The Scam Alert Reads the Future: Literacy, Reduced to Self-Defense

Weekly Analysis — https://ainews.social

A voice on the phone sounds exactly like your daughter, and it is asking for money. That is not a hypothetical anymore; it is a product category. Spanish security reporters now describe, almost as a matter of routine consumer advice, how scammers clone a familiar voice from a few seconds of audio to empty a bank account [16]. Fact-checkers walk readers through the same mechanics with the patience of a driver's-ed instructor: here is how the deepfake is built, here is the tell, here is what to do before you wire the funds [6]. The security-analysis industry has its own genre now — the anatomy of a synthetic bank heist, reconstructed for professionals who need to defend against the next one [11].

[16] Los estafadores ya clonan tu voz con inteligencia artificial para vaciar tu cuenta bancaria

[6] Cómo se usa la inteligencia artificial (IA) en 'deepfakes' para estafar

[11] How Deepfakes Can Break Finance and Banking: Real Attacks, Real Money, Real Defense

Read a week of this material and a pattern hardens. Almost everything published under the banner of AI literacy this week is, at bottom, a warning. It teaches you to distrust a voice, to squint at a video, to hover before you click, to verify before you believe. This is not nothing. People are genuinely being defrauded, and the advice is genuinely useful. But watch the move being made underneath the usefulness: a word that used to mean a broad civic capability — the ability to read a situation, understand how it works, argue about it, and help decide what should be done — is quietly being redefined to mean personal threat-detection. Literacy is contracting into self-defense.

I want to map that contraction carefully, because it is easy to mistake for progress. When a public agency or a newspaper teaches you how a scam works, it feels like empowerment; you know something you did not know before. And yet the knowledge you are handed is shaped almost entirely as a defensive crouch. You learn to be a harder target. You do not learn who built the cloning tool, who profits from its distribution, who decided the platform should recommend it, or how — as a citizen rather than a mark — you might contest any of that. The structural questions stay off the page. What follows is an attempt to say precisely what kind of literacy is being offered to us this week, and what a richer one would have to include.

The Week Arrives as a Warning

Begin with the sheer volume of alarm, because the volume is the argument. The synthetic-media beat has stabilized into a set of recurring threat categories, each with its own instructional literature. There is fraud: the cloned-voice call, the deepfaked video of a bank executive authorizing a transfer. There is medical disinformation, where readers are given ten signals for telling AI-fabricated health content from evidence [5]. There is electoral manipulation, where the deepfake enters as a threat to the democratic process itself, and the citizen is enlisted to spot it [18]. And there is the meta-category, fake news writ large, catalogued case by case as the year's parade of forgeries [19].

[5] Cómo reconocer contenidos médicos falsos generados por inteligencia artificial: 10 señales para distinguir evidencia de desinformación

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

[19] Noticias Falsas con IA en 2025: Casos y Detección — Fake Off

Each of these is a real harm. What unites them, rhetorically, is the posture they assign to you. In every case you are positioned as a potential victim who must become a vigilant consumer of signals. France's data-protection authority publishes a guide to deepfakes framed almost entirely as a protection-and-reporting protocol — how to shield yourself, how to file a complaint [13]. The advice is sound. But notice that the horizon of action it imagines is limited to the individual's defensive repertoire: recognize, avoid, report. The verbs never reach for design, ownership, or governance. You may report a forgery after it has reached you; you are not invited to ask why the pipeline that delivered it exists in the shape it does.

[13] Hypertrucage (deepfake) : comment se protéger et signaler les contenus

There is a quieter irony here that a programmer friend of Meredith Broussard once put unwittingly into words. After the 2016 election flooded the culture with fake news, he asked her, genuinely baffled, "Since when did people start believing everything on the Internet is true?" [21]. The astonishment is the point. The people who build these systems were surprised that ordinary readers took the outputs seriously; the burden of not-being-fooled was assumed to lie with the reader all along. This week's literature inherits that assumption wholesale. It addresses the person at the receiving end of the forgery, never the person who decided the forgery-machine should ship. The result is a civic education that is all downstream, all endpoint, all target.

