Trained for the Tool, Not for the Question: The Literacy Crisis Behind the AI-Scam Panic

Weekly Analysis — https://ainews.social

Somewhere this week, a well-meaning public-safety bulletin told you that AI-enabled crime is exploding — the figure most often quoted is a 148 percent surge — and that the responsible thing to do is to slow down, call your bank back on a number you trust, and never believe a voice on the phone just because it sounds like your daughter. The advice is not wrong. Voice-cloning fraud is real, and the criminal economy around it is maturing quickly, as the industry mapping in reports like [13] makes clear. But notice what the bulletin does to you as it protects you. It turns you into a lone sentry. It hands you a checklist and calls the checklist "literacy."

[13] El auge de la delincuencia facilitada por la IA - Informe

This is the quiet inversion at the heart of the current moment. The most visible, most funded, most widely distributed AI-literacy materials of the past year are, overwhelmingly, scam warnings and detection tips. They teach a citizen to spot a deepfake, to verify a caller, to prompt a chatbot more cleverly, to distrust a video until proven otherwise. Meanwhile, the numbers that justify all this vigilance circulate with impressive decimal precision and almost no visible methodology — the kind of statistical fog that a genuinely literate public would be equipped to interrogate, and that the current literacy curriculum trains us instead to repeat. The 2026 edition of the field's most-cited scoreboard, [27], documents the scale of AI's diffusion into daily life, but the diffusion of alarming crime statistics has outpaced the diffusion of any means to evaluate them.

[27] The 2026 AI Index Report | Stanford HAI

My argument is that "AI literacy" has been quietly narrowed — from a civic capacity into a personal-defense skill — at precisely the moment it needed to expand. The literacy this moment demands is not private skepticism, the mental habit of a consumer guarding a wallet. It is a collective understanding of who builds these systems, who profits from them, who is excluded by them, and who gets to decide what they are for. To see how far the narrowing has gone, we first have to notice that the word "literacy" is doing at least two incompatible jobs at once.

The Two Faces of a Single Word

Ask a dozen institutions what AI literacy means and you will get two families of answer that sound similar and point in opposite directions. The first family is operational: literacy as competent use. It is the vision you find in most teaching guides — the ability to describe what a model does, to write effective prompts, to recognize hallucination, to fact-check an output. Stanford's own overview, [10], organizes literacy largely around understanding capabilities and using tools responsibly. This is genuinely useful. It is also, tellingly, a definition that a vendor could love: it locates the entire problem inside the user's head and hands, and never once looks up at the company that shipped the system.

[10] Comprender la alfabetización en IA | Teaching Commons

The second family is structural: literacy as the capacity to understand AI as a social and political arrangement, not just a gadget. This is the tradition that Kate Crawford insists on when she writes, in [28], that once we connect AI to broader social systems we "escape the notion that artificial intelligence is a purely technical domain" — that AI is, at a fundamental level, "technical and social practices, institutions and infrastructures, politics and culture." A person literate in that sense does not merely detect a deepfake; they ask who owns the model that made it, whose faces trained it, and which institutions profit from a general climate of doubt.

[28] The Atlas of AI

The problem is not that one definition is right and the other wrong. Both are needed. The problem is the ratio. When you survey what actually reaches people — the bulletins, the tip sheets, the mandatory modules — the operational definition swamps the structural one almost completely. The more careful assessments know this is a problem. The field's own diagnostic work, gathered in the ETS review [24], acknowledges how hard it is to measure the critical, judgment-heavy dimensions of literacy — and how easily assessment drifts toward the countable skill and away from the harder question. What gets measured gets taught; what is hard to measure quietly vanishes from the definition.

[24] PDF Opportunities and Challenges for Asessing Digital and AI Literacies - ETS

The MIT Press primer on the subject gestures at the richer alternative. In [28], the authors argue that "doing AI" or "doing data science" should simply include ethics as part of the practice, and imagine a "more diverse and holistic kind of Bildung" — a formation of the whole person, radically interdisciplinary in its methods and its media. That word, Bildung, is the tell. It marks the distance between training and formation: between learning to operate a tool and learning to live, as a citizen, in a world the tool is remaking. Almost everything sold as AI literacy this year is training. Almost none of it is Bildung.

