The Arms Race of Suspicion: How Authentication Tools Are Reengineering Academic Judgment

Weekly Analysis — https://ainews.social

In late April, lawyers for an Adelphi University student filed a complaint that should be read in every dean's office in the country. The student, accused of having used ChatGPT to write an essay, says she did not, and the evidence against her is essentially this: a software vendor's number, a faculty member's intuition trained by that number, and an institutional process that treated the number as probable cause [9]. The case is one of a growing docket; a running tally maintained by independent observers now lists dozens of similar suits and grievances, almost all of them turning on the question of whether a probability score is the kind of thing a university should be allowed to use to ruin an undergraduate's transcript [6].

This is the moment higher education has actually entered, beneath the louder discourse about AI literacy and "transformation." The classroom has acquired a forensic apparatus. Essays go through detectors; detectors emit confidence intervals; confidence intervals produce hearings; hearings produce settlements. What used to be a relationship — instructor reading student, student reading instructor — has been reorganized around a third party that neither of them trusts but that both must now answer to. The Adelphi suit is not an outlier but a leading indicator: the legal system is beginning to do, slowly and expensively, the conceptual work that the academy declined to do at the speed the technology required [3].

What follows is an attempt to map this arms race honestly, in the institution's own terms. Not the optimistic terms of vendor brochures, not the apocalyptic terms of op-eds, but the terms that a university actually inhabits when it decides — quietly, through procurement and policy — that the cheapest answer to a pedagogical question is to buy a piece of software and train its faculty to be its operators.

[9] An Adelphi University student was accused of using AI to ... - Newsday

[6] AI Detection Lawsuits: Every Student Case, Outcome, and What the Data ...

[3] Adelphi University accused a student of using AI to plagiarize. He ...

The Quiet Procurement of Doubt

The instinct to detect is older than ChatGPT, but the scale at which detection has been industrialized is new. Surveys of secondary and post-secondary administrators tracked through 2025 and into 2026 show detection products now embedded in the default workflow of grading at a majority of institutions, with adoption rates climbing steeply at every tier from community colleges to the elite research universities [4]. The faculty member uploads the essay; the platform returns a percentage; the percentage becomes a fact. The phrase "AI-generated" floats free of any defensible technical claim and becomes a verdict.

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

The technical literature on the detectors themselves has, meanwhile, grown blunt to the point of embarrassment. A widely circulated paper published this spring argues that the use of generative-AI detection in higher-education assessment is not merely unreliable but indefensible — that the false-positive rates, the systematic disadvantage to non-native English writers, and the impossibility of independent audit make the whole enterprise unfit for the high-stakes decisions universities are using it to underwrite [13]. Other systematic reviews of writing assessment in the AI era reach a similar conclusion through gentler language: the detection paradigm is incompatible with what writing instruction is actually for, and the fix is not better detectors but different assignments [20].

[13] Contra generative AI detection in higher education assessments

[20] Reimagining Writing Assessment for the AI Era: A Systematic Review on Balancing AI Support and Authentic Skill Growth

The institution, however, does not buy a detector because it believes the detector works. It buys a detector because the detector permits a particular kind of administrative speech — defensible, auditable, scalable speech — about a problem the institution does not wish to address pedagogically. Meredith Broussard's term for this kind of move, the substitution of a software artifact for the harder social work the artifact is supposed to be doing, was *technochauvinism*, and her catalogue of the way that substitution misfires across domains reads now like a pre-write of the present moment [5]. What is being procured is not detection. It is the appearance of due process at a price the budget can absorb.

[5] Artificial Unintelligence

The Forensic Turn in Assessment

To grasp how strange this is, set it next to the way assessment used to be defended. The traditional argument for the essay was that it externalized thought — that you could see, in the rhythm of an undergraduate's sentences, whether she had read what she said she had

read, whether she had thought about it, whether she had earned the conclusion. Reading an essay was an act of professional judgment, undertaken by someone trained to do it. That judgment was fallible but it was *answerable*: the instructor could be asked why, could justify, could be argued with.