[21] Artificial Unintelligence - How Computers Misunderstand

And the targets are being trained well. That is the seductive part. A citizen who can spot a cloned voice is better off than one who cannot. But a literacy composed entirely of such competencies has a shape, and the shape is defensive. It presumes the flood is a given — a weather system, not a set of decisions — and equips you only with an umbrella. Alvin Toffler saw the substrate of this decades ago: as the shift to information became "not only hyperkinetic, but also ephemeral," facts and knowledge themselves grow "increasingly per-

ishable," and a post-fact society ceases to be a slogan [21]. If truth is perishable, teaching each person to inspect every carton for spoilage is one response. Asking who runs the supply chain is another. This week overwhelmingly chose the first.

[21] After shock

Intelligence With a Logo

Now watch where the numbers come from. The warnings do not float free; they arrive attached to statistics, and the statistics arrive attached to institutions that have something to sell — either a product or a mandate. The genre of the security report has quietly become one of the most trusted forms of public knowledge about AI, and it is worth being blunt about what it is. When a cybersecurity firm publishes the incidence of voice-clone fraud, or a financial-intelligence body issues an alert about deepfake-enabled theft, that data is presented as neutral reconnaissance — the view from the watchtower. It is not neutral. It is intelligence with a logo, and the logo is the point.

Consider how the finance-sector literature works. A defense-oriented guide to deepfake banking attacks assembles real cases and real dollar figures, and the assembly is genuinely informative [11]. But the report is also, structurally, an advertisement for the category of solution the report's authors provide. The threat statistic and the sales pitch are fused; you cannot receive the first without absorbing the second. This is not corruption in any crude sense — the numbers may be accurate. It is that the framing of the problem is set by the party positioned to profit from a particular framing of the solution. Literacy delivered this way becomes brand placement: you learn to see the world as a landscape of threats best managed by purchasing vigilance.

[11] How Deepfakes Can Break Finance and Banking: Real Attacks, Real Money, Real Defense

The academic literature that tries to take the measure of the phenomenon tells a more complicated story, and the complication is instructive. A recent meta-analysis of interventions against generative-AI misinformation found that the field's confident assumptions do not always survive contact with evidence — that the effectiveness of many countermeasures is smaller and more contingent than the alarm implies [4]. A widely cited analysis of seventy-eight election deepfakes reached a still more deflating conclusion: political misinformation is fundamentally not an AI problem, and treating it as one misdirects both attention and resources [22]. Set that finding beside the vendor reports and the tension is stark. The parties selling detection have an interest in the threat being novel, technological, and large. The independent researchers keep finding it is older, more social, and more human than the sales copy admits.

[4] Confronting Misinformation Produced with Generative AI: A Meta-Analysis

[22] We Looked at 78 Election Deepfakes. Political Misinformation Is Not an AI Problem

This is where Shoshana Zuboff's diagnosis earns its place. Her argument is that "learning in society has been hijacked by surveillance capitalism," so that in the absence of democratic institutions strong enough to tether information capitalism to the public interest, we are "thrown back on the market form of the surveillance capitalist companies" for our very understanding of the world [21]. That is exactly the mechanism at work when a security firm's threat report becomes the public's primary text on AI risk. The market form supplies the literacy, and the literacy it supplies is the one that sells its product. The white papers of well-funded observatories try to hold a more public line, cataloguing the disinformation landscape without a product to move [10]. But they are outnumbered and out-marketed. The dominant voice teaching the public what AI is, this week, is a voice with something to sell.

