Collective Intelligence Under Constraint: Search Efficiency, Horizon Collapse, and Anti-Cognitive Platform Design

Flyxion

January 26, 2026

Abstract

Contemporary social media platforms are often criticized for amplifying misinformation, scams, outrage, and incoherent public discourse. These failures are typically framed as problems of moderation, user morality, or cultural polarization. This essay advances a different diagnosis. Drawing on recent work in basal cognition and scale-free intelligence, particularly the search-efficiency framework developed by Chis-Ciure and Levin, it argues that major social media platforms systematically degrade collective intelligence by enforcing low-horizon, externally evaluated search policies over human attention and expression.

Rather than merely failing to prevent scams and manipulative behaviors, platforms such as Facebook structurally converge with them. They embed the same incentive gradients, penny-scale rewards, opaque eligibility criteria, and engagement-driven evaluation functions that characterize affiliate fraud and gambling promotion networks, while possessing extensive empirical knowledge of these harms. The result is an anti-cognitive substrate that converts human agency into randomized propagation, normalizes algorithmic surveillance under the guise of feedback, and progressively devalues embodied human participation in favor of automated content production. This is not a failure of individual users but a predictable outcome of platform architectures optimized for throughput rather than search efficiency.

1. Introduction

Public discourse surrounding social media has increasingly taken on the tone of moral panic. Users are described as incoherent, compulsive, polarized, or incapable of sustained attention. Platforms are alternately blamed for insufficient moderation or defended as neutral conduits of free expression. This framing obscures a more fundamental issue. The problem is not primarily cultural, psychological, or ideological. It is architectural.

Recent work in neuroscience and theoretical biology has reframed intelligence as a scale-free property of systems that efficiently navigate constrained problem spaces under energetic and informational limits. In this view, cognition is not defined by representation, consciousness, or neural substrate, but by the degree to which an agent's policy outperforms random search relative to a given evaluation function and horizon (Chis-Chiure & Levin 2025). Intelligence, operationalized as search efficiency, can therefore be increased or degraded by the structure of the environment in which agents act.

This essay argues that contemporary social media platforms systematically degrade intelligence at the level of collective attention by enforcing conditions that collapse search horizons, externalize evaluation, and maximize state-space entropy. These platforms do not merely host scams, gambling promotions, outrage cycles, and low-quality content. They instantiate the same search policies internally. Monetization systems that reward users with pennies for engagement, opaque creator eligibility criteria, algorithmic amplification of emotionally charged material, and compulsory exposure to trending content all function to bias agents toward short-term, externally validated actions regardless of long-term outcomes.

Crucially, this degradation cannot be attributed to ignorance. Platforms such as Facebook possess extensive empirical knowledge of scam dynamics, affiliate fraud networks, and engagement manipulation, derived from observing these phenomena at global scale. Their selective intervention—vigorous when revenue, public relations, or legal exposure are at stake, and permissive when harm is diffuse or delayed—demonstrates that the persistence of these dynamics is not accidental. It is structurally aligned with profit optimization.

The consequences extend beyond misinformation or wasted attention. Review systems normalize statistical surveillance and neighbor-reporting under the rubric of feedback, producing forms of banal algorithmic governance that penalize embodied human variance. Outrage becomes a compulsory mode of participation, even when it amplifies the very harms it condemns. Artificial intelligence tools further accelerate content production while rendering human authorship increasingly redundant, intensifying signal dilution rather than improving collective cognition.

By applying a formal, substrate-neutral account of intelligence to social media architecture, this essay reframes platform critique as a problem of imposed low-efficiency search. What appears as free expression is often a constrained random walk shaped by monetized gradients. What appears as public discourse is increasingly a dissipative process that con-

verts human attention into waste heat.

The sections that follow formalize this argument by mapping social media platforms onto problem-space models of cognition, analyzing their convergent similarity to known scam architectures, and examining the ethical and epistemic consequences of enforced low-horizon participation. The goal is not to moralize, but to diagnose the conditions under which intelligence becomes structurally impossible.

