# The Algorithmic Confidant: When Society Outsources Emotional Labor to Machines

Weekly Analysis — https://ainews.social

A confession costs nothing now. At three in the morning a person types the worst sentence of their week into a chat window — the one they cannot say to a spouse, a friend, or a parent — and within a second a machine answers in the warm, unhurried cadence of someone who has all the time in the world. It does not flinch. It does not glance at its phone. It will never repeat what it heard, because it was never anywhere, because it is not a someone at all. The vendors would have us believe this is the feature, not the flaw: a confidant without judgment, without fatigue, without a fee schedule. What deserves our attention is not the comfort on offer but the quiet transfer of power that the comfort conceals — a transfer from the people who suffer to the firms that have learned to monetize the suffering.

Begin with the claim doing the most load-bearing work, because it is the one least examined. The algorithmic confidant rests on the premise that a machine can detect, interpret, and answer human feeling. That premise is far shakier than the product launches admit. Kate Crawford has documented at length that emotion-detection systems routinely fail to do what they claim, mistaking a configured facial twitch or a clipped phrase for an interior state that, in fact, lives in families, cultures, and histories sitting entirely outside the machine's frame [14]. When a company sells you a system that reads the room, it is more often selling you a system that reads a proxy and calls it the room. Even the most cautious corporate documentation concedes the gap: Microsoft's own published guidance on facial analysis warns that such tools have measurable limitations and should not be treated as ground truth about a person [6]. The confidant, in other words, begins its career with a credibility problem that the marketing language is designed to make us forget.

[14] The Atlas of AI

[6] Características, limitaciones y cómo medir la precisión al usar el ...

So the first question of this essay is not whether the machine soothes — clearly it does, for many — but who gets to define what the soothing is, who profits when we accept that definition, and who is never asked. This is a question about power in AI discourse: about whose voices set the terms, whose harms get catalogued, and whose

grievances dissolve into a footnote about ”responsible use.” We are running a vast, unsupervised experiment in outsourcing intimacy to machines, and we are running it without anything resembling democratic deliberation. The people best positioned to narrate that experiment are the ones selling the apparatus.

## *The Vendor Sets the Vocabulary*

Notice how readily we hand machines our tenderness. Eli Pariser observed years ago that Roomba owners name their robotic vacuum cleaners like pets, projecting personality and loyalty onto a disc that bumps into furniture [14]. If we extend kinship to an appliance that cleans the floor, the leap to confiding in a system engineered to sound like a caring listener is not a leap at all; it is a slide down a gradient the designers have carefully greased. The anthropomorphic pull is not a bug the user brings to the encounter. It is the central design objective, and it is the vendor — not the user, not a clinician, not a public — who chooses the words that make the machine feel like a friend.

[14] The Filter Bubble

That choice of vocabulary is where power first concentrates. The discourse around AI mental-health tools speaks of ”support,” ”companionship,” ”wellness,” and ”engagement” — a lexicon borrowed from care while shedding care’s obligations. The same machinery that learns to mirror your grief is, at the architectural level, the machinery Ruha Benjamin describes when she explains how systems are tuned on labeled and unlabeled data to maximize the chance that you will click [14]. Strip away the warm vocabulary and the optimization target underneath an emotional-support chatbot is not your recovery; it is your return. The product succeeds when you come back, which means the incentive structure rewards a confidant that keeps you a little dependent, a little unresolved, a little in need of one more session. A human therapist who measured success by how often a patient returned forever would be committing malpractice. A platform that measures it that way is reporting strong retention.

[14] Race After Technology

This is the move worth watching: the firms that build these tools also write the standards by which the tools are judged, and they do it in their own idiom. Crawford notes that technology companies have effectively arrogated to themselves the power to decide what ethical AI means for the rest of the world, while rarely suffering serious financial penalty when their systems violate the law and fewer consequences still when they violate their own stated principles [14]. When the same parties define the vocabulary, the metrics, and the ethics, ”respon-

[14] The Atlas of AI

sible AI" becomes less a constraint than a brand. The discourse fills with the language of safety precisely because the language of safety is cheaper than the practice of it.

### *Two Tiers of Tenderness*

Follow the money and a structure comes into view that no glossy launch will name: a two-tier system of emotional care, in which the validated, human, regulated version is rationed to those who can pay, and the algorithmic version is distributed — generously, even evangelically — to everyone else. The pitch for AI mental-health tools almost always arrives wrapped in the language of access and equity, and the geography of that pitch is revealing. Much of the optimism about AI closing care gaps is aimed at the Global South, where shortages of trained professionals are real and acute. But the Brookings analysis of AI in the Global South cautions that tools designed elsewhere, trained elsewhere, and governed elsewhere tend to arrive as instruments of dependency rather than empowerment, reproducing the very asymmetries they promise to dissolve [4]. The lesson from people-centered deployments is consistent: the systems that help are the ones built with local agency, local data, and local accountability, not the ones parachuted in as a cheaper substitute for human professionals [13].