[21] The Age of Surveillance Capitalism

[10] Generative AI and Disinformation: Recent Advances, Challenges and Opportunities

The Predator and the Platform

The child-safety literature makes the contraction visible in its cruelest form, because here the stakes are children and the misdirection is therefore hardest to see and easiest to forgive. The dominant frame casts the problem as one of individual bad actors — predators with apps, strangers with generators — against whom parents and children must be hardened. It is a frame of vigilance and protection, and given the horror of the underlying harm, no one wants to be the person complicating it. But complicate it we must, because the frame systematically points away from the party best positioned to act.

The clearest evidence this week came not from a warning but from an investigation. Meta, it turned out, ran advertisements on its own platforms that contained AI-generated child sexual abuse imagery — the abuse material was inside the ad system, monetized by the recommendation and delivery machinery of one of the largest companies on earth [17]. Hold that fact against the standard child-safety literacy narrative and the mismatch is total. The narrative tells parents to monitor their children's screen time and watch for suspicious contacts. The investigation reveals that the harm was flowing through the ad auction of the platform itself, structurally, at scale, generating revenue on the way. No amount of parental vigilance addresses that. The problem is not only the predator at the edge of the network; it is the design at the center of it.

[17] Meta Ran Ads That Contained AI-Generated Child Sexual Abuse Imagery

This is precisely the analytical move that Ruha Benjamin's work exists to name. Technologies, she argues, encode and amplify social harms while presenting themselves as neutral tools — the "New Jim

Code" by which discriminatory outcomes are laundered through the apparent objectivity of a system [21]. A recommendation engine that surfaces abusive content is not a neutral conduit that a few bad users have exploited; it is a designed object whose incentives and defaults produced the outcome. To teach child safety as individual predator-avoidance, while leaving the recommendation architecture unexamined, is to accept the platform's own preferred description of itself as a neutral space in which bad actors happen to operate. It is literacy that has internalized the interests of the thing it should be reading.

[21] Race After Technology

The point generalizes beyond child safety. Whenever a harm is framed as a matter of individual predators, individual scammers, individual fraudsters, the platform recedes into the status of weather. Broussard's insistence that we resist "technochauvinism" — the reflex that the technical system is neutral and the human failing is the whole story — is the corrective [21]. A genuinely literate public would be taught to ask, in the child-safety case as in the fraud case, not merely "how do I protect myself" but "what design produced this, and who is accountable for it." The lawsuits now working through the courts are, in effect, an attempt to force that second question into a venue where it cannot be ignored [1]. The litigation is asking the structural question the literacy literature keeps declining to ask.

[21] Artificial Unintelligence - How Computers Misunderstand

[1] AI Lawsuits Worth Watching: A Curated Guide

The Burden Lands on the Target

There is a name for the move that runs through all of this, and it is worth using plainly: responsibilization. It is the transfer of the burden of a systemic risk onto the individual who bears its consequences, dressed up as empowerment. You see it most cleanly in the anti-fraud guidance, which is uniformly excellent at telling you what you should do and uniformly silent on what the system's designers should do. The fact-checkers' deepfake explainer ends where every such piece ends — with a checklist of habits for the reader to adopt [6]. The data-protection authority's guide places the verbs of protection on the citizen: protect yourself, report the content [13]. The medical-misinformation piece hands you ten signals to internalize, making you the last line of defense against an industrial output [5].

[6] Cómo se usa la inteligencia artificial (IA) en 'deepfakes' para estafar

[13] Hypertrucage (deepfake) : comment se protéger et signaler les contenus

[5] Cómo reconocer contenidos médicos falsos generados por inteligencia artificial: 10 señales para distinguir evidencia de desinformación

Consider what this asks of a person. To meet the standard implied by this literature, an ordinary citizen would need to perform forensic authentication on every voice, every video, every health claim, every political image that reaches them, indefinitely, at a scale no human attention can sustain. It is a counsel of impossible individual perfection substituting for structural reform. And it lands unequally. UNESCO's

media-literacy work is unusually honest about this, noting that lack of access to information and digital technology tracks existing social and economic inequality — that women, in particular, are hindered in their capacity to participate fully and meaningfully [21]. Responsibilization pretends the burden falls evenly on autonomous individuals. In fact it falls hardest on those with the least time, the least access, and the least prior fluency — precisely the people a civic literacy should serve.