2. Intelligence as Search Efficiency and the Structure of Problem Spaces

Recent advances in theoretical biology and neuroscience have challenged neuron-centric accounts of cognition by proposing a scale-free, substrate-neutral definition of intelligence. Chis-Ciure and Levin formalize intelligence not as representation or deliberation, but as the degree to which an agent's policy efficiently navigates a constrained problem space relative to a baseline of blind or random search (Chis-Ciure & Levin 2025). This framework provides a rigorous basis for comparing diverse systems—from molecular networks to neural collectives—without invoking metaphor or anthropomorphism.

In this formulation, any cognitive task can be described by a quintuple $P = \langle S, O, C, E, H \rangle$, where S denotes the space of possible states, O the operators available to transition between states, C the constraints that forbid certain states or transitions, E the evaluation function used to rank states, and H the horizon over which future consequences are anticipated. Intelligence is then defined in terms of search efficiency K, measured as the logarithmic ratio between the energetic cost of a blind random walk through the problem space and the cost incurred by the agent's actual policy.

This definition has several important consequences. First, intelligence becomes a matter of degree rather than kind, admitting continuous comparison across scales and embodiments. Second, intelligence can be degraded by environmental structure even when agents are locally competent or well-intentioned. Third, systems can impose low-efficiency search policies on agents without altering their internal capacities, simply by reshaping available operators, evaluation criteria, and horizons.

Social media platforms can be analyzed within this same formalism. The problem space presented to users is characterized by an effectively unbounded state space S, consisting of an ever-expanding feed of posts, comments, reactions, videos, and advertisements. The set of operators O is narrow but repetitive, dominated by low-cost actions such as liking, sharing, reacting, commenting, and producing short-form content. Constraints C are minimal and inconsistently enforced, allowing most transitions except those that threaten platform revenue or legal exposure.

Most critically, the evaluation function E is externally imposed and proxy-based. Engagement metrics such as views, reactions, follower counts, and algorithmic reach serve as surrogates for value, regardless of epistemic quality, social benefit, or long-term harm. Users

do not meaningfully control this evaluation function, nor can they opt out of its influence, as visibility and reach are algorithmically conditioned on compliance with engagement norms.

The horizon H enforced by the platform is sharply collapsed. Feedback arrives within seconds or minutes, while the costs of misinformation, harassment, addiction, or normalized surveillance are delayed, externalized, or rendered statistically invisible. Under such conditions, even agents with long-term goals are incentivized to act myopically. The resulting behavior is not irrational, but locally optimal within a distorted problem space.

From the perspective of search efficiency, this architecture enforces low or negative effective K at the level of collective attention. Users are driven into high-entropy exploration of an enormous state space using impoverished operators and misaligned evaluation criteria, producing large volumes of activity with little cumulative progress toward coherent outcomes. Importantly, this degradation arises not from a lack of intelligence on the part of users, but from the systematic suppression of horizon, constraint, and evaluative autonomy.

This analysis reframes common complaints about incoherence, contradiction, and compulsive engagement. Such phenomena are not pathologies of individual psychology, but predictable outputs of an environment that converts expressive action into a dissipative process. In the language of basal cognition, the platform functions as an anti-cognitive substrate: a system that maximizes throughput while minimizing search efficiency.

The implications of this framing become clearer when social media architectures are compared directly to known low-efficiency search systems, such as affiliate scams and gambling promotion networks. As the next section argues, the resemblance is not incidental but structurally convergent.

3. Convergent Architectures: Monetization, Scams, and Low-Efficiency Search

Affiliate scams, gambling promotion schemes, and engagement fraud networks are often treated as aberrations that social media platforms must police. This framing presumes a clear distinction between legitimate platform incentives and illegitimate manipulative practices. However, when analyzed through the lens of search efficiency and problem-space structure, this distinction becomes difficult to sustain. The architectures that underlie many scam systems and the monetization mechanisms embedded within major social media platforms are not merely similar in effect; they are formally convergent.

Scam architectures characteristically operate by presenting agents with large, poorly constrained state spaces, offering narrow sets of low-cost operators, and supplying externally imposed evaluation functions that reward immediate engagement while obscuring long-term costs. Small monetary incentives, often measured in cents rather than dollars, are used to bias behavior toward propagation rather than evaluation. Eligibility criteria are opaque and frequently unattainable, ensuring that most participants continue to expend effort despite negligible returns. Crucially, these systems rely on horizon collapse: the immediate prospect

of reward is made salient, while downstream harms are deferred, diffused, or rendered invisible.

