The Authentication Racket: Who Profits When Trust Becomes a Technical Problem?

Weekly Analysis — https://ainews.social

There is a particular kind of market that opens up whenever a society loses confidence in something it used to take for granted. The market for bottled water expanded with anxiety about municipal pipes. The market for home security swelled with the disinvestment of public safety. The market for private tutoring expanded as public schooling frayed. We are now watching a similar enclosure unfold around something stranger and more elemental: the market for trust itself. Did a human write this? Was that voice on the phone really your daughter? Is the photograph of the protest authentic, or stitched together by a model trained on ten million other protests? The question is genuine. The answers are increasingly being sold to us, and the people doing the selling are not neutral.

What we are calling, somewhat mildly, "the AI authentication problem" is in fact the construction of a new infrastructure of suspicion — and like every infrastructure, it has owners, tollbooths, and people who will be priced out. The dominant editorial framing treats provenance and detection as a technical race against synthetic media, a contest between forgers and forensic tools. That framing is not wrong, but it is conveniently incomplete. It elides the more telling question, which is not *can we authenticate?* but *whose authentication will count, and who pays the rent on the verification stack?* As one Wall Street Journal columnist asked, only half-rhetorically, in a piece worrying over the substitution of "intelligence" with its synthetic twin, the public is being trained to fear AI as a thing that happens *to* it rather than as a system *built by* identifiable people [15]. The fear is the product. The remediation is the upsell.

[15] Opinion | Would You Be Afraid of IA?

This essay is an attempt to look directly at the racket — meaning not a crime, exactly, but the older sense of the word: a noisy, lucrative system whose participants are too busy with the noise to notice what is being arranged behind it. The argument is that the rush to solve AI-generated content with detection and provenance is producing a new economy of trust gatekeepers, that this economy is materially shifting power away from independent creators, defendants, students, job applicants, and protesters, and that the very definition of what

counts as ”authentic” is now a contested asset class. As Kate Crawford put it in a passage worth keeping in view, tech companies ”rarely suffer serious financial penalties when their AI systems violate the law and even fewer consequences when their ethical principles are violated” [20]. The authentication economy inherits that asymmetry and extends it.

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

The Provenance Gold Rush

Begin with the vendors, because they are the most legible part of the story. A class of companies has formed around the proposition that AI-generated content threatens institutional trust and that the appropriate response is for institutions to buy the threat-mitigation product. Adobe’s Content Credentials, Microsoft’s media provenance work, the C2PA consortium standards, school-side detection products like Turnitin’s AI score, hiring-side validators like Eightfold and Workday — these are not neutral plumbing. They are commercial infrastructure with revenue models, board accountability, and a stake in continuous problem persistence.

The pattern is most exposed in employment. A wave of lawsuits has begun to crack open how algorithmic hiring tools assemble and sell judgments about who is ”really” qualified, with plaintiffs alleging that systems silently rank, downgrade, and exclude in ways that no candidate can audit and no employer is obliged to explain [3]. The Eightfold case is instructive because the plaintiffs’ core claim is not that the algorithm misfires randomly but that it operates as a ”secret algorithm” — a closed proprietary scoring system, sold as an objective talent evaluator, that ranks applicants on logic the applicants themselves cannot inspect [10]. The Workday litigation is structurally similar: a vendor’s screening machinery, embedded in hundreds of HR pipelines, determines who is and is not authentically employable, with disparate impact on older applicants and applicants of color [4].

[3] AI Hiring Bias Lawsuits Are Reshaping Recruiting in 2026: What ...

[10] Eightfold AI Lawsuit Claims Secret Algorithm Ranking Applicants

[4] AI Hiring Discrimination: How Algorithms Reject Millions of Qualified ...

What is the verification claim being sold here? It is that the candidate is, or is not, ”real” enough — credentialed enough, fluent enough, articulate enough, neurotypical enough — to be considered. Recent data suggest these systems systematically fail neurodivergent applicants, whose communication patterns are precisely the kind of signal an alignment-to-norms classifier was built to penalize [5]. A French analysis of recruitment AIs found a self-preference bias in which models trained on prior hiring outcomes effectively re-authenticate the demographic profile of yesterday’s winners as the proper template for tomorrow’s candidates [13]. Authentication, here, is not detection of

[5] AI Hiring Is Excluding Neurodivergent Candidates—The Data ... - LinkedIn

[13] Le biais d’auto-préférence des IA de recrutement

forgery; it is a forgery of detection. The system pretends to be reading the candidate when it is reading itself.