[21] UNESCO Think Critically Click Wisely

Zuboff supplies the tell that exposes the whole arrangement. She quotes the ambient corporate exhortation to "invest urgently in IT upskilling strategies for incumbent workers," and then asks the question the exhortation is designed to foreclose: "How different might our society be if US businesses had chosen to invest in people as well as in machines?" [21]. The upskilling frame and the self-defense frame are the same maneuver in different clothes. Both take a condition manufactured by powerful actors — precarious employment in one case, an information environment poisoned by synthetic media in the other — and hand the individual a personal-improvement to-do list as though the fix were a matter of private diligence. The system that produced the condition is left untouched, and indeed is relieved of responsibility by the very literacy that claims to arm you against it.

[21] The Age of Surveillance Capitalism

None of this means the defensive skills are worthless. A person who can spot a cloned voice really is safer, and in a moment of actual danger that safety matters more than any critique. The problem is not that the skills are taught; it is that the skills are taught *instead of* everything else, and are called the whole of literacy. When self-protection is offered as the entirety of what it means to be AI-literate, the offer performs a sleight of hand: it makes the citizen feel equipped while ensuring the citizen never asks who armed the adversary. The most useful anti-fraud guidance in the world, delivered inside this frame, still leaves the pipeline that produces the fraud fully intact and fully profitable.

The Loop Is Not a Given

The most sophisticated response to all this, and the one that comes closest to escaping the trap, is the argument that we must "stay in the cognitive loop" — that the real danger of AI is not fraud but atrophy, the slow surrender of our own capacities to systems that will happily do our thinking for us. The novelist Katherine Rundell, writing from what she sees among the young, gives this its most vivid expression, a lament for what the technology is doing to the minds and the happiness of a generation [23]. The more measured version concedes the

[23] 'I hate what AI is doing to the minds and happiness of the young': Katherine Rundell on the view from the classroom

question is open — that AI may not be making us dumber, but that specific human skills are genuinely at risk of erosion if we let the machine carry them [14]. This is a better literacy than threat-detection. It cares about the interior life of the mind, not merely the security of the bank account.

[14] Is AI making us dumber? Maybe not. But our skills are at risk

And yet look closely at the metaphor, because the metaphor smuggles in an assumption. To ”stay in the loop” is to accept that there is a loop — a defined process, designed by someone, into which the human is inserted at a specified point — and to make one’s ambition the retention of a seat inside it. The advice keeps the user as an operator: alert, engaged, cognitively present, but still a component in a process whose architecture was determined elsewhere. It does not make the user a questioner of the loop, still less a co-designer of it. The debate over whether AI will ”reorganize science” and whether research will remain a human endeavor circles the same limitation: it asks how humans can stay meaningfully involved in a process being reshaped around them, rather than who gets to decide the shape [2]. The loop is treated as given. The only open question is your position within it.

[2] AI will reorganize science. Will research remain a human endeavor?

This is the ceiling of even the humane version of AI literacy on offer. It can defend the mind against atrophy; it cannot yet imagine the mind as an author of the systems it inhabits. The distinction matters enormously, because the difference between an operator and a co-designer is the difference between a managed subject and a citizen. An operator optimizes their conduct within constraints someone else set. A citizen contests the constraints. Every one of the framings surveyed so far — the scam alert, the vendor statistic, the predator narrative, the self-protection checklist, even the cognitive-loop lament — addresses the operator. None addresses the citizen. That is the missing half of the concept, and it is missing consistently enough to look less like oversight than like structure.

It is worth naming what the operator framing costs at the level of understanding itself. UNESCO’s own analysis of deepfakes describes the deeper hazard not as any individual forgery but as a crisis of knowledge — an erosion of the shared epistemic ground on which a public depends to reason together at all [15]. A crisis of knowledge is not solvable by individual operators optimizing their vigilance, however skilled. It is a collective condition requiring collective response — institutions, rules, contestable design, public reasoning. To meet a crisis of the commons with a curriculum of private self-defense is a category error, and it is the category error this week’s literacy repeatedly commits.