[28] AI Ethics

Precision Without Method

Start with the numbers, because the numbers are where the narrowing does its most elegant work. The scam-and-fraud genre runs on statistics — a 148 percent rise here, an X-fold increase in synthetic-media crime there — and these figures travel with a confidence that their sourcing rarely earns. The underlying reports, including the criminal-economy analysis in [13], are often serious pieces of work, but by the time a number reaches a public bulletin it has been stripped of its denominators, its baselines, its definitions of what counts as "AI-enabled." A percentage without a base is not information; it is atmosphere.

[13] El auge de la delincuencia facilitada por la IA - Informe

This matters more than pedants usually claim, because the scholarly literature on AI deception is itself careful to distinguish categories that the panic collapses. The survey [2] draws lines between systems that deceive as a designed function, systems that deceive as an emergent side effect, and humans who deceive using AI tools — three very different problems with three very different remedies. A literate public would hold those distinctions. The scam bulletin erases them, folding everything into a single undifferentiated threat that only heightened personal vigilance can meet.

[2] AI deception: A survey of examples, risks, and potential solutions

Here is the exquisite irony. The genre that presents itself as literacy — as the thing that will make you a smarter, safer information consumer — routinely commits the exact epistemic failure it claims to cure. It asks you to distrust a viral video while trusting an unsourced statistic. It trains suspicion of the image and credulity toward the number. Even the most authoritative aggregations, like [27], model good practice precisely by showing their methods — and the gap between that transparency and the bulletins that cite such reports is where literacy is actually lost. The material meant to build critical judgment instead rehearses the reader in accepting authoritative-sounding claims at face value. That is not a marginal defect. It is a literacy failure embedded inside literacy training itself.

[27] The 2026 AI Index Report | Stanford HAI

Crawford has a name for the mechanism that makes this work: "enchanted determinism," the framing in [28] that tells the public to "focus on the innovative nature of the method rather than on what is primary: the purpose of the thing itself." Enchantment, she writes, "obscures power and closes off informed public discussion, critical scrutiny, or outright rejection." The scam-panic statistic is enchanted in exactly this way. It dazzles with magnitude, forecloses the question of who assembled it and why, and leaves the reader with nothing to do but be afraid and be careful — the two emotions least likely to produce a demand for structural change.

[28] The Atlas of AI

The Verification Reflex and Its Hidden Costs

Follow the advice these materials give and you arrive at a single instruction repeated in a hundred registers: verify everything. Do not trust the caller; call back. Do not trust the video; check the source. Do not trust the chatbot; confirm elsewhere. Taken one at a time, each instruction is sound. Taken together, as a theory of how a society should function, they describe something close to a nightmare: a world in which the burden of authenticating every message falls entirely on the individual receiving it, and in which trust — the connective tissue of any working public — has been privatized into a personal risk-management task.

UNESCO's framing is unusually honest about the stakes, if not about the remedy. Its essay [11], and its French counterpart [17], name the real danger correctly: not that any single fake will fool us, but that the ambient possibility of fakery corrodes the very idea that shared knowledge is possible. That is a collective, epistemic injury. And yet the response on offer keeps collapsing back into individual technique — teach people the tells, teach them to slow down — as though a crisis of shared knowing could be solved one careful viewer at a time.

[11] Deepfakes and the crisis of knowing

[17] Les deepfakes et la crise du savoir - UNESCO

The costs of the verification reflex are not merely philosophical; they are concrete, and they fall unevenly. Consider the detection tools that the verify-everything ethos inevitably summons. They do not work reliably, and the people they fail are not random. Reporting collected in [4] documents how detectors are deployed despite known unreliability, and the investigation [5] shows the false positives clustering on non-native English writers — people whose prose the machine reads as "too perfect" to be human. The legal-research overview [18] catalogs the same structural unreliability. When verification is automated, it does not become neutral; it acquires a bias and then launders that bias as objectivity. A literacy worthy of the name would teach people to be suspicious of the detector, not just the deepfake.