The detector cannot be argued with. It produces a number, and the number is a product, and the product was trained on a corpus the instructor cannot inspect, by a vendor whose business model requires the corpus to remain proprietary. A faculty member at the front of an academic-integrity hearing now finds herself in the position of expert witness for evidence she cannot examine and a method she cannot describe. NBC's reporting on undergraduate behavior captures the perverse equilibrium this produces: students, knowing the detectors are unreliable, are running their own writing through "humanizer" tools to lower their detection scores, sometimes paying the same vendors who sell the detectors to schools [26]. The arms race is no longer a metaphor. It is a market, with the university as captive purchaser on one side and the undergraduate as captive purchaser on the other, and the same handful of firms collecting the rents on both ends.

[26] To avoid accusations of AI cheating, college students turn to AI - NBC News

This is what the Probe-4 stake of this essay names: the institution's function shifts from educator to evidence-prosecutor. A Harvard Gazette feature this winter on "preserving learning in the age of AI shortcuts" framed the crisis in the right terms — as a question about what kind of cognitive labor we still believe a degree is meant to certify — but the framing has not yet penetrated the procurement decisions [19]. Faculty are told to "have a conversation" with students about integrity while the integrity office is buying tools that make conversation unnecessary. The pedagogical and the forensic are talking past each other inside the same building.

[19] Preserving learning in the age of AI shortcuts — Harvard Gazette

What the Detection Frame Forecloses

The most damning fact about the detection apparatus is not that it produces false positives, though it does. The most damning fact is what its existence permits the institution to *not* do. As long as detection is plausible, the curricular question — what should an assignment in this discipline look like now that a coherent five-paragraph essay can be generated for free — can be deferred. Detection is the alibi that allows the syllabus to remain unchanged.

The literature on what a serious response would look like is, at this point, abundant and largely unread by the people setting policy. A widely cited MDPI paper proposes redesigning assessment around

process artifacts — drafts, annotations, oral defenses, in-class iteration — rather than the singular polished output that detectors and generators are both trained on [10]. A complementary framework circulating among Australian and UK institutions argues that "authentic assessment" must mean assessment that is constitutively impossible to outsource because it is grounded in the specific evidence of the student's own situated thinking [18]. A working paper out of the open-access repositories proposes treating student work in the AI era as a collaborative artifact with the machine, foregrounding the student's editorial judgment over the machine's generative output [28].

[10] Beyond Detection: Redesigning Authentic Assessment in an AI ... - MDPI

[18] PDF Authentic Assessment in the Age of AI - marcbowles.com

[28] Writing with machines? Reconceptualizing student work in the age of AI

None of this is exotic. All of it is more labor-intensive than uploading a PDF to a vendor's API. That, and not any genuine doubt about its pedagogical merit, is why it is not the default response. Resource-strapped writing programs running adjuncts at fifteen contact hours a section cannot redesign their assessment infrastructure on a semester's notice. The detector, in this light, is not the institution's answer to AI; it is the institution's answer to the cost of teaching writing properly, with AI providing the cover story.

The Governance Layer Eats the Pedagogy Layer

Look at the headlines from the last quarter and a pattern emerges that the discourse has not quite named: governance is moving fast, pedagogy is moving slowly, and the gap between them is being filled by vendors. The California State University system signed a system-wide deal with OpenAI to put ChatGPT Edu in front of every student and every faculty member, a procurement decision made above the pedagogical level and now being absorbed, with visible discomfort, by the academics expected to use it [11]. Arizona State has rolled out an "AI course builder" that lets faculty generate course materials at scale, with faculty publicly worrying that the tool's defaults will collapse the deliberative work of syllabus design into vendor-shaped templates [14]. The University of Surrey has announced that AI will be embedded in every degree from September 2026, framed as a pedagogical commitment but executed, in the university's own announcement, as an institutional procurement [23].

[11] Cal State struck a deal with OpenAI. Some students and ...

