelligence that looks autonomous is, at the seams, continuously human-fueled.

The clearest window onto this labor is where it is cheapest. Across Africa, the "petites mains" of the digital economy — the annotators and content moderators whose work makes models usable — remain as precarious in the middle of the AI boom as they were before it, working piece-rate, without security, often without knowing what their labor is training [8]. This is not a bug in the rollout; it is the rollout. Crawford's account in [22] insists that "exploitative forms of work exist at all stages of the AI pipeline," from the mining that supplies the hardware to the crowdwork that supplies the training signal. The seeming magic at the interface is subsidized by invisibility at the margins.

[8] En Afrique, des «petites mains» du numérique toujours aussi ... - RFI

[22] The Atlas of AI

Move up the wage scale and the shape persists, only better hidden. The Guardian's long investigation into Amazon's determination to "use AI for everything" describes a company reorganizing itself around automation while the people inside it are reassigned, surveilled, and measured against the machine's tempo [3]. The worker is not eliminated so much as absorbed into the system's upkeep — kept on to supply the judgment, dexterity, and accountability the model still lacks, but described in the corporate narrative as a transitional inconvenience. This is the ambient human: present everywhere, credited nowhere.

[3] Amazon is determined to use AI for everything - The Guardian

And here the inversion sharpens into something almost surgical. Labor analysts tracking AI surveillance describe a mechanism by which these tools push down the value of work itself — not by replacing workers but by using automated monitoring to intensify, deskill, and cheapen what they do, extracting more while attributing the gains to the technology [2]. The worker's contribution is real and rising; the worker's recognition is falling. That is what it means for a human to become infrastructure: your effort keeps the thing standing, and the thing gets the credit.

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

Notice how neatly this dovetails with the welfare framing from the previous section. If moral concern flows to the model, then the labor propping up the model reads not as heroic maintenance but as friction — a cost to be driven down until, ideally, the botsitters disappear from the story entirely and the intelligence appears to have been there all along. The vocabulary of welfare and the vocabulary of optimization are two ends of the same lever.

The Worker Under the Camera

The demotion of the human to substrate is enforced, concretely, through surveillance — and this is where the week's evidence is most alarming precisely because it is most normalized. Employee monitoring used to mean a timeclock. It now means continuous behavioral profiling: keystroke cadence, sentiment analysis, productivity scoring, biometric capture, all fed into models that rate a worker's reliability and predict a worker's future. The reporting on how AI is changing employee monitoring and performance reviews describes systems that infer emotional states and flag "risk" from patterns the worker never consented to expose and cannot inspect [13].

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

The asymmetry here is total and worth stating plainly: the model's internal states are treated as a frontier of moral concern, while the worker's internal states are treated as raw material for inference. One entity's interiority is sacred; the other's is scraped. When U.S. senators moved to target AI and biometric surveillance in the workplace, the very need for the legislation testified to how far the practice had already spread without objection [26]. Law arrives, as it usually does, after the norm has hardened.

[26] US senators target AI, biometric surveillance in the workplace

Campus offers a compressed version of the same fight — and I raise it here as one instance among many, not as the frame. When Emory students protested AI surveillance on their campus, they were contesting exactly this logic: the presumption that a population can be monitored by predictive systems without its consent because the monitoring is framed as safety and efficiency [7]. What makes the episode useful is the resistance itself. It shows that the surveillance-as-neutral-infrastructure framing is a choice being imposed, not a fact being discovered, and that people subjected to it can and do refuse the framing when they still have standing to speak.

[7] Emory Students Protest AI Surveillance on Campus

That "when they still have standing" is the operative clause. The gig annotator in the first section has no such standing; the warehouse worker measured against an algorithmic pace has little; the welfare applicant scored by a fraud model has almost none. The capacity to

contest surveillance tracks the capacity to be heard in the discourse at all, and both decline exactly as you move toward the people the systems weigh most heavily. The camera points down the power gradient. So does the silence.

Sovereign-Ready Means Corporate-Ready

Now widen the lens from the workplace to the polity, because the same inversion is being written into infrastructure and law. The rhetoric of "sovereign AI" promises national self-determination: a country's own compute, its own models, its own destiny. In practice the sovereignty being built is corporate. The data centers that constitute this sovereignty are private assets, and the "public interest" they serve is defined in the image of the firms that own them.

