

The Surveillance Pivot: Why Society Must Choose Between Policing and Educating AI Use

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

Watch the sequence, because it repeats with the regularity of a liturgy. A new technology arrives, its capabilities poorly understood; a wave of anxiety follows about who might abuse it; and then, before anyone has thought hard about what they actually want, an industry materializes to sell the anxiety back to us as a product. The product is always the same. It watches. It scores. It flags. It promises to separate the authentic from the fraudulent, the safe from the dangerous, the deserving from the suspect. And it fails — not occasionally, not at the margins, but structurally, in ways that fall hardest on the people least able to contest the verdict. The story of AI detection tools, now quietly being abandoned by the very schools that rushed to adopt them, is not really a story about cheating. It is a story about a society reaching, by reflex, for surveillance when it could have reached for understanding.

That reflex is worth interrogating, because the decision to police rather than to educate is never made in the open. It is made through procurement contracts, vendor demonstrations, and the soft coercion of institutional risk-aversion. Detection tools remain unreliable — teachers know this, administrators have been told this, and yet adoption continues because the tool offers something more valuable than accuracy: it offers the appearance of control [4]. The appearance of control is a commodity, and someone is always selling it. The question this essay wants to pursue is not whether detection works. It manifestly does not. The question is who benefits from our continued willingness to pretend it might — and whose voices are structurally absent from the room where that pretense gets ratified.

Because the AI discourse, taken as a whole, has a strange center of gravity. A remarkable share of it — by some accounts nearly two in five of the arguments circulating in any given week — concerns ethical failures: bias, harm, discrimination, breach. This looks, at first, like a healthy critical culture holding power to account. But look again. The dominance of failure-documentation may reveal less about our capacity for accountability than about our inability to convert docu-

[4] AI detection tools are unreliable.
Teachers are using them anyway :
NPR

mentation into anything else. We have become fluent in describing the wound and mute on the subject of the cure. That asymmetry is not accidental. It is a feature of who holds power in the conversation.

The Reflex to Watch

The surveillance pivot did not begin with AI, and it does not end with essays. It is a general posture toward populations deemed risky, and its infrastructure has been quietly assembling for years in the places where the least powerful people have the least ability to say no. Consider the surveillance of children, which now runs continuously through the school day and beyond. Monitoring software scans what students type, read, and search, ostensibly to protect them, and in the process constructs a permanent behavioral record of minors that would be politically unthinkable if proposed for adults [20]. The rationale is always safety. The result is a generation habituated to being watched.

What makes this a study in power rather than merely in policy is the pattern of who bears the errors. When these systems misfire — and they misfire constantly — the consequences do not fall on the vendors who sold them or the districts that deployed them. Surveillance products like Gaggle have generated false alarms that escalate into police contact, wellness checks, and in some cases arrests, disproportionately affecting the students already most surveilled by the world outside the school gates [21]. In Houston, districts have deployed AI tools to scan for mental-health red flags, framing constant monitoring as care [15]. The framing matters, because it converts a question of power — who gets to watch whom, and with what consequences — into a question of benevolence, where objection sounds like a refusal of help.

This is the deep grammar of the surveillance pivot: it recasts the extension of institutional power as an act of protection, and thereby immunizes itself against the ordinary language of rights. Children are watched twenty-four hours a day across an expanding array of devices and platforms, and the normalization proceeds precisely because the watched cannot vote, cannot lobby, and cannot easily narrate their own experience into the public record [22]. When advocacy organizations attempt to reframe these deployments as questions of detection, surveillance, and student rights rather than as neutral safety tools, they are pushing against a discourse already structured to treat the watched as objects of concern rather than subjects of consent [6]. The asymmetry is total. One side builds the systems and writes the press releases. The other side experiences the false alarm.

[20] Public Schools, Private Eyes: How EdTech Monitoring Is Reshaping Public Schools

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

[15] Houston schools use AI tools to watch for mental health red flags

[22] School Surveillance: Your Kids Are Being Watched 24/7

[6] AI in Schools: Surveillance, Detection, and Student Rights - CPAI

Who the Errors Are For

Move from the school to the labor market, and the machinery is identical; only the stakes change. The largest study to date of AI hiring algorithms found clear racial disparities in how these systems score candidates — the same tools sold as instruments of objectivity, cleansing hiring of human prejudice, turn out to reproduce and launder that prejudice at scale [18]. This is the crucial move to watch, because it reveals the true function of the detection reflex across every domain it colonizes. The promise is neutrality. The delivery is bias with a veneer of mathematics, which is worse than ordinary bias because it is harder to see and harder to appeal.