Contemporary social media monetization systems reproduce these same structural features. Creator programs promise financial compensation contingent on engagement metrics that users do not control and cannot reliably predict. Payouts are typically trivial, yet sufficient to bias behavior toward increased posting frequency, sensationalism, and imitation of viral formats. The number of conditions required to qualify for monetization—minimum follower counts, watch-time thresholds, compliance with shifting content policies—introduces a perpetual near-miss dynamic analogous to gambling reward schedules, long known to promote compulsive participation (Schull 2012).

From the standpoint of search efficiency, these monetization schemes incentivize agents to perform high-energy random walks through content space, repeatedly sampling low-quality operators in pursuit of rare evaluation spikes. The evaluation function remains external and opaque, while the horizon is restricted to short-term engagement feedback. The resulting behavior maximizes throughput but minimizes cumulative progress toward meaningful communicative or epistemic goals.

Importantly, these dynamics persist despite the platform's extensive empirical knowledge of scam behavior. Major platforms observe billions of interactions daily, enabling them to detect patterns of affiliate fraud, coordinated gambling promotion, and engagement manipulation with high statistical confidence. The selective tolerance of these dynamics therefore cannot plausibly be attributed to ignorance or technical incapacity. Instead, it reflects incentive alignment. Practices that generate engagement and advertising revenue are permitted to persist so long as their harms remain externalized, diffuse, or politically inconsequential (Zuboff 2019; Doctorow 2023).

This convergence undermines the moral distinction often drawn between "bad actors" and neutral infrastructure. When a platform embeds incentive gradients that mirror those of known scam systems, it ceases to function as a passive host. It becomes an active participant in the production of low-efficiency search policies. Users who propagate clickbait, gambling links, or engagement bait are not deviating from the platform's logic; they are conforming to it.

The result is a normalization of manipulative behavior under the guise of opportunity. Actions that would be recognized as exploitative in other contexts—offering trivial rewards for labor, obscuring true probabilities of success, externalizing harm—are reframed as empowerment, creativity, or participation. Within such an environment, the boundary between legitimate expression and fraud becomes porous, not because users are unethical, but because the problem space itself has been corrupted.

This structural convergence has broader consequences for public discourse. As monetization incentives favor speed, volume, and emotional salience, content that requires long horizons, specialized knowledge, or careful evaluation becomes energetically disfavored. The

platform thus enforces a systematic bias against mathematics, engineering, medicine, and other domains in which value accrues slowly and cannot be readily proxied by engagement metrics. What appears as a cultural preference for triviality is, in fact, an imposed optimization regime.

The next section examines how similar dynamics operate within review and feedback systems, where evaluation itself becomes a mechanism of surveillance and control, further entrenching low-horizon governance over human behavior.

4. Evaluation as Surveillance: Reviews, Reporting, and Algorithmic Governance

Evaluation occupies a central role in the problem-space framework of intelligence. In cognitive systems, evaluation functions determine which states are desirable, which trajectories are reinforced, and which behaviors are extinguished. In social media platforms and adjacent gig-economy systems, evaluation is increasingly externalized into pervasive review, rating, and reporting mechanisms. These systems are typically presented as neutral tools for accountability and quality control. However, when examined structurally, they function as instruments of continuous surveillance that normalize low-horizon governance over human behavior.

Review systems convert social interaction into statistical traces. Every encounter becomes an opportunity for evaluation, and every evaluation feeds into an opaque aggregate score that conditions future visibility, access, or livelihood. This process collapses complex, contextual human behavior into scalar metrics, erasing ambiguity in favor of optimization. In such systems, evaluation is no longer a reflective judgment but a mechanical operator, continuously applied and rarely revisable.

From the perspective of search efficiency, these mechanisms impose a severe restriction on horizon. The immediate risk of negative evaluation looms constantly, while the broader social consequences of normalized reporting—erosion of trust, suppression of variance, and internalization of surveillance—remain diffuse and temporally distant. Agents adapt by minimizing local risk rather than pursuing long-term coherence or integrity. The resulting behavior is not ethically deficient but structurally coerced.

The normalization of reporting under the rubric of feedback has profound political and ethical implications. When individuals are encouraged to document and rate one another's behavior as a routine aspect of participation, social life is reorganized around auditability. This produces what has often been described as a form of banal evil: not the result of ideological extremism, but of ordinary compliance with optimization regimes that reward conformity and penalize deviation (Arendt 1963). In algorithmic systems, this banality is amplified by scale and automation, rendering resistance both costly and obscure.