And note the political economy: at the moment state-level fairness rules tried to force these vendors to open their boxes, the federal posture moved in the opposite direction, with the Trump administration aligning with Elon Musk to attack state AI hiring fairness laws as obstacles to innovation [22]. The vendors do not even need to defend themselves; the political superstructure does it for them. This is what a successful racket looks like: the fee is collected, the audit is denied, and the regulator is converted into a brand ambassador.

[22] Trump administration joins Musk to take aim at US state AI hiring ...

What "Authentic" Now Means in the Boilerplate

The deeper move under the authentication economy is semantic. "Authentic" used to mean something close to: produced by a person, with intention, in a context recoverable to other persons. Inside the new infrastructure, "authentic" means: bearing a verifiable cryptographic claim issued by a trusted vendor's pipeline. These are not the same thing. The first is a property of human conduct. The second is a property of corporate certification.

This is a familiar enclosure pattern. Shoshana Zuboff has described how surveillance capitalism trades on "the volatile conditions of existence" — uncertainty, isolation, informational vertigo — by offering "solutions to individuals in the form of social connection, access to information, time-saving convenience, and, too often, the illusion of support" [20]. The illusion of authentication is the latest entry in that catalog. What is offered is a sense that one's content, one's voice, one's identity has been *verified*. What is extracted is the right to define what verification means and to charge for inclusion in it.

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

Ruha Benjamin has been even more precise about the rhetorical scaffolding. Systems "that seem to objectively rank people on the basis of merit and things we like, such as trustworthiness, invoke 'efficiency' and 'progress' as the lingua franca of innovation" [20]. She points to China's social credit policy as an unembarrassed instance — "It will forge a public opinion environment where keeping trust is glorious." Western deployments are more decorous, but the structure is the same: a private or quasi-private system anoints some performances as trustworthy, demotes others, and naturalizes the ranking as merit. The Content Credentials seal on a New York Times photograph and the dimmer-but-still-real ranking that an applicant tracking system assigns to a resume are the same kind of object: a vendor-issued legitimacy stamp, monetized.

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

The Two-Tier Information Ecosystem

If we extrapolate the logic of the racket outward — beyond hiring and beyond newsroom photo desks — the shape of the resulting information ecosystem becomes legible, and worrying. There will be a verified premium tier: news organizations with budgets for C2PA-compliant capture devices, creators on platforms that pay for provenance integration, institutions whose press releases ship with cryptographic signatures, and the political and commercial actors with enough resources to purchase placement in the trusted ledger. There will be an unverified remainder: independent journalists, community media, public-records leakers, citizen documentarians of police violence, and most of the global majority of internet users, whose materials will be flagged, downranked, or quietly treated as suspect by default.

The educational sector is already prefiguring this stratification. Detector tools have been wrongly accusing students of cheating, with serious consequences, and the false-positive rate falls disproportionately on writers whose English is non-native or whose prose patterns deviate from the model's idea of human normality [9]. A Hechinger Report investigation found that an AI essay grader penalized Asian American students more than white peers, lowering their scores in ways the vendor could neither fully explain nor reliably correct [17]. Translate this dynamic from the classroom — where it is bad enough — to the labor market, the asylum interview, the credit application, the news-aggregator's trust filter. The same vendor logic operates: pattern-match to a normative template; flag deviation as inauthenticity; let the costs of being non-default fall on those least able to dispute them.

[9] Detectores de IA acusan falsamente a estudiantes de hacer trampa, con ...

[17] PROOF POINTS: Asian American students lose more points in an AI essay ...

The harm is compound, because once a piece of content or a person is flagged as low-trust by one infrastructure, downstream systems treat the flag as a feature of the subject rather than a feature of the classifier. A literature review on algorithmic bias in education makes this point in academic register: prediction systems fold past inequities into present scoring, which then determines future access [7]. The technical name is feedback loop. The political name is caste.

[7] Algorithmic Bias in Education | International Journal of Artificial ...

The structural silence here, worth naming aloud, is that nobody is building a publicly funded, democratically governed authentication commons. There is no civic provenance utility in the way that there is a postal service or a public library. Every serious provenance proposal currently in deployment has a corporate sponsor, a licensing model, or a state-security entanglement. Meredith Broussard's work has long argued that the technocratic instinct is to mistake a narrow computational definition of a problem for the problem itself, and to sell that

narrow definition as a comprehensive solution [20]. The authentication market is a textbook case: the social problem is the erosion of shared epistemic ground; the marketed solution is a metadata standard with subscription tiers.