Human-resources professionals are beginning to be warned about the hidden biases baked into recruitment algorithms, but the warning arrives after the infrastructure is already load-bearing — after millions of applications have already been filtered by systems whose logic no applicant can inspect [8]. The rejected candidate never learns why. The bias is invisible from below and deniable from above. And this is precisely where the question of power becomes a question of geography and identity: in Latin America, researchers have documented how these systems encode gender, racism, and xenophobia, importing the prejudices of their training data into contexts where the harmed populations have almost no representation in the rooms where the models are designed [14].

The pattern generalizes to prediction itself. When institutions deploy models to forecast who will succeed and who will fail, the forecasts systematically disadvantage students already at the margins, because the models learn from a past shaped by inequality and then project that inequality forward as if it were destiny [19]. A prediction that the disadvantaged will underperform becomes, through the allocation of attention and resources it triggers, a mechanism for ensuring they do. This is the self-fulfilling architecture of algorithmic suspicion, and it operates everywhere the detection reflex takes hold. Ruha Benjamin named this dynamic precisely: technologies marketed as neutral or even benevolent frequently deepen the discrimination they claim to transcend, a phenomenon she called the New Jim Code [5].

What unites the surveilled child, the filtered job applicant, and the pre-judged student is that they are all being *scored by systems they did not consent to and cannot contest*, and that the errors of those systems are not distributed randomly. The false positive has a demographic. It has an accent, a zip code, a name that the training data flagged as unfamiliar. Meredith Broussard has argued that our faith in computational objectivity — what she calls technochauvinism,

[18] Largest study of AI hiring algorithms to date finds 'clear racial disparities

[8] Biais cachés des algorithmes de recrutement : ce que les RH doivent

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

[19] Predictive models in higher ed disadvantage some students

[5] Race After Technology

the belief that a technical solution is always superior — blinds us to exactly this, to the way that automated systems misunderstand the social world and then impose their misunderstanding as fact [5]. The discourse documents these errors abundantly. What it does not do is ask why we keep buying the systems that produce them.

[5] Artificial Unintelligence

The Vanishing of Accountability

Here is the question the failure-documentation industry rarely reaches: when a detection system falsely accuses someone, who pays? The answer, with brutal consistency, is that the accused pays and the accuser walks. When a high schooler in Palo Alto was flagged by an AI detector for cheating, the burden of proof inverted entirely — the family had to mount a defense, hire help, and fight to clear a record, while the tool that made the accusation faced no scrutiny at all [1]. The vendor's error became the family's problem. This inversion is the accountability vanishing act in miniature, and it scales all the way up.

[1] A Palo Alto high schooler was accused of AI cheating. His family filed

Kate Crawford has documented the structural version of this immunity: tech companies rarely suffer serious financial penalties when their AI systems violate the law, and even fewer consequences when they violate their own stated ethical principles [5]. The principles are, in this sense, a form of insurance — public documents that signal virtue while binding no one. The same welfare and eligibility systems Crawford describes, systems that track anomalous patterns to cut people off from benefits and accuse them of fraud, embody the identical logic: the algorithm accuses, the human must exonerate, and the institution that deployed the accusing machine bears no cost when it is wrong. Virginia Eubanks traced how these automated eligibility regimes function as a kind of digital poorhouse, converting the mere fact of being poor into a permanent condition of algorithmic suspicion [5].

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

Notice how this reshapes the entire structure of the conversation. The AI discourse produces an enormous volume of harm documentation, and that volume can feel like accountability — surely, if we are describing the failures this thoroughly, someone is being held responsible? But documentation and accountability are different things, and the gap between them is where power hides. To describe a harm is to perform concern; to assign consequence is to threaten an interest. The discourse does a great deal of the former and almost none of the latter. We have detailed, sophisticated accounts of algorithmic bias in hiring, in schools, in welfare — and we have almost no cases where the builders of those systems faced anything commensurate with the damage. Shoshana Zuboff's warning is apt here: without a robust

[5] Automating Inequality

countervailing movement in which democratic institutions tether raw information capitalism to public interest, we are thrown back on the market form of the surveillance companies themselves, where democracy and social trust are superseded by what she calls the certainty machines, their priests, and their owners [5].