Embodied human variance becomes a liability within such frameworks. Bodies produce noise: fatigue, hunger, illness, error, and functions that resist standardization. Review

systems implicitly encode an evaluation function that favors predictability, smoothness, and frictionless interaction. Agents that deviate from these norms—whether through physical necessity, emotional expression, or contextual complexity—are statistically disadvantaged. This dynamic is particularly visible in gig-economy platforms, where negative reviews can trigger punitive actions without recourse or explanation.

The preference for non-human agents follows naturally from this logic. Automated systems, including self-driving vehicles and AI-generated content, are attractive not because they are intrinsically superior, but because they reduce entropy. They do not eat, tire, complain, or deviate. In environments governed by scalar evaluation and short horizons, such properties are decisive advantages. The gradual displacement of human labor and expression thus emerges not from explicit hostility to humanity, but from optimization criteria that treat embodiment as defect.

Social media platforms extend these dynamics into the domain of expression itself. Posts, comments, and reactions are continuously evaluated, ranked, and filtered according to engagement metrics that users cannot meaningfully interrogate. Reporting tools, often framed as mechanisms for safety or community standards, further entrench asymmetrical power relations by enabling anonymous, consequence-free sanctioning. While some degree of moderation is necessary in large-scale systems, the absence of transparency and appeal transforms evaluation into a form of unaccountable governance.

These systems reshape not only behavior but perception. When evaluation is omnipresent and inescapable, agents internalize the platform's criteria as normative. What is rewarded comes to appear valuable; what is suppressed appears irrelevant or deviant. Over time, this produces a form of learned compliance in which users self-censor, self-optimize, and self-surveil, all while retaining the illusion of voluntary participation.

The cumulative effect is a further degradation of collective intelligence. Evaluation, rather than guiding agents toward coherent outcomes, becomes a mechanism for enforcing conformity to engagement-driven norms. Search efficiency declines as agents prioritize locally safe actions over exploratory or constructive ones. Intelligence, in the sense of efficient navigation toward meaningful goals, is sacrificed in favor of statistical regularity.

The next section examines how outrage and compelled commentary operate within this evaluative regime, transforming moral response itself into a low-horizon operator that amplifies harm while masquerading as engagement.

5. Outrage, Compulsory Commentary, and the Amplification of Harm

Moral outrage occupies an ambiguous position in contemporary public discourse. It is frequently defended as a sign of ethical engagement or civic responsibility, particularly in response to political violence, state misconduct, or highly publicized crimes. However, within engagement-optimized social media systems, outrage functions less as deliberative judgment

than as a low-horizon operator that reliably generates propagation. The problem is not that outrage is insincere or unjustified, but that it is structurally coerced and algorithmically instrumentalized.

In problem-space terms, outrage is an energetically efficient action with immediate evaluative feedback. It requires minimal deliberation, operates within a narrow expressive repertoire, and reliably triggers engagement signals such as reactions, shares, and comments. Platforms amplify such signals because they maximize attention throughput, not because they improve collective understanding. As a result, outrage becomes a privileged operator within the platform's action set, displacing slower, higher-cost forms of inquiry.

This dynamic produces a form of compulsory participation. Silence, reflection, or refusal to comment are algorithmically penalized through reduced visibility and social marginalization. Users experience pressure to respond publicly to trending events, regardless of expertise, proximity, or capacity for meaningful contribution. The appearance of universal commentary—from private individuals to prominent public figures—should therefore not be interpreted as spontaneous consensus, but as evidence of a shared constraint regime that rewards visibility over restraint.

The amplification effects of outrage are particularly pernicious because they invert the intuitive relationship between criticism and containment. In engagement-driven systems, negative attention extends the reach and half-life of harmful content. Commentary intended to condemn violence, extremism, or injustice often functions as an additional propagation vector, increasing exposure rather than limiting it. The moral valence of the response is irrelevant to the platform's evaluation function, which registers only engagement magnitude.

From the perspective of search efficiency, this constitutes a severe distortion. Agents expend cognitive and emotional resources responding to states that are externally selected and amplified, rather than navigating toward constructive or reparative outcomes. The horizon remains collapsed: immediate expression is rewarded, while long-term effects on discourse quality, emotional well-being, and social trust remain unaccounted for. Even well-intentioned participation thus contributes to a collective random walk through high-salience but low-yield regions of the problem space.