[20] Artificial Unintelligence

When Authentication Wears a Badge

The most dangerous version of the racket is the one in which the vendor's customer is the state and the product authenticates not content but people. Here the trust-engineering rhetoric reveals what it always was: a power tool.

Consider the pipeline that ran through the Columbia University protests. The New York Police Department, which is technically subject to a ban on facial recognition, used a workaround through the Fire Department to run images of pro-Palestinian student protesters through Clearview AI's database — a private, scraped-faces vendor selling itself as identity ground truth [14]. This is authentication as enforcement. The state purchases from a vendor the right to declare *who that person really is*, in defiance of the local rule that was supposed to prevent exactly that. The vendor's business model is to be the one whose claim about identity is treated as final.

[14] NYPD Bypassed Facial Recognition Ban to ID Pro-Palestinian Student ...

The pattern is global. A *Rest of World* investigation tracked how governments from Russia to India deploy facial recognition specifically against protesters, with the technology marketed as crowd-safety infrastructure but operationalized as a dissent-suppression filter [12]. The Associated Press has documented a quieter but no less aggressive variant: private groups in the United States working to identify and report student protesters to immigration authorities, leveraging the same off-the-shelf identity-resolution stack [16]. The infrastructure does not care who buys it. The infrastructure does what it is built to do.

[12] How governments use facial recognition for protest surveillance - Rest ...

[16] Private groups work to identify and report student protesters for ...

This is what Crawford means when she writes that AI systems are now performing "capture that were once reserved for extralegal espionage. Welfare decision-making systems are used to track anomalous data patterns in order to cut people off from unemployment benefits and accuse them of fraud" [20]. Authentication infrastructure, in its enforcement mode, generalizes the technique. The protester is authenticated as a deportation candidate. The unemployment claimant is authenticated as a fraudster. The student is authenticated as a cheater. In each case the verb *to authenticate* has slipped its meaning: it no longer answers *who are you?* but rather *which category does your pattern justify enrolling you in?*

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

Even the school-monitoring industry — Gaggle, GoGuardian, Securly, Bark — operates on this logic. These systems ostensibly protect students from harm by scanning their school-issued devices for signs of crisis, but reporting has shown they generate false alarms that have led to arrests of children and exposure of LGBTQ kids to families that did not know [18]. A separate investigation found that the same surveillance products carry significant security vulnerabilities of their own, exposing the data they were sold to protect [19]. The market sells *authenticity-of-distress detection* to school districts; the externalities — frightened families, criminalized children, leaked data — accrue elsewhere [11]. The vendor does not bear them. The students do.

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

[19] Schools use AI to monitor kids, hoping to prevent violence. Our ...

[11] How AI monitors school Chromebooks and what it means for privacy ...

Who Pays the Energy Bill of Trust?

There is a material substrate to this entire economy that the authentication discourse almost never discusses, and that omission is itself a tell. Trust infrastructure — model training, continuous detection inference, cryptographic ledger maintenance, real-time content scanning at platform scale — runs on a staggering and growing electrical and water footprint. *MIT Technology Review*'s detailed reporting on AI's energy footprint laid out the scale of the demand surge and the degree to which the social costs of that demand fall on the communities nearest the data centers, not on the firms operating them [23].

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

The political economy here is exact. The trust infrastructure is sold as a public good. The energy infrastructure that powers it is paid for, partly, through publicly subsidized utility rates, public water allocations, and locally borne pollution. The verification fees flow to the vendor. The grid stress flows to the host town. A peer-reviewed analysis of generative AI's socioeconomic effects has begun documenting the distributional consequences — that productivity gains accrue narrowly while environmental and labor costs disperse widely [21]. Authentication is a particularly clean illustration of this asymmetry, because it is sold as a service to society as a whole — *we all need to know what is real* — and is built on infrastructure whose costs are quietly taxed onto somebody else.

[21] The impact of generative artificial intelligence on socioeconomic ...

The Accountability Gap

A useful diagnostic, when looking at any technological discourse, is to ask: where is the harm documented, and where is the solution authored? In a healthy field these would be in conversation. In the authentication economy they are in different rooms.

The harm is documented in employment law filings, in immigration cases, in school disciplinary appeals, in journalism investigations. The Hechinger Report on essay grading, the *AP* on school surveillance, *The City* on the NYPD workaround, the Workday and Eightfold complaints — these are produced by reporters, plaintiffs, civil-rights lawyers, and affected workers. The solution-building, by contrast, happens almost entirely inside vendor consortia, standards bodies funded by those vendors, and procurement contracts that flow back to the same firms whose tools generated the harm in the first place. This is not a coincidence. It is how the racket renews itself: the firms that are paid to detect the problem are the firms paid to fix it.