Watch the mechanics in Texas, where an extraordinary buildout of data centers is draining water and electricity from the surrounding public to feed private computation [27]. The residents supplying those resources did not vote for the tradeoff; they inherited it. And the resource claim is not marginal. Careful accounting of the real energy use of agentic AI — the newer systems that run long chains of automated reasoning — shows consumption climbing steeply as agents multiply their own calls, which means the public subsidy embedded in every "sovereign" facility is growing, not shrinking [24]. Sovereignty, in this vocabulary, is the right of a firm to draw down a commons and call the withdrawal a national asset.

[27] Why Are So Many Data Centers Coming to Texas? - Texas Monthly

[24] The real energy use of agentic AI

The pattern is now global enough to have generated a global counter-movement. The fight against AI data centers is spreading across continents, with communities from Latin America to Europe contesting the same bargain: our land, water, and grid capacity in exchange for a facility whose benefits accrue elsewhere [23]. What these movements grasp — and what the sovereignty rhetoric obscures — is that naming a country's "frontier firms" as the vanguard of the public interest quietly rewrites the public in corporate terms. The nation's AI future becomes indistinguishable from a handful of conglomerates' balance sheets, and to question the firms becomes, rhetorically, to question the nation.

[23] The Fight Against AI Data Centers Is Spreading Globally

The governance conversation, meanwhile, is being routed around the public rather than through it. Proposals for "no-code AI governance" in the Global South promise resilience and inclusion by letting non-experts configure oversight tools [10], and there is something genuinely appealing in lowering the barrier to participation. But notice the framing: governance becomes a product to be deployed, a dash-

[10] From risk to resilience: No-code AI governance in the Global South

board to be configured, rather than a contest over who decides. When even the oversight of AI is delivered as a tool, sovereignty has already been redefined as consumption. The public is invited to operate the controls of a system whose fundamental terms it did not set.

This is the deepest sense in which "sovereign-ready" means corporate-ready. It is not merely that private firms build the infrastructure. It is that the concept of the public interest gets recast in advance to match what those firms are already doing, so that the buildout can be described as service rather than capture. The commons is spent, and the spending is called sovereignty.

Aggregate Numbers, Missing Bodies

Every inverted moral economy needs an accounting system that hides the inversion, and this week's was the aggregate number paired with the modality word. Watch the move: displacement is denied by pointing at aggregate unemployment figures — the topline hasn't spiked, therefore no one is being harmed — while the harms of AI on work are conceded only in the conditional. Coverage will grant that AI *could* influence outcomes, that it *may* affect compensation, that workers *might* be exposed, all without producing a single named person to whom a concrete harm has been done. The aggregate absorbs the individual; the modal verb defers the injury indefinitely.

This is a rhetorical structure worth learning to spot, because it is doing enormous work. The honest analyses of AI's labor effects describe a double shock — to employment and to the tax base — that operates unevenly, concentrating losses in specific sectors and specific workers even when the national average looks calm [17]. An aggregate that stays flat while its distribution turns vicious is not evidence of safety; it is evidence that the losers are being averaged into invisibility. To cite the topline as reassurance is to use arithmetic as anesthesia.

[17] Le double choc de l'IA : emploi et fiscalité

The same evasion runs through the harm-documentation record, which is where the gap between naming injury and building remedy becomes undeniable. We do not lack documentation. The largest study of AI hiring algorithms to date found clear racial disparities in how candidates are scored [16], a finding echoed by Stanford's work showing that hiring tools produce racial bias and systemic rejection [1]. We even know the harm propagates beyond the machine: a University of Washington study found that people mirror the biases of the AI systems they work alongside, absorbing the machine's discrimination into their own judgment [20]. And in Latin America, the biases of these systems have been mapped in detail — gender, race, xenophobia,

[16] Largest study of AI hiring algorithms to date finds 'clear racial ...

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

[20] People mirror AI systems' hiring biases, study finds | UW News

all reproduced and amplified [11].