The priests, in this framing, are the experts who validate the systems; the owners are those who profit; and the rest of us are invited to trust the certainty machine while absorbing its errors. When schools respond to AI by declaring an arms race — treating the classroom as a courtroom where every student is a potential defendant — they are enacting the priesthood’s logic, adopting the posture of prosecution rather than pedagogy [24]. The move to ban and block the technology outright, urged in some quarters as the only responsible response, is the same reflex in a harder shell: control through prohibition, suspicion as default, the assumption that the watched are guilty until the machine says otherwise [9].

Cataloguing Wounds, Building Nothing

If the discourse is so good at documenting harm, why does the harm persist and even accelerate? Part of the answer is that documentation is cheap and construction is expensive, and the incentives of everyone with a platform favor the cheap thing. It costs little to write another account of algorithmic bias; it costs a great deal — politically, financially, institutionally — to build the public infrastructure that would make bias contestable. So the discourse fills with wounds and empties of blueprints. The people harmed appear in the discourse mostly as evidence, as case studies, as statistics in someone else’s argument. They rarely appear as authors of the solution.

This is the structural silence that ought to unsettle us. When the global trends in AI and schools are surveyed, the documentation of bias risk is thorough and the account of remedy is thin, a gap that is treated as normal rather than scandalous [2]. Even the earnest playbooks for detecting and mitigating bias in automated grading operate within the frame of the deploying institution — they ask how to make the surveillance machine fairer, not whether the machine should exist, not who decided to install it, not what the surveilled would choose if asked [12]. The reform imagination is bounded by the tool. We debate the calibration of the trap and forget to ask who set it.

The people best positioned to redesign these systems from first principles are, tellingly, the ones with the least access to the discourse

[5] The Age of Surveillance Capitalism_ The Fight for a Human Future at the New Frontier of Power

[24] The Next AI Wave: Classrooms as Courtrooms (opinion)

[9] Colleges And Schools Must Block And Ban Agentic AI Browsers Now Here’s Why

[2] AI Cheating in Schools: 2026 Global Trends & Bias Risks

[12] Detecting & Mitigating Bias in AI Grading: A Practical Playbook

that shapes them. There is a movement to decolonize AI, to center human dignity over technological dependency, that begins from precisely this observation — that the communities most affected by AI systems are the least represented in their design, and that genuine remedy requires shifting power, not merely auditing outputs [23]. This is a different kind of intervention from harm-documentation. It does not merely describe the wound; it asks who holds the knife and proposes to take it from them. That the decolonial framing remains marginal in the mainstream conversation, while failure-cataloguing dominates, tells you something exact about where power sits.

[23] The Movement to Decolonize AI: Centering Dignity Over Dependency

Consider the geography of that power in its rawest form. The surveillance and data infrastructures shaping African conflicts are frequently owned, trained, and controlled elsewhere, a dynamic that scholars have named data colonialism: the extraction of behavioral information from populations who neither consent to nor benefit from its use, governed by interests continents away [7]. And even the geopolitics of the technology itself follows the logic of control — when China moves to close off access to its own AI technologies, treating capability as a strategic resource to be hoarded, it demonstrates that the fundamental question everywhere is not *what can this do* but *who gets to decide* [17]. The failure-documentation industry, busy cataloguing the outputs, tends to leave the ownership question politely unexamined.

[7] AI surveillance and data colonialism shape African conflicts

[17] La Chine veut fermer l'accès à ses technologies d'IA

The Material Underside of Suspicion

There is a temptation to treat all of this as a matter of information, of bits and models and abstract flows of data. But the surveillance pivot has a physical body, and that body is being built in specific places, on specific land, drawing specific water and power, over the objections of specific people who were not consulted. To follow the power in AI discourse all the way down is to arrive, eventually, at a data center — and at the community fighting it.

In Memphis, residents have battled the construction of Elon Musk's xAI data center, contending with the environmental burden of a facility that arrived in their neighborhood as a fait accompli, its emissions and energy draw imposed on a community whose consent was never sought [16]. This is not a peripheral story. It is the same story as the surveilled child and the filtered applicant, told in concrete and kilowatts. The pattern of community opposition to AI data centers has grown into a national phenomenon, with hundreds of billions of dollars in projects meeting resistance from the people who live nearest the infrastructure and furthest from its profits [3]. The power flows both

[16] Inside Memphis' Battle Against Elon Musk's xAI Data Center - TIME

[3] AI Data Center Community Opposition Statistics 2026: \$228 Billion

ways: electrical power into the machines, political power away from the communities.