When a recruitment AI is found to discriminate, the remedy is usually another vendor product — bias auditing software, often sold by the same family of firms, or a "fairness layer" bolted onto the original system [2]. When a school's AI grader misranks Asian American students, the remedy proposed is rarely "stop using AI graders for high-stakes assessment" but rather a "playbook" for mitigating bias, sold by people in the same trade [8]. The structural feature here is that *failure of authentication does not threaten the authentication market; it expands it.* Each documented harm becomes a new line item: detection-of-detection, audit-of-audit, the verifier's verifier.

[2] AI Hiring Bias Exposed in New Study | X0PA AI posted on ...

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

This is the move worth watching most carefully. In a properly functioning accountability regime, repeated failure produces consequences for the producer — fines, exit, criminal liability for executives. In the authentication regime, repeated failure produces consequences for the *subject* — the rejected applicant, the suspended student, the deported protester — and produces revenue for the producer. As Crawford observes, the consequences for the systems themselves are vanishingly few [20].

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

The accountability gap is also a discursive gap. Notice how the pronoun structure works in vendor literature: *we* face an authenticity crisis; *we* must adopt provenance; *we* must defend democracy from synthetic media. The "we" performs a kind of populism. But when a specific harm is named — a wrongful AI cheating accusation, a deportation enabled by a face match, a hiring rejection by a black-box ranker — the pronoun mysteriously fragments. Suddenly the harmed party is an individual with an individual grievance, and the system is a complex multi-stakeholder ecosystem about which no specific actor can be held responsible. The collective noun was load-bearing. It hid the asymmetry.

What the Discourse Won't Say

It is worth marking what is absent from the dominant authentication conversation, because the silences are organized.

There is, first, very little discussion of *whether the problem as defined is the right problem*. The authentication crisis is consistently framed as: AI generates indistinguishable fakes; we must build provenance to detect them. The framing presupposes that the appropriate response to synthetic media is a technical one. There is barely any serious public discussion of non-technical responses — slowing the deployment of generative tools, requiring liability for foreseeable misuse, restricting certain uses of synthetic likeness, redirecting public investment toward independent journalism rather than toward detection systems. Even sober pieces asking whether schools should ban AI altogether are framed as eccentric rather than as one rational option among several [24].

[24] ¿Deberían las escuelas prohibir la IA? La pregunta que no podemos ...

There is, second, very little voice in the discourse from the people most affected. The discourse is conducted by vendors, regulators, academics, and journalists. It is conducted *about* job applicants, students, gig workers, immigrants, protesters. The asymmetry is structural: the subjects of authentication systems do not own the systems, do not see the criteria, do not have access to the decisions made about them, and do not have standing to compel disclosure. As Benjamin observes about identity systems generally, even where the surface promises openness, "different valuations for skins, social groups and categories" are quietly being formed [20]. The valuations are not chosen by those they govern.

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

There is, third, almost no engagement with the labor question specific to this turn. Teaching contracts, for instance, have barely begun to address how AI authentication tools change the conditions of teachers' work — what counts as a teacher's evaluation, what claims a school can make on a teacher's drafts, who owns the lesson-plan provenance trail [6]. The same is broadly true across professions. The infrastructure is being installed faster than the social contracts around it can be renegotiated, and the vendors prefer it that way.

[6] AI Is Changing Teaching, But Few Labor Contracts Reflect It

There is, fourth and most quietly, a gap between the geographies of authentication's costs and the geographies of its language. Most editorial discussion of provenance is conducted in English, in the United States and the European Union, by professionals employed in jurisdictions where there is at least some prospect of legal recourse. The infrastructure, however, is global. A pilot in Jodhpur is running AI-generated assessments across seventy thousand students in a thousand

schools, with results returning in seconds [1]. The students whose academic futures are being shaped by these systems are not party to the consortium meetings in San Francisco where the standards are drafted. The authentication economy is being built on populations whose objections will not be in the room.

[1] 70,000 students, 1,000 schools, results in seconds: Jodhpur begins ...

The Politics of "Trust"

The word *trust*, used as a noun in vendor materials, has a specific function: it converts a relational concept into a tradeable asset. In ordinary speech, trust is what people extend to each other based on history, accountability, and the possibility of repair. In the racket's lexicon, trust is what a verification stack issues, what a platform displays, and what a competitor cannot offer because they have not paid the licensing fee. This semantic shift is the deepest stake of the moment.