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

[5] AI Detection Tools Falsely Accuse International Students of Cheating

[18] Los problemas de los detectores de IA: falsos positivos y ...

And the verification reflex has a cost that its promoters never mention: it teaches distrust as a lifestyle. The prompt-injection literature, surveyed in accessible form in [25], demonstrates that even the systems we are told to consult for verification can be manipulated through their own inputs — that the "check elsewhere" instruction assumes an "elsewhere" that is itself compromised. At some point the reasonable citizen realizes that they cannot personally authenticate the informational world, that no individual can, and that the entire premise of literacy-as-vigilance was a category error. The authentication of public communication is an infrastructural responsibility. It belongs to plat-

[25] Prompt Injection Attacks: Examples and Defences

forms that profit from distribution, to telecoms that carry the calls, to states that regulate both. Handing it to the individual and calling that "empowerment" is one of the more successful sleights of hand of the decade.

When the State Answers with Surveillance

If the market's answer to the crisis of knowing is a tip sheet, the state's answer is increasingly a monitoring system — and the difference between those two responses tells you almost everything about how power understands literacy. Faced with the prospect of synthetic media distorting elections, governments have reached, by reflex, for detection and surveillance rather than for the slower work of civic education and trust repair.

The election-year literature makes the pattern legible. Analyses like [15] and the French municipal coverage in [20] frame generative AI primarily as a threat vector to be countered, monitored, flagged. In the United States, the survey [3] documents a wave of state laws aimed at synthetic political media — a legislative energy directed almost entirely at detection and disclosure mandates rather than at building a public capable of situating such media in a wider understanding of who owns the megaphone. The instinct everywhere is the same: install a verification apparatus over the electorate rather than cultivate judgment within it.

[15] IA générative et élections européennes : opportunité ou péril ?

[20] Municipales 2026 : IA, deepfakes et désinformation, la démocratie ...

[3] AI Deepfakes Are Flooding the 2026 Midterms. 26 States Scramble

This is where the choice of response becomes a choice about citizenship. A state that answers a civic-epistemic crisis with a monitoring system has decided that its people are threats to be screened, or victims to be shielded, rather than participants to be equipped. The alternative exists and is legible: the Brennan Center's [6] maps ways AI could deepen participation — could help people understand and shape governance — rather than merely defend elections from contamination. That the participatory vision is so much quieter than the surveillance vision is itself a data point about whose literacy the state considers worth building.

[6] Artificial Intelligence, Participatory Democracy, and Responsive Government

Even the more educational state responses tend to route AI back through the operational, tool-centered definition. France's official framework, [8], is a serious attempt to structure appropriate use — and it is, characteristically, a document about use: about deploying tools well and safely. Nothing in it is wrong. But a framework organized around usage cannot, by its nature, ask the prior question that Crawford insists on in [28] — the question of "the purpose of the thing itself," of the power relations the system encodes before any citizen

[8] Cadre d'usage de l'IA en éducation

[28] The Atlas of AI

ever touches it. When the state's imagination of literacy stops at competent, safe use, it has already conceded that the architecture of AI is settled and only individual conduct within it remains to be managed. Surveillance for the many, use-guidance for the diligent, and for no one the standing to ask whether the system should exist in its current form at all.

The Vulnerable as a Test of the Whole Framework

Nothing exposes the limits of literacy-as-personal-defense more sharply than the populations for whom personal defense is least available. Children, adolescents, the cognitively vulnerable, the elderly targeted by voice-cloning — these are the people the scam-panic framework claims to protect, and they are precisely the people its individualized model cannot reach. You cannot tip-sheet your way to safety for someone who lacks the developmental capacity, the linguistic fluency, or the sheer time to run the verification protocol on every incoming message.