[14] Faculty Concerned About ASU's New AI Course Builder

[23] Surrey embeds AI in every degree from 2026

The OpenAI product page for ChatGPT Edu reads, with admirable clarity, as a B2B governance pitch: SOC 2 compliance, data controls, admin dashboards, a workspace the institution can monitor [12]. It is not, primarily, a teaching tool; it is an enterprise procurement aimed at the office of the CIO. The pedagogical claims are downstream of the compliance claims. Institution-branded variants like UC Irvine's

[12] ChatGPT Edu at OpenAI - OpenAI Help Center

ZotGPT follow the same logic — a campus license, a campus interface, a layer of governance assurance laid over a frontier model that the campus does not control [1].

[1] #AnteaterIntelligence: Designing Smarter Classes with ZotGPT

This is a familiar institutional shape: governance becomes the vehicle by which the institution can claim to be doing something while postponing the question of what it is actually doing. A widely circulated framework for "AI leadership in education" promises to scale safely through governance design, and reads, in places, like the kind of document that exists so that nothing has to change in the classroom [7]. A Canadian policy paper takes this further and argues, accurately, that AI is functioning as an institutional response to a pre-existing crisis of retention and risk — a way of automating the management of students rather than the education of them [21]. The detector and the chatbot are the two ends of the same pincer: surveillance on one side, productivity on the other, both routed around the academic.

[7] AI Leadership in Education: A Governance Framework to Scale Safely

[21] Risk, Retention, and the Algorithmic Institution: Artificial Intelligence as a Policy Response to Higher Education in Crisis

Kate Crawford's argument that AI infrastructures are continuous with state and corporate surveillance regimes — that the systems we treat as neutral computation are in fact heavily contoured by who built them, who pays for them, and what they were originally for — has a sharp application here that the higher-education conversation has barely begun to register [5]. When a campus signs a license with a frontier-model vendor, it is not merely buying a productivity tool. It is integrating its assessment, advising, and curriculum-design workflows into the data architecture of a private firm whose strategic interests are not the university's, and whose corpus, model weights, and update cycle are outside the institution's governance entirely.

[5] The Atlas of AI

The Inequity Already Built In

The detection apparatus does not fail randomly. Its failures have a demographic shape. Multiple peer-reviewed studies and surveys this year have documented systematically higher false-positive rates for non-native English writers, for students writing in a register adjacent to the genres the detectors were trained on, and for students whose prose has been edited by accessibility tools or grammar assistants of the kind that universities themselves have promoted [4]. The pattern is not new. It belongs to a longer history of automated decision systems whose statistical confidence was uncorrelated with the lives they ranked, and whose errors fell predictably on populations already vulnerable to institutional misreading [5].

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

[5] Automating Inequality

Race After Technology, Ruha Benjamin's name for the cluster of dynamics by which ostensibly neutral tools encode historical injustice,

captures the asymmetry visible in the Adelphi-style cases: a Black or international or first-generation student accused on the basis of a probability score has fewer resources to mount the kind of counter-investigation a wealthier or better-networked student can afford [5]. The detector's output is the same; the consequences are not. The university is using a tool whose error structure mirrors the inequities the institution claims, in its mission statements, to be redressing.

[5] Race After Technology

The discipline-specific evidence is grimmer still. Comparative analytics studies of how students seek help from chatbots versus human tutors find that the patterns of use diverge sharply along socioeconomic and language lines, with the students the institution most wants to retain often the ones whose interactions with AI tools — both as users and as targets of detection — produce the most distorted institutional reads of their work [27]. When the institution then uses detection scores to triage academic-integrity referrals, it compounds the distortion. The system is not merely fallible. It is fallible in ways that recapitulate the institution's existing failures.

[27] Unpacking help-seeking process through multimodal learning analytics:A comparative study of ChatGPT vs Human expert

The Pedagogical Alternative the Discourse Keeps Deferring

There is a counter-current in the literature, and it has been there from the start, but it is harder to procure. A growing body of work — the AACSB's recent essay on the integrity dilemma, the open-access "Teaching and Generative AI" volume, the series of pieces from writing-studies and assessment researchers — argues that the only stable response to generative AI in student work is to redesign the student's relationship to writing as a process, not a product [25]. The product can always be generated; the process cannot, or at least not cheaply, and the assignment can be designed so that the process is what is assessed [24].