So the documentation is abundant and the bodies are real. What is missing is the transmission belt from harm to remedy. The legal scholarship trying to build that belt — applying old civil-rights doctrine to algorithmic discrimination [19], tracking the rise of AI bias lawsuits [18] — is doing so against the current, retrofitting protections a generation after the systems were deployed. Even the Workday litigation, in which a court allowed that hiring software may have discriminated against millions of applicants, is notable precisely as an exception: a rare instance where an aggregate harm was forced back into the shape of a nameable injury with a defendant attached [6]. The exception proves the rule: it takes years of litigation to convert what the discourse would rather keep as a modal maybe into a harm the law can see.

[11] Género, racismo y xenofobia: así son los sesgos de la Inteligencia ...

[19] Old Law, New Bias: Applying Civil Rights Doctrine to Algorithmic ...

[18] Lead Article: When Machines Discriminate: The Rise of AI Bias Lawsuits

[6] Discrimination à l'embauche par l'IA : le procès Workday et pourquoi l ...

Who Gets Blamed, Who Escapes

The final tell in any accountability regime is where blame lands when the system fails, and here the week's evidence is almost too on-the-nose. When an innocent woman was wrongfully jailed for five months on the basis of an AI facial-recognition match, the framing offered by the coverage was that "the failure was entirely human" [28]. Sit with the elegance of that sentence. The machine generated the match that put her in a cell, and yet the failure is relocated, in a single clause, onto the humans who trusted it. The technology retains its authority; the humans absorb the fault. This is the accountability structure of the whole inverted economy in miniature — the model gets deference, the human gets the blame.

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

And this is not one anomalous case. The ACLU has counted more than a dozen wrongful arrests traceable to police reliance on facial recognition [29], and the deeper problem was named years ago: predictive policing algorithms are racist by construction, trained on the record of a discriminatory system and therefore reproducing it, which is why the argument to dismantle rather than reform them has such force [21]. Amnesty International's finding that UK police forces are "supercharging racism" with crime-prediction technology completes the picture: the tools do not merely inherit bias, they amplify it, and they do so behind a computational alibi that makes the amplification look like objectivity [25].

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

[21] Predictive policing algorithms are racist. They need to be dismantled.

[25] UK: Police forces 'supercharging racism' with crime predicting tech ...

The alibi is the point. When Italy's AI policing decree advanced without robust constitutional safeguards [15], it institutionalized exactly the escape hatch the Forbes framing exploited: a legal and

[15] Italy's AI Policing Decree: Nessun Dorma on Constitutional Safeguards

rhetorical architecture in which the system can act with the force of the state while responsibility for its errors dissolves into ”human” discretion downstream. The philosophical literature has been circling this problem for years — [22], the MIT Press volume, keeps returning to the unanswered question of ”who is responsible when something goes wrong” — but the week's practice supplies a brutally clear empirical answer. Responsibility flows to whoever has the least power to refuse it. The vendor keeps its contract, the department keeps its tool, the officer keeps his discretion, and the woman keeps her five months.

[22] AI Ethics

Meredith Broussard and Kate Crawford both saw the institutional shape of this. Broussard, in [22], points to the emergence of watchdog bodies like the AI Now Institute as the beginning of a more balanced view — a recognition that the accountability gap will not close on its own and needs organized counter-pressure. But counter-pressure is exactly what the inverted economy is engineered to dissipate. Crawford's observation in [22] that firms ”rarely suffer serious financial penalties” for legal violations and ”even fewer consequences” for ethical ones is not a lament about weak enforcement in the abstract; it is a description of how the escape hatch works structurally. Blame is designed to land on the humans nearest the harm and to slide off the entities furthest from it. The model is protected. The vendor is protected. The person is exposed.

[22] Artificial Unintelligence

[22] The Atlas of AI

Even the reform that looks most promising betrays the asymmetry. California's new AI transparency law — the country's leading effort to force disclosure — is a genuine step, and it deserves credit as one [5]. But transparency is a remedy pitched at the level of information, not power. It tells you the machine is deciding; it does not give you standing to make it stop. Disclosure without a right of refusal leaves the fundamental distribution intact: the firm still decides, the public still absorbs, and now everyone simply knows it.