The evidence here is stark and growing. The American Psychological Association's [7] treats adolescent exposure to AI systems as a matter of health, not merely of skill — a framing that quietly concedes the inadequacy of the checklist. Research into [9] shows that the developmental capacity to distrust a fluent, confident machine is not evenly distributed and cannot simply be assumed. And the survey work in [22] documents a generation forming its relationship to trust and information under conditions no prior cohort faced — and forming it, crucially, before the institutions around them have decided what protection they are owed. The United Nations framing in [19] puts the sequencing plainly: children are growing up inside these systems before the systems have been made accountable to them.

[7] Aviso de salud: Inteligencia artificial y bienestar adolescente

[9] Children's susceptibility to content generated by artificial ...

[22] PDF 2025 Teens, Trust, and Technology in the Age of AI - Common Sense Media

[19] Millones de niños crecen con la inteligencia artificial antes de que ...

The design side is worse than the exposure side, because it reveals harm that no individual literacy could ever forestall. Work on [1] and the Australian regulator's analysis [14] document how generative systems are being turned to the mass production of exploitation material — a harm located entirely in the architecture and deployment of the models, not in the vigilance of any potential victim. The engineering response, such as the detection approach described in [21], is welcome and necessary, but it is a structural intervention — a change to the system — which is exactly the point. No amount of "teach children to spot fakes" touches a harm manufactured by the system itself.

[1] Abuse at scale: Artificial intelligence is reshaping child ...

[14] Generative AI and child safety: A convergence of ...

[21] New method aims to keep kids safe from illegal AI- ...

There is a further, colder finding that should end the individual-defense model outright. The study reported in [26] shows that these

[26] Study: AI chatbots provide less-accurate information to ...

systems deliver worse information to the users least equipped to detect the degradation — that the machine's failures are correlated with, and therefore compounded by, the vulnerability of the person facing them. This is the precise inverse of what a literacy-as-personal-skill model assumes. That model assumes the burden can be met by the individual and that meeting it improves with effort. The reality is that the systems fail hardest against exactly those who can least carry the burden, which means the burden was never theirs to carry in the first place. The gendered dimension UNESCO underscores in [28] — that women's unequal access to digital technology "hinders their access and opportunity to express themselves and participate fully" — is one more instance of the same structure: literacy that presumes an already-empowered user cannot help the user who was never given power to begin with.

[28] UNESCO Think Critically Click Wisely

The Questions Not Being Asked

There is a canonical set of questions that the critical tradition has been asking about technology for years, and the strange thing about this year's literacy corpus is how completely those questions have gone missing from the materials most people will ever see. The MIT Press primer [28] frames them with almost brutal economy — who has access, who benefits, who is excluded — as the questions that any real formation in AI must hold. They are present in the scholarship. They are absent from the tip sheet.

[28] AI Ethics

Ask them, and the whole picture reorganizes. Who benefits when literacy is defined as personal vigilance? The platforms and vendors do, because a public trained to guard itself is a public that never thinks to demand the guardrail be built into the product. Who is excluded when detection is automated? The people the detector misreads — a bias no longer hypothetical but litigated, as the survey [16] documents, tracking a rising body of law around systems that discriminate in hiring, lending, and beyond. A literacy that never mentions these lawsuits has decided that the citizen needs to know how to spot a fake video but not how to recognize an algorithm quietly deciding their creditworthiness. That is a curriculum built around the harms that keep you scrolling and away from the harms that shape your life.

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

Who has access to the understanding required to contest any of this? Here the corpus contains its own rebuke. Documents like the Argentine [23] show what it looks like to organize literacy around responsibility and transparency — around the obligations of the deployer as much as the caution of the user. Notice how rare that inversion is.

[23] PDF Manual de uso responsable y transparencia de la Inteligencia Artificial ...

Most materials address the receiver of AI; a genuinely civic literacy would spend at least as much energy on the builders, the buyers, the institutions procuring these systems on the public's behalf without the public's understanding.

The deepest question the tip sheet cannot ask is the one about definition itself: who decides what counts as literacy, and whose literacy counts. When a bank defines literacy as your ability to detect its fraud liability, literacy becomes an instrument of the bank. When a state defines it as compliance with a monitoring regime, literacy becomes an instrument of the state. When a vendor defines it as skilled use of the vendor's product, literacy becomes marketing. The reason this matters is that literacy has always been, in every prior information revolution, a site of power — a threshold that some are helped across and others are kept behind. To let the definition be set by the very institutions whose power the literacy should be able to scrutinize is to guarantee a literacy that scrutinizes everything except them.