[25] The AI Dilemma: When Innovation Outpaces Integrity | AACSB

[24] Teaching and Generative AI

A French essay this season makes the underlying epistemic point with a useful clarity: AI can produce, but it cannot yet judge — and the work of an undergraduate education, properly understood, has always been to learn to judge, with production as the trace evidence of the judgment [17]. If the production is what we test, of course we have a crisis; the production has been commodified. If the judgment is what we test, the crisis recedes into a curriculum problem, demanding but tractable. UNESCO's competency framework for teachers, in its own dryer language, says the same thing: in education, where most AI applications are by their nature high-risk, the institution's task is independent validation of what these systems do, not deference to their outputs [5].

[17] L'IA sait tout produire… mais pas encore juger

[5] AI competency framework for teachers - UNESCO

This is the missing collaborative framing the present discourse keeps gesturing at and failing to inhabit. It is also harder than detection in exactly the way teaching is harder than grading. It requires that faculty be given time, training, and pedagogical authority to redesign their courses, which requires that institutions stop spending the relevant budgets on detection contracts and start spending them on professional development. The Probe-4 stakes question — whether we want an institution that *educates* or one that *prosecutes* — is, at the operational level, a question about which line in the budget the AI money goes into.

When Authentication Replaces Authority

Three further data points illustrate how thoroughly the authentication frame has begun to substitute for academic authority. First, MIT Sloan's reporting on "persuasion bombs" — the ways generative models exploit user susceptibility through tone, flattery, and confidence calibration — describes a class of failure that detectors are constitutively unable to address; the model that produces a persuasive false answer leaves no fingerprint of falsehood, only of fluency, which is the trait detectors are trained to flag as suspect [15]. The detector cannot tell the truth from the lie; it can only tell the smooth from the rough. Authentication tooling, in other words, mistakes register for honesty.

[15] How generative AI 'persuasion bombs' users

Second, the South African government's national AI policy was found this year to have cited fake research that the policy's own drafters had generated with AI — a perfect demonstration of what happens when the institution's authentication apparatus is itself the system being authenticated [22]. If a national ministry can publish hallucinated citations under its own seal, the question of whether an undergraduate's footnote is "AI-generated" begins to look like the wrong question to be litigating with such ferocity.

[22] South Africa's AI policy cited fake research, created by AI

Third, the University of Staffordshire course taught largely by AI provoked a now-famous student revolt — "we could have asked ChatGPT" — that should be read as an authentication crisis from the other direction [2]. When the institution itself outsources its instructional voice to a model, the student's question is no longer "how do I prove I wrote this" but "how does the institution prove it taught me?" The authentication arrow runs both ways. The student demanding human teaching from a university that has substituted a chatbot is making the same argument the faculty member ought to be making against the detection vendor: the relationship is supposed to be with a person, and a number is not a person.

[2] 'We could have asked ChatGPT': students fight back over course taught by AI

The economic context tightens the screw. Reporting on entry-level labor markets this spring suggests AI is already foreclosing the bottom rungs of professional employment that universities exist, in part, to feed [8]. The institution that replaces its judgment with a vendor's score is producing graduates whose own first encounter with the labor market is a vendor's score. The pedagogy of suspicion trains a generation to expect, and to game, the authentication systems they will spend their working lives inside.

[8] AI won't kill your job — it will kill the path to your first one

What an Institution Could Choose Instead

The argument here is not that the university should ignore generative AI in student work. It is that the institution has a choice between two postures, and it is currently making the wrong one by default.

The first posture is the one we have inventoried: detection contracts, governance frameworks, vendor licenses, and an academic-integrity apparatus retooled as a forensic operation. It is cheap in the short run and ruinous in the long run, because it teaches every student that the institution does not trust them and every faculty member that judgment has been delegated upward. Education Week's reporting on the parallel debate over AI grading — whether faculty themselves should use AI to mark student work — captures the symmetry: an institution that does not trust its students to write without AI is also, increasingly, an institution that does not trust its faculty to grade without it [16]. The endpoint of the authentication race is mutual surveillance between people the institution claims to be the relationship between.