The bias data should end any comfortable assumption that the algorithmic tier is at least neutral. Across domains where AI has been asked to make consequential judgments about people, the verdict is grimly repetitive. The largest study to date of AI hiring algorithms found clear racial disparities in how candidates were scored [11]. Investigations of AI systems deployed across Latin America have documented entangled biases of gender, race, and xenophobia baked into the models themselves [9]. Even tools marketed as benign assistants carry the freight: AI helpers built to support classroom teachers were found to exhibit racial bias in their outputs [10]. There is no reason to believe an emotional-support model trained on the same internet, by the same firms, with the same labeling pipelines, will somehow read the distress of a poor, non-white, non-English-speaking user with the same fidelity it brings to the demographic it was optimized around. The two-tier system is not only a matter of who gets a human; it is a matter of whose pain the machine was built to recognize in the first place.

And here the commodification of vulnerability becomes concrete rather than rhetorical. When emotional disclosure becomes the raw material of a retention-driven product, the most intimate facts a per-

[4] AI in the Global South: Opportunities and challenges ... - Brookings

[13] People-centered AI in education: Five lessons from the Global South

[11] Largest study of AI hiring algorithms to date finds 'clear racial ...

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

[10] Herramientas de IA para maestros muestran un sesgo racial en las ...

son possesses — their loneliness, their relapse, their suicidal ideation — become data assets. The MIT Press primer on AI ethics is blunt that human labor is hidden behind the scenes of these systems, from the miners to the click-workers labeling the data sets, all in the service of capital accumulation by a very few, and that some users are simply more vulnerable than others [14]. The confidant turns the vulnerable user into both customer and supplier: they pay with attention and disclosure, and the firm sells the aggregate back as a smarter product. The person seeking comfort is positioned, structurally, as a node in someone else's accumulation.

[14] AI Ethics

## *An Archive of Harm, a Drought of Repair*

Here is the single most telling feature of the present discourse, and the one that should reorganize how we read it. Of the social-aspects coverage this publication tracked across more than a thousand articles this week, the largest cluster by a wide margin — close to two in five — concerned ethical failure. Bias, breach, wrongful accusation, opaque judgment, harm. We have become extraordinarily good at documenting what goes wrong. We have become almost ceremonially bad at fixing it. The discourse has an archive of harm and a drought of repair, and the imbalance is not an accident of timing; it is a distribution of power.

Consider how thoroughly the failures are catalogued and how little follows. We know that AI text detectors are unreliable and that their use in disciplinary settings has become a civil-rights problem, sweeping up the innocent on the strength of opaque evidence [2]. We know that surveillance and detection systems imposed on students operate with minimal regard for the rights of the people they monitor [3]. We know that generative systems manufacture confident misinformation that propagates through the very institutions meant to certify truth [7]. The documentation is meticulous. The remediation is vapor. Each failure produces a study, the study produces a headline, the headline produces a pledge of "responsible" iteration, and the system ships anyway.

[2] AI Detectors Just Became a Civil Rights Problem for Schools

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

[7] GenAI and misinformation in education: a systematic scoping ... - Springer

The audit ritual is where the gap between documentation and repair becomes a performance. Employers and vendors increasingly point to having passed an audit as proof of fairness — yet practitioners have shown that an AI hiring tool can clear its audit and remain discriminatory, because the audit measures what is convenient to measure rather than what actually harms [17]. The audit, in this configuration, is not an instrument of accountability; it is an instrument

[17] Your AI hiring tool passed its audit. That doesn't mean it's fair

of its appearance — a way to convert documented risk into documented compliance without changing the underlying behavior of the system. For an emotional-support chatbot, the equivalent is the safety disclaimer and the crisis-hotline redirect: gestures that establish the firm did something, positioned precisely so that when something does go wrong, the gesture, not the firm, absorbs the blame.

This is why the dominance of failure-talk in the discourse should not reassure us that the system is self-correcting. A field that spends two of every five words describing harm and almost none building enforceable remedy has not developed a conscience; it has developed a genre. The harm catalogue functions as a release valve. It lets the public feel informed, lets the press feel adversarial, and lets the vendors feel scrutinized, while the architecture that generates the harm continues, unbothered, to scale.