Literacy as a Collective Capacity

The way out is not to abandon the useful skills. Learning to recognize a hallucination, to slow down before a suspicious call, to read a synthetic image with a skeptical eye — these are real goods, and in a year of accelerating fraud they are worth teaching. The error is not the skill. The error is mistaking the skill for the whole, letting personal defense stand in for civic understanding, and thereby leaving the architecture of AI unexamined while congratulating the public on its new alertness.

The more honest voices in the field already sense the shortfall. The field-listening exercise in [12] keeps returning to a gap between what literacy programs deliver and what participation in a democracy actually requires — between operating the tool and understanding the order the tool belongs to. Even the data on how people are actually reaching for these systems, gathered in [29], shows adoption racing ahead of any shared framework for judging what the adoption means. Usage is not understanding, and the gap between them is where a public becomes managed rather than sovereign.

[12] Digital Literacy in the Age of AI: Voices from the Field

[29] Uso de la IA generativa por parte de jóvenes y docentes ...

What would a collective literacy look like? It would treat the questions from [28] — access, benefit, exclusion — not as an advanced module for specialists but as the first thing anyone learns, because they are the questions that turn a user into a citizen. It would teach people to read a statistic's methodology as readily as they read a video's artifacts. It would insist, with Crawford in [28], that AI is "in-

[28] AI Ethics

[28] The Atlas of AI

stitutions and infrastructures, politics and culture," and that understanding it therefore means understanding power, not just probability. And it would locate responsibility where the harm is manufactured — with the platforms, the telecoms, the states, the vendors — rather than dispersing it, atomized and unwinnable, across a population of anxious individual sentries.

The scam panic will pass, or rather it will be replaced by the next panic, and the one after that. What will remain is the definition of literacy we allow to harden while the panics distract us. If we let it harden around personal defense, we will have trained a generation to guard itself expertly against symptoms while never acquiring the standing to ask about the disease. Trained for the tool, not for the question — and a public that cannot ask the question is not, in any sense the word once carried, literate at all.

References

1. Abuse at scale: Artificial intelligence is reshaping child ...
2. AI deception: A survey of examples, risks, and potential solutions
3. AI Deepfakes Are Flooding the 2026 Midterms. 26 States Scramble
4. AI detection tools are unreliable. Teachers are using them anyway : NPR
5. AI Detection Tools Falsely Accuse International Students of Cheating
6. Artificial Intelligence, Participatory Democracy, and Responsive Government
7. Aviso de salud: Inteligencia artificial y bienestar adolescente
8. Cadre d'usage de l'IA en éducation
9. Children's susceptibility to content generated by artificial ...
10. Comprender la alfabetización en IA | Teaching Commons
11. Deepfakes and the crisis of knowing
12. Digital Literacy in the Age of AI: Voices from the Field
13. El auge de la delincuencia facilitada por la IA - Informe
14. Generative AI and child safety: A convergence of ...
15. IA générative et élections européennes : opportunité ou péril ?

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

17. Les deepfakes et la crise du savoir - UNESCO

18. Los problemas de los detectores de IA: falsos positivos y ...

19. Millones de niños crecen con la inteligencia artificial antes de que ...

20. Municipales 2026 : IA, deepfakes et désinformation, la démocratie ...

21. New method aims to keep kids safe from illegal AI- ...

22. PDF 2025 Teens, Trust, and Technology in the Age of AI - Common Sense Media

23. PDF Manual de uso responsable y transparencia de la Inteligencia Artificial ...

24. PDF Opportunities and Challenges for Asessing Digital and AI Literacies - ETS

25. Prompt Injection Attacks: Examples and Defences

26. Study: AI chatbots provide less-accurate information to ...

27. The 2026 AI Index Report | Stanford HAI

28. The Atlas of AI

29. Uso de la IA generativa por parte de jóvenes y docentes ...