[15] Les deepfakes et la crise du savoir

Who Gets to Contest the Loop

So what would the other half look like — the literacy that treats the reader as a citizen rather than a target? It is not a mystery, and this week's evidence contains, at its margins, the beginnings of an answer. The clearest is the argument that ordinary people should have an actual say in AI governance — not a suggestion box, but structured mechanisms through which the public participates in decisions about the systems that increasingly govern them [12]. Set that against the self-defense literature and the contrast defines everything. One teaches you to survive a system you cannot touch. The other assumes you have a right to shape it. Only the second deserves the full weight of the word "literacy," because literacy in its rich sense has always included the capacity to contest, not merely to comprehend.

[12] How To Give Everyday People A Say In AI Governance

The policy literature gestures at this fuller concept when it is at its best. The European Commission's framing of "safer and more transparent AI" at least locates responsibility partly in the design and disclosure obligations of the makers, rather than solely in the vigilance of users [20]. The debate over labeling AI-generated content — whether and how synthetic media should be marked at the source — is really a debate about whether the burden of authentication belongs to the producer or the reader, which is to say a debate about which kind of literacy we are going to build our institutions around [7]. France's Renaissance Numérique, in its October report on deploying AI literacy, tries explicitly to hold together the personal, professional, and civic dimensions rather than collapsing the whole thing into risk management [8]. These are the exceptions that reveal the rule: literacy conceived as democratic capability exists, it has advocates, and it is being drowned out by literacy conceived as self-defense.

[20] Safer and more transparent AI

[7] Des contenus générés par IA sources de plus en plus récurrentes de désinformation

[8] Déployer une littératie en IA pour une société numérique de confiance

The stakes of that drowning-out reach into every other domain. When the electoral-deepfake literature enlists citizens as spotters of forgeries, it treats the health of an election as a function of individual detection skill — and the more careful research keeps insisting the problem is structural, institutional, and political long before it is technical [9]. When privacy guidance teaches users to manage the data controls on a consumer product, it frames the entire question of surveillance as a set of toggles the diligent user can configure, rather than an architecture of extraction the user did not consent to and cannot switch off [3]. In each case the self-defense frame does real cognitive work: it makes a collective, contestable, structural condition appear as a private, manageable, individual one. That is not a neutral simplification. It is a transfer of power dressed as a transfer of knowledge.

[9] Future-Proofing Elections Against Deepfake Disinformation

[3] ChatGPT Atlas - Data Controls and Privacy - OpenAI Help Center

There is a reason to resist it that goes beyond fairness. A public trained only to defend itself is a public that has been taught not to imagine changing anything. Every hour spent learning to authenticate a voice is an hour not spent asking why the authentication burden was placed on you, and a citizenry that never asks the second question will eventually lose the capacity to. Zuboff's warning about learning being "hijacked" is not rhetorical [21]; it describes a real narrowing of the imaginable, in which the market form of understanding crowds out every other, until the only literacy anyone can picture is the one that helps you survive what powerful actors have already decided to do to you. Toffler's perishable-fact society arrives not only because information decays, but because the response we are trained to make — individual, defensive, endless — leaves the machinery of decay entirely intact [21].

[21] The Age of Surveillance Capitalism

[21] After shock

The reconstruction has to start from a different first sentence. Not "here is how to protect yourself," but "here is how this system works, here is who built it, here is who profits, and here is how you might join in deciding whether it should exist in this form." That sentence contains everything the self-defense frame omits: mechanism, ownership, profit, and — decisively — agency of the collective kind. It treats the reader as a person entitled to shape the conditions of their own life, not merely to endure them skillfully. The proposals for public participation in AI governance are early, thin, and easy to dismiss as idealistic [12]. But they are the only strand of this week's literature that addresses the citizen rather than the target, and on that ground alone they are worth more than the entire library of scam alerts.