[16] Is It Ethical to Use AI to Grade? - Education Week

The second posture is harder and has been clearly described, repeatedly, in the literature this essay has cited. It begins with the admission that detection is unfit for high-stakes academic-integrity decisions and that its continued use is an injustice [13]. It proceeds with assessment redesign that puts the student's process — drafts, oral defenses, in-class iteration, situated reasoning — at the center of what the institution evaluates [10]. It funds professional development at the scale this redesign actually requires, which is not the scale of a memo. And it treats the question of AI in the curriculum as a pedagogical question owned by faculty, not a procurement question owned by the CIO, with vendor relationships subordinated to the academic mission instead of structuring it.

[13] Contra generative AI detection in higher education assessments

[10] Beyond Detection: Redesigning Authentic Assessment in an AI ... - MDPI

This is not a technological program. It is an institutional one. The technology will keep doing what technology does — improving on some axes, degrading on others, generating scores that vendors will sell

as the truth. What an institution can decide is whether it metabolizes the technology through its pedagogy or substitutes the technology *for* its pedagogy. The detection arms race is the visible artifact of that choice being made, mostly without deliberation, in the wrong direction.

The Adelphi suit will be settled or it will not. Other suits will follow; the running list of cases makes that mathematically certain [6]. What will not settle, absent a decision the academy has so far avoided, is the underlying question. An academic integrity grounded in suspicion is not academic integrity. It is the bureaucratic residue of a pedagogical relationship the institution has stopped maintaining. The arms race of suspicion is winnable in the short run only by the side that sells the weapons, and the longer the academy stays in it, the less of itself it will have left to defend when the next wave of tools arrives. The choice the university actually faces is not which detector to buy. It is whether to remain the kind of place where the question of what a student knows is answered by another human being, on the record, in conversation, with reasons that can be argued with — or to become a place where that question is answered by a vendor's number, and the conversation is over before it starts.

[6] AI Detection Lawsuits: Every Student Case, Outcome, and What the Data ...

References

1. #AnteaterIntelligence: Designing Smarter Classes with ZotGPT
2. 'We could have asked ChatGPT': students fight back over course taught by AI
3. Adelphi University accused a student of using AI to plagiarize. He ...
4. AI Cheating in Schools: 2026 Global Trends & Bias Risks
5. AI competency framework for teachers - UNESCO
6. AI Detection Lawsuits: Every Student Case, Outcome, and What the Data ...
7. AI Leadership in Education: A Governance Framework to Scale Safely
8. AI won't kill your job — it will kill the path to your first one
9. An Adelphi University student was accused of using AI to ... - Newsday
10. Beyond Detection: Redesigning Authentic Assessment in an AI ... - MDPI

11. Cal State struck a deal with OpenAI. Some students and ...

12. ChatGPT Edu at OpenAI - OpenAI Help Center

13. Contra generative AI detection in higher education assessments

14. Faculty Concerned About ASU's New AI Course Builder

15. How generative AI 'persuasion bombs' users

16. Is It Ethical to Use AI to Grade? - Education Week

17. L'IA sait tout produire… mais pas encore juger

18. PDF Authentic Assessment in the Age of AI - marcbowles.com

19. Preserving learning in the age of AI shortcuts — Harvard Gazette

20. Reimagining Writing Assessment for the AI Era: A Systematic Review on Balancing AI Support and Authentic Skill Growth

21. Risk, Retention, and the Algorithmic Institution: Artificial Intelligence as a Policy Response to Higher Education in Crisis

22. South Africa's AI policy cited fake research, created by AI

23. Surrey embeds AI in every degree from 2026

24. Teaching and Generative AI

25. The AI Dilemma: When Innovation Outpaces Integrity | AACSB

26. To avoid accusations of AI cheating, college students turn to AI - NBC News

27. Unpacking help-seeking process through multimodal learning analytics:A comparative study of ChatGPT vs Human expert

28. Writing with machines? Reconceptualizing student work in the age of AI