The energy question sharpens the point. As efforts to benchmark the electricity consumption of AI models make plain, these systems carry a material cost that someone must pay, and the accounting of who pays it — in higher utility bills, in strained grids, in local pollution — is systematically obscured by a discourse that prefers to talk about capabilities and risks in the abstract [10]. Crawford’s insistence in the Atlas that AI is not artificial and not immaterial but a planetary extractive industry, drawing on minerals, labor, and energy from the least powerful to serve the most powerful, is the necessary corrective. The surveillance pivot is not just an epistemic choice; it is an infrastructural commitment, and infrastructure has a way of foreclosing the futures we might otherwise have chosen.

[10] Consumo energético de la IA: el benchmark ML.ENERGY

This is why the choice between policing and educating is not merely a metaphor. To build detection infrastructure is to build data centers, to sign contracts, to train models on extracted behavior, to habituate a population to being watched — and each of these commitments makes the alternative harder to reach. Every dollar spent on the machinery of suspicion is a dollar not spent on the slower, more distributed, less profitable work of building public understanding. The material bias of the pivot runs toward surveillance because surveillance is what can be sold, financed, and scaled. Education cannot be packaged and licensed in the same way, which is precisely why it keeps losing the budget fight.

Literacy as the Redistribution of Judgment

What would it mean to reach for understanding instead? Not the pallid version that appears in vendor brochures — a training module bolted onto the same surveillance regime — but the substantive version, in which the capacity to judge AI is redistributed away from the priests and toward the public. This is the real content of the phrase AI literacy, and it is worth stripping the jargon to see what it actually claims. It claims that ordinary people, equipped with the right knowledge and the right tools, can make collective judgments about these systems that are better than the automated verdicts imposed on them. It claims that competence can substitute for surveillance. It claims, in short, that the people currently positioned as objects of the discourse could become its authors.

The evidence that this is possible comes, revealingly, not from institutions but from streets. During the protests in Kenya, citizens used

AI tools on their own terms — to translate, to document, to organize, to make sense of official narratives and construct their own — demonstrating a form of grassroots technological agency that no detection regime anticipated or could contain [11]. This is the literacy-first society in embryo: not people being scored by machines, but people wielding machines toward ends they chose themselves. The difference is entirely a difference of power. In the surveillance model, the technology acts on the population. In the literacy model, the population acts through the technology. The same tools, radically different politics.

The critical scholarship on AI in learning has begun to insist on exactly this distinction, asking whether AI tutors empower or enslave — whether they build the learner’s capacity for independent judgment or cultivate a dependency that leaves the learner more governable, not less [13]. The question is not rhetorical, and its answer is not determined by the technology. It is determined by who designs the deployment and toward what end. A tool that explains its reasoning and invites contestation empowers; a tool that issues verdicts and forbids appeal enslaves. Detection is the second kind of tool. Literacy is the demand for the first.

None of this happens without public infrastructure, which is the part the market cannot supply and will not fund. A genuine literacy-first society requires media literacy campaigns, accessible and independent evaluation of the tools, cultural norms around transparency, and — crucially — a policy architecture that treats these as public goods rather than private products. When national education authorities finalize AI priorities, the choice they face is exactly the fork this essay has traced: they can codify the surveillance regime, funding detection and monitoring, or they can fund the slower work of building judgment [25]. The direction of that policy will shape which future becomes cheaper to inhabit. The instinct toward automated eligibility and algorithmic sorting has already appeared in public administration elsewhere — in Gothenburg, the automated assignment of pupils to schools produced a documented case of algorithmic injustice, a preview of what happens when consequential decisions are handed to systems that cannot be reasoned with [26].