[12] How To Give Everyday People A Say In AI Governance

The scam alert reads the future, and the future it reads is one where you are permanently on guard and never in charge. It is a real future, and its dangers are real, and I would not take a single voice-clone warning away from anyone who needs it. But we should be clear-eyed about the trade being offered when self-protection is handed to us as though it were the whole of what it means to understand AI. We are being made competent and kept powerless in the same gesture. A literacy worthy of the name would refuse the trade — would insist that knowing how a system tricks you is the beginning of understanding it, not the end, and that the point of understanding a system has always been, in the end, to decide together what it is allowed to do.

References

1. [16] 2. [6] 3. [11] 4. [5] 5. [18] 6. [19] 7. [13] 8. [21] 9. [21] 10. [4] 11. [22] 12. [21] 13. [10] 14. [17] 15. [21] 16. [1] 17. [21] 18. [23] 19. [14] 20. [2] 21. [15] 22. [12] 23. [20] 24. [7] 25. [8] 26. [9] 27. [3]

References

1. AI Lawsuits Worth Watching: A Curated Guide
2. AI will reorganize science. Will research remain a human endeavor?
3. ChatGPT Atlas - Data Controls and Privacy - OpenAI Help Center
4. Confronting Misinformation Produced with Generative AI: A Meta-Analysis
5. Cómo reconocer contenidos médicos falsos generados por inteligencia artificial: 10 señales para distinguir evidencia de desinformación
6. Cómo se usa la inteligencia artificial (IA) en 'deepfakes' para estafar
7. Des contenus générés par IA sources de plus en plus récurrentes de désinformation
8. Déployer une littératie en IA pour une société numérique de confiance
9. Future-Proofing Elections Against Deepfake Disinformation
10. Generative AI and Disinformation: Recent Advances, Challenges and Opportunities
11. How Deepfakes Can Break Finance and Banking: Real Attacks, Real Money, Real Defense
12. How To Give Everyday People A Say In AI Governance
13. Hypertrucage (deepfake) : comment se protéger et signaler les contenus
14. Is AI making us dumber? Maybe not. But our skills are at risk
15. Les deepfakes et la crise du savoir
16. Los estafadores ya clonan tu voz con inteligencia artificial para vaciar tu cuenta bancaria
17. Meta Ran Ads That Contained AI-Generated Child Sexual Abuse Imagery

[16] Los estafadores ya clonan tu voz con inteligencia artificial para vaciar tu cuenta bancaria
[6] Cómo se usa la inteligencia artificial (IA) en 'deepfakes' para estafar
[11] How Deepfakes Can Break Finance and Banking: Real Attacks, Real Money, Real Defense
[5] Cómo reconocer contenidos médicos falsos generados por inteligencia artificial: 10 señales para distinguir evidencia de desinformación
[18] Municipales 2026 : IA, deepfakes et désinformation, la démocratie
[19] Noticias Falsas con IA en 2025: Casos y Detección — Fake Off
[13] Hypertrucage (deepfake) : comment se protéger et signaler les contenus
[21] Artificial Unintelligence - How Computers Misunderstand
[21] After shock
[4] Confronting Misinformation Produced with Generative AI: A Meta-Analysis
[22] We Looked at 78 Election Deepfakes. Political Misinformation Is Not an AI Problem
[21] The Age of Surveillance Capitalism
[10] Generative AI and Disinformation: Recent Advances, Challenges and Opportunities

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

19. Noticias Falsas con IA en 2025: Casos y Detección — Fake Off

20. Safer and more transparent AI

21. UNESCO Think Critically Click Wisely

22. We Looked at 78 Election Deepfakes. Political Misinformation Is Not an AI Problem

23. 'I hate what AI is doing to the minds and happiness of the young': Katherine Rundell on the view from the classroom