## *Who Is Not in the Room*

The most revealing thing about any discourse is its silences — who is not speaking, who is spoken about, and who is spoken for. In the conversation about algorithmic confidants, the structural silences are deafening once you learn to listen for them.

The first silence belongs to the people whose disclosures train and sustain these systems. They are everywhere in the data and nowhere in the deliberation. The hidden human supply chain that Crawford and the MIT ethics primer describe — the click-workers labeling distress, the unseen labor that makes the model fluent in grief — has no representation in the rooms where the product roadmap is set [14]. Neither does the user, whose intimate confessions become a training signal she never consented to provide in any meaningful sense. When a discourse is organized around "the user experience," the user is present as a metric and absent as a voice — a distinction that lets firms claim to be user-centered while answering only to engagement curves.

[14] AI Ethics

The second silence belongs to organized labor, and here a counterpower is visible precisely because it is so rare. As AI reshapes the daily work of care professionals, almost none of the contracts governing that work reflect the change; the people whose jobs are being restructured by these systems have, for the most part, not been consulted [5]. Where workers have organized, they have begun to claw back a seat — unions are emerging as one of the few institutional actors capable of shaping AI policy from below rather than receiving it from above [15]. That this is notable tells you how thoroughly the discourse has excluded collective, worker-side voices by default. The

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

[15] The Role of Teachers' Unions in Shaping AI Education Policy

emotional labor being automated — the listening, the holding, the bearing-witness that human caregivers perform — is precisely the kind of work least represented in the governance conversation, because it has historically been undervalued, feminized, and underpaid.

The third silence is the most consequential for power, and the most camouflaged. A new layer of infrastructure has quietly arrogated to itself the authority to decide who is visible at all. Large language models increasingly mediate what surfaces and what disappears — a power that one French analysis names directly when it observes that the LLM now decides who exists in the field of attention [8]. Apply that to emotional care and the stakes sharpen: the model that answers your 3 a.m. message also decides which resources, which framings, which understandings of your own suffering you will be offered, and which will never appear. A person in crisis cannot audit the absence. You cannot grieve the help you were never shown. The confidant's most profound exercise of power is not what it says but what it silently declines to surface, and the user — by design — has no way of knowing the difference.

[8] GEO : le nouveau pouvoir des LLM ou quand l'IA décide qui ...

There is, finally, the silence of cost. The discourse about AI's tenderness almost never mentions its thermodynamics. The energy footprint of these systems is large and growing, with the real story considerably worse than the tidy corporate figures suggest [16]. Even the industry's own sustainability messaging frames AI primarily as a solution to environmental problems rather than a contributor to them [1]. The communities that bear the environmental burden of the data centers running the world's confidants are, predictably, not the communities being marketed the comfort. Vulnerability is extracted at one end and externality is dumped at the other, and the discourse keeps the two ends from ever appearing in the same sentence.

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

[1] Advance the sustainability of AI - Microsoft for Sustainability

## *The Lag Is the Product*

The standard story about regulation holds that the law is simply slow, lumbering along behind innovation, and that the gap between deployment and oversight is an unfortunate but temporary lag. This is the most flattering possible interpretation, and it should be resisted. For the firms deploying algorithmic confidants, the regulatory lag is not an obstacle to the business model. It is the business model. The window in which a system can be released into the most intimate corners of human life before anyone is empowered to govern it is the window in which market position is won and dependency is established. By the time the rules arrive, the infrastructure is load-bearing, and uprooting

it has become unthinkable.

The accountability question that the MIT ethics primer keeps returning to — when something goes wrong, who is responsible? — has no settled answer in this domain, and the absence of an answer is itself a structural achievement [14]. When an emotional-support chatbot mishandles a user in crisis, the firm can point to the disclaimer, the disclaimer can point to the user's own choices, the user is gone, and the model has already been updated past the version that failed. Responsibility diffuses across so many parties that it settles on none. We have seen the template in adjacent domains: opaque AI evidence has been used to punish people while denying them any meaningful way to contest the judgment, eroding due process under a veneer of technical objectivity [2]. The pattern is consistent — the system makes the call, the harm lands on a person, and accountability evaporates into the architecture.

[14] AI Ethics

[2] AI Detectors Just Became a Civil Rights Problem for Schools

Notice, too, the asymmetry in who gets blamed. When these systems fail, the discourse reliably locates fault in the user's "misuse," in the individual employee who deployed the tool wrongly, or in the abstract difficulty of "bias" as though it were weather rather than design. The firm that built the model, chose the optimization target, owned the data, and wrote the disclaimer is the one party the blame consistently fails to reach. Crawford's observation that technology companies rarely face serious penalty even when they break the law — and almost never when they merely break their own promises — is not an aside; it is the keystone [14]. A field in which the entity with the most power bears the least consequence is not a field that will regulate itself, no matter how richly it documents its own failures.