Zuboff again, more sharply: in the surveillance regime, "democracy and social trust are superseded by the certainty machines, their priests, and their owners" [20]. The phrase "certainty machines" is exact. Authentication systems are sold as certainty about identity, certainty about provenance, certainty about authorship. The certainty they provide is partial, brittle, and deeply contingent on whose model is doing the deciding. But the experience of using them — the green checkmark, the verified badge, the candidate score — feels like certainty, and that feeling is the product.

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

The political consequence is a slow conversion of public epistemic life into a commercial subscription. The civic question *is this true?* is being replaced by the consumer question *which provider's seal does this carry?* These are not equivalent questions. The first one belongs to citizens reasoning together. The second one belongs to a market.

Closing: Naming the Toll Road

The authentication racket should be named for what it is, not because the underlying anxieties are illegitimate — they are not, the synthetic-media problem is real and the harms are real — but because the framing currently dominant in the discourse routes those anxieties into a particular kind of solution that benefits a particular set of incumbents. A reader who finishes this essay should be able to recognize the move. When a vendor says that trust requires their infrastructure, ask who built the infrastructure, who pays the rent on it, who is excluded by it, and what happens when it errs. When a school district adopts an AI

cheating detector, ask which students will be falsely accused and what the appeal process looks like. When a state agency procures a face-matching system, ask which protesters will be the first to be matched and which laws will be reinterpreted to allow it [16].

[16] Private groups work to identify and report student protesters for ...

The provenance question is not going to disappear, and refusing to engage with it is not a serious posture. But the terms on which it is engaged are themselves a political choice. We could, as a public, decide that authentication is too important to be a private market — that it should be governed the way we govern voting registries, weights and measures, or vital records, with public accountability, statutory rights of contest, and the assumption that errors fall on the system rather than on the subject. Or we can let it continue to be built as it is being built now: as a tiered subscription service, with the wealthy and institutional inside the moat and the rest of us outside it, our content and our faces and our resumes flagged-by-default in a system whose criteria we are not allowed to see.

The clarifying observation, finally, is that the people most insistently telling us that trust is in crisis are the same people most insistently selling us the cure. That is not, on its face, evidence of bad faith. But it is evidence of a structural alignment that the editorial discourse should at minimum name. As one peer-reviewed assessment of generative AI's wider socioeconomic impact noted, the gains and the losses of this transition are not evenly distributed, and pretending they are is itself a political act [21]. The authentication economy is the latest enclosure. The job, for readers, is to refuse to mistake the toll booth for the road.

[21] The impact of generative artificial intelligence on socioeconomic ...

References

1. 70,000 students, 1,000 schools, results in seconds: Jodhpur begins ...

2. AI Hiring Bias Exposed in New Study | X0PA AI posted on ...

3. AI Hiring Bias Lawsuits Are Reshaping Recruiting in 2026: What ...

4. AI Hiring Discrimination: How Algorithms Reject Millions of Qualified ...

5. AI Hiring Is Excluding Neurodivergent Candidates—The Data ... - LinkedIn

6. AI Is Changing Teaching, But Few Labor Contracts Reflect It

7. Algorithmic Bias in Education | International Journal of Artificial ...

8. Detecting & Mitigating Bias in AI Grading: A Practical Playbook

9. Detectores de IA acusan falsamente a estudiantes de hacer trampa, con ...

10. Eightfold AI Lawsuit Claims Secret Algorithm Ranking Applicants

11. How AI monitors school Chromebooks and what it means for privacy ...

12. How governments use facial recognition for protest surveillance - Rest ...

13. Le biais d'auto-préférence des IA de recrutement

14. NYPD Bypassed Facial Recognition Ban to ID Pro-Palestinian Student ...

15. Opinion | Would You Be Afraid of IA?

16. Private groups work to identify and report student protesters for ...

17. PROOF POINTS: Asian American students lose more points in an AI essay ...

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

19. Schools use AI to monitor kids, hoping to prevent violence. Our ...

20. The Age of Surveillance Capitalism_ The Fight for a Human Future at the New Frontier of Power

21. The impact of generative artificial intelligence on socioeconomic ...

22. Trump administration joins Musk to take aim at US state AI hiring ...

23. We did the math on AI's energy footprint. Here's the story you haven't ...

24. ¿Deberían las escuelas prohibir la IA? La pregunta que no podemos ...