The deepest argument for literacy over policing is not that detection fails technically, though it does. It is that surveillance and education produce different kinds of citizens. A society that responds to a new technology by watching everyone teaches its members that they are suspects, that trust is naive, that judgment is best delegated to machines. A society that responds by teaching everyone teaches the opposite: that people can be trusted with understanding, that judg-

[11] Cómo están usando IA los kenianos durante las protestas

[13] Do AI tutors empower or enslave learners? Toward a critical use of AI

[25] The US Department of Education Just Finalized Its AI in Education Priority Here’s What It Means

[26] Un cas pratique d’injustice algorithmique : l’attribution automatisée

ment is a shared civic capacity, that competence is the proper answer to complexity. The MIT account of AI ethics keeps returning to the same governing question — how many of our decisions, and how much of each decision, do we want to delegate, and who bears responsibility when the delegated decision goes wrong [5]. The surveillance pivot answers by delegating maximally and diffusing responsibility to nobody. The literacy alternative answers by keeping judgment human and responsibility locatable.

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

The Fork, Named Plainly

Return, at the end, to the question of power, because everything in this essay has been an attempt to make it visible. Who has power in the AI discourse? The vendors who sell the certainty machines, the institutions that buy them to manage risk, the experts who validate them, and the owners who profit — Zuboff’s priests and their patrons, escaping the consequences that Crawford documents they almost never face. Who does not? The surveilled child, the filtered applicant, the pre-judged student, the Memphis resident breathing the data center’s exhaust, the Kenyan protester improvising agency from tools built to govern her, the Latin American whose name the model flagged as foreign. They appear in the discourse abundantly — as evidence. They appear in its decisions rarely — as authors.

The dominance of failure-documentation in the conversation, that striking concentration of energy on cataloguing ethical harm, turns out on inspection to be a symptom of this asymmetry rather than a cure for it. We document harm so fluently because documentation is the one form of participation available to those without power, and we build remedy so poorly because remedy requires the redistribution of power that the powerful have no incentive to grant. The gap between the wound and the blueprint is not an oversight. It is the shape of the current distribution of voice.

The pivot toward surveillance, then, is not really a technical decision that happens to have social consequences. It is a social decision — about trust, about who counts as a suspect, about whether judgment is a civic capacity or a proprietary service — that hides inside a technical vocabulary to escape democratic scrutiny. The abandonment of detection in schools is the first crack, the first place the brittle thing broke in public. The society that notices the crack has a choice about what to build in its place. It can rebuild the surveillance regime with better calibration, funding a more sophisticated version of the same suspicion. Or it can choose the harder and more democratic path of

building public competence, treating its members as citizens capable of judgment rather than as populations in need of management. The tools that Kenyans turned toward their own liberation are the same tools that watch children twenty-four hours a day; the difference lies entirely in who holds the power to decide. That decision is still open. It will not stay open long.

References

1. A Palo Alto high schooler was accused of AI cheating. His family filed
2. AI Cheating in Schools: 2026 Global Trends & Bias Risks
3. AI Data Center Community Opposition Statistics 2026: \$228 Billion
4. AI detection tools are unreliable. Teachers are using them anyway : NPR
5. AI Ethics - The MIT Press Essential Knowledge series
6. AI in Schools: Surveillance, Detection, and Student Rights - CPAI
7. AI surveillance and data colonialism shape African conflicts
8. Biais cachés des algorithmes de recrutement : ce que les RH doivent
9. Colleges And Schools Must Block And Ban Agentic AI Browsers Now Here's Why
10. Consumo energético de la IA: el benchmark ML.ENERGY
11. Cómo están usando IA los kenianos durante las protestas
12. Detecting & Mitigating Bias in AI Grading: A Practical Playbook
13. Do AI tutors empower or enslave learners? Toward a critical use of AI
14. Género, racismo y xenofobia: así son los sesgos de la Inteligencia
15. Houston schools use AI tools to watch for mental health red flags
16. Inside Memphis' Battle Against Elon Musk's xAI Data Center - TIME
17. La Chine veut fermer l'accès à ses technologies d'IA

18. Largest study of AI hiring algorithms to date finds 'clear racial disparities
19. Predictive models in higher ed disadvantage some students
20. Public Schools, Private Eyes: How EdTech Monitoring Is Reshaping Public Schools
21. School AI surveillance like Gaggle can lead to false alarms, arrests
22. School Surveillance: Your Kids Are Being Watched 24/7
23. The Movement to Decolonize AI: Centering Dignity Over Dependency
24. The Next AI Wave: Classrooms as Courtrooms (opinion)
25. The US Department of Education Just Finalized Its AI in Education Priority Here's What It Means
26. Un cas pratique d'injustice algorithmique : l'attribution automatisée