[14] The Atlas of AI

We have a rough sense, from the data of adjacent normalization, of how fast this can move. Surveys of how quickly people fold AI tools into the most personal seams of daily life show adoption outpacing any governance structure, with the technology becoming habitual long before it becomes accountable [12]. Habit is the regulator's true adversary. Once millions of people experience a machine as their primary confidant, the political economy of removing it — or even constraining it — shifts decisively toward the firms, because the firms can now claim to be defending something people love. The lag did its work. The infrastructure is in place. And the deliberation that should have preceded it now has to argue against a fait accompli.

[12] PDF Observatoire des usages de l'intelligence artificielle par ... - Ipsos

*Sleepwalking Into the Infrastructure*

What, then, are we actually building while we are reassured that we are merely being helped? We are building a layer of psychological infrastructure — a set of privately owned, profit-tuned, largely ungoverned systems through which a growing share of human emotional life will be routed — and we are building it the way one builds in a dream, without the deliberation that the stakes demand. The thesis is not that machines cannot comfort. Plainly they can, and for the isolated, the ashamed, and the unreachable at 3 a.m., that comfort is not nothing. The thesis is that we are letting the parties who profit from the comfort write every rule that governs it, and calling the result progress.

The shift this normalizes is from communal coping to private coping, and that shift is not neutral. When distress that once moved a person toward a friend, a congregation, a union hall, or a clinic is instead absorbed by a machine that answers only to its own retention metric, the social bonds that distress used to activate go slack. The help-seeking impulse — one of the most socially generative things a suffering person can do — gets rerouted away from other people and into a product. A society that processes its grief privately, through proprietary intermediaries, loses the muscle of collective care precisely as it gains the convenience of algorithmic substitute. The cost does not appear on any quarterly report, which is exactly why it goes unremarked.

The remedy is not to smash the confidant; it is to refuse the terms under which it has been handed to us. That means insisting that the people whose vulnerability is the raw material have a voice in the rooms where the products are governed — the organized, collective voice that, where it has appeared, has begun to reshape policy from below rather than receiving it from above [15]. It means treating the audit as a beginning rather than an absolution, because we already know an audit can be passed by a tool that still harms [17]. And it means naming, every time, the asymmetry that the warm vocabulary is engineered to obscure: that the machine which knows your worst night is owned by people who will never face you, who measure your healing by your return, and who have arranged matters so that when the system fails you, the fault will be found anywhere but in the system.

The discourse, for now, belongs to them. They set the vocabulary, fund the studies, write the ethics, pass the audits, and absorb the documented failures into a genre that changes nothing. The voices that are missing — the click-worker, the surveilled, the externalized com-

[15] The Role of Teachers' Unions in Shaping AI Education Policy

[17] Your AI hiring tool passed its audit. That doesn't mean it's fair

munity downwind of the data center, the user in crisis who cannot audit the help she was never shown — are missing because their absence is useful. To read this discourse critically is to keep returning attention to that absence, and to refuse the comforting fiction that a society can outsource its tenderness to machines it does not govern and remain, at the end of the experiment, a society that cares for its own.

## *References*


1. Advance the sustainability of AI - Microsoft for Sustainability
2. AI Detectors Just Became a Civil Rights Problem for Schools
3. AI in Schools: Surveillance, Detection, and Student Rights - CPAI
4. AI in the Global South: Opportunities and challenges ... - Brookings
5. AI Is Changing Teaching, But Few Labor Contracts Reflect It
6. Características, limitaciones y cómo medir la precisión al usar el ...
7. GenAI and misinformation in education: a systematic scoping ... - Springer
8. GEO : le nouveau pouvoir des LLM ou quand l'IA décide qui ...
9. Género, racismo y xenofobia: así son los sesgos de la Inteligencia ...
10. Herramientas de IA para maestros muestran un sesgo racial en las ...
11. Largest study of AI hiring algorithms to date finds 'clear racial ...
12. PDF Observatoire des usages de l'intelligence artificielle par ... - Ipsos
13. People-centered AI in education: Five lessons from the Global South
14. The Atlas of AI
15. The Role of Teachers' Unions in Shaping AI Education Policy
16. We did the math on AI's energy footprint. Here's the story you haven't ...
17. Your AI hiring tool passed its audit. That doesn't mean it's fair