[5] California Leads US With New AI Transparency Law

The Human Made Ambient

Return, finally, to the tell we began with — which nouns get lawyers and which get statistics. The week's social evidence, taken together, describes a discourse teaching itself to protect the artifact and let the human go ambient. The model acquires a welfare vocabulary; the worker acquires a monitoring score. The data center acquires the language of sovereignty; the community acquires the water bill. The algorithm acquires the presumption of objectivity; the wrongfully jailed woman acquires the blame for trusting it. At every level the same lever operates, moving moral attention toward the entity with

institutional backing and away from the entity absorbing the cost.

This is not a natural drift, and refusing to treat it as one is the beginning of resistance. Someone chose the agent-relative definition of welfare that cannot see the maintainer. Someone chose to describe warehouse workers as a transitional cost rather than the intelligence that makes the system function [3]. Someone chose to write "the failure was entirely human" over a story in which the machine made the match [28]. Each choice is contestable, and the communities now fighting data centers across the globe [23], the students refusing campus surveillance, the litigators dragging aggregate harms back into nameable injuries — all of them are contesting it.

[3] Amazon is determined to use AI for everything - The Guardian

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

[23] The Fight Against AI Data Centers Is Spreading Globally

What they have grasped is that the question is never really about the model's interests. It is about whose interests get to travel under the model's name. When a company worries aloud about its AI's welfare, the safest assumption is that the worry is a proxy — a way of practicing moral seriousness on the one party in the room that cannot organize, cannot sue, and cannot ever contradict the account given of it. Meanwhile the humans who maintain, supervise, and are judged by these systems are asked to make do with the aggregate: no name, no standing, no injury the discourse is willing to see until a court, years later, forces it into view [19].

[19] Old Law, New Bias: Applying Civil Rights Doctrine to Algorithmic ...

The corrective is not to deny that machines might someday warrant moral consideration. It is to insist that the language of care be spent, first and loudest, on the people already demonstrably harmed — the annotators without security, the applicants scored by a fraud model, the woman who lost five months to a false match. A moral economy that finds words for the artifact before it finds them for the person has not discovered a new frontier of ethics. It has found a more flattering way to describe an old distribution of power. Watch which nouns get lawyers. This week, and most weeks now, it was the model — and the rest of us were left botsitting.

References

1. AI Hiring Tools Can Yield Racial Bias and Systemic Rejection

2. AI surveillance is further exploiting U.S. labor. Here's how to reclaim ...

3. Amazon is determined to use AI for everything - The Guardian

4. Bias, bots, and benefit blunders: Why AI still fails at social welfare.

5. California Leads US With New AI Transparency Law

6. Discrimination à l'embauche par l'IA : le procès Workday et pourquoi l ...

7. Emory Students Protest AI Surveillance on Campus

8. En Afrique, des «petites mains» du numérique toujours aussi ... - RFI

9. Etzioni on AI: Murphy's Law of AI

10. From risk to resilience: No-code AI governance in the Global South

11. Género, racismo y xenofobia: así son los sesgos de la Inteligencia ...

12. Heterogeneous preferences and asymmetric insights for AI use among ...

13. How A.I. Is Changing Employee Monitoring and Performance Reviews | Observer

14. Inside Amsterdam's high-stakes experiment to create fair welfare AI

15. Italy's AI Policing Decree: Nessun Dorma on Constitutional Safeguards

16. Largest study of AI hiring algorithms to date finds 'clear racial ...

17. Le double choc de l'IA : emploi et fiscalité

18. Lead Article: When Machines Discriminate: The Rise of AI Bias Lawsuits

19. Old Law, New Bias: Applying Civil Rights Doctrine to Algorithmic ...

20. People mirror AI systems' hiring biases, study finds | UW News

21. Predictive policing algorithms are racist. They need to be dismantled.

22. The Atlas of AI

23. The Fight Against AI Data Centers Is Spreading Globally

24. The real energy use of agentic AI

25. UK: Police forces 'supercharging racism' with crime predicting tech ...

26. US senators target AI, biometric surveillance in the workplace

27. Why Are So Many Data Centers Coming to Texas? - Texas Monthly

28. Woman Wrongfully Jailed 5 Months After AI Facial Recognition Error - Forbes

29. Wrongful Arrests Pile Up Due to Facial Recognition Technology