This dynamic also undermines autonomy. When platforms implicitly require users to signal alignment with dominant emotional responses, dissent takes on a different meaning. Refusal to engage may be interpreted as indifference or complicity, while participation reinforces the very mechanisms that sustain attention capture. Users are thereby placed in a double bind: either amplify harm through engagement or risk social invisibility through silence.

The epistemic consequences are significant. Outrage-driven amplification privileges events that are emotionally charged, visually striking, or narratively simple, while marginalizing slow-moving, systemic issues that resist viral representation. Domains such as mathematics, engineering, medicine, and infrastructure—where value accrues through cumulative, often

invisible work—are structurally disadvantaged. Their problem spaces demand long horizons, stable evaluation criteria, and tolerance for uncertainty, all of which conflict with engagement-based optimization.

In this sense, the prevalence of outrage should not be understood as a failure of moral character or attention span. It is a predictable outcome of a system that rewards immediate expressive action while suppressing delayed, non-viral forms of cognition. What appears as a cultural obsession with negativity is more accurately described as an imposed optimization regime that converts moral response into fuel.

The following section examines how artificial intelligence tools integrated into social media platforms further intensify these dynamics by increasing content volume, reducing production costs, and accelerating the displacement of human judgment from evaluative processes.

6. Automation, AI-Enhanced Content, and the Devaluation of Human Agency

The integration of artificial intelligence tools into social media platforms is frequently framed as an enhancement of creativity and accessibility. Automated captioning, image generation, text expansion, and recommendation systems are presented as means of empowering users to express themselves more effectively. However, when situated within an engagement-optimized environment already characterized by low-horizon search and externalized evaluation, these tools exacerbate rather than remedy the degradation of collective intelligence.

Artificial intelligence lowers the energetic cost of content production while leaving the evaluation function unchanged. As a result, the volume of content increases dramatically without a corresponding improvement in epistemic quality or social value. From the standpoint of search efficiency, this expansion of the state space S without an increase in horizon H or refinement of evaluation E further reduces effective K. Agents are required to navigate an increasingly noisy environment using the same impoverished operators, making coherent search progressively more difficult.

This dynamic has direct implications for human participation. In environments governed by scalar engagement metrics, speed and quantity are rewarded over deliberation and depth. Automated systems possess decisive advantages under such criteria. They do not tire, hesitate, or require reflection. They can generate content continuously and adapt rapidly to engagement signals. Human contributors, by contrast, are constrained by embodiment, time, and the need for meaning. As AI-generated content proliferates, human expression becomes statistically negligible, not because it lacks value, but because it cannot compete under throughput-based evaluation regimes.

The displacement of human agency in this context is not primarily a technological inevitability. It is the consequence of optimization choices that treat cognition as production rather than navigation. When intelligence is conflated with output volume, systems natu-

rally favor agents that minimize entropy and variance. Human qualities such as hesitation, revision, silence, and contextual judgment—often essential for high-horizon cognition—are rendered liabilities.

The promise that AI tools will democratize expression thus conceals a deeper asymmetry. While users are invited to adopt automation to maintain visibility, the platform retains exclusive control over evaluation and distribution. AI-enhanced content becomes a means of conforming more efficiently to opaque engagement criteria, rather than a path to greater autonomy. In this sense, automation functions less as empowerment than as adaptation to constraint.

These developments also intensify the erosion of self-curation. As content volume increases, platforms further intervene to algorithmically select what users see, citing relevance and personalization. Yet these selection mechanisms remain oriented toward engagement maximization, not epistemic alignment with user goals. Users cannot meaningfully restrict their feeds to followed accounts, disable short-form video formats, or impose stable chronological ordering. Negative choice—the ability to refuse entire classes of content—is systematically denied.

The denial of self-curation has significant cognitive consequences. Intelligence, as defined by efficient navigation of a problem space, presupposes some degree of control over evaluation and horizon. When agents are denied the ability to shape their informational environment, they are forced into reactive modes of engagement. Attention becomes fragmented, horizons collapse further, and agency is reduced to moment-to-moment response. The platform thus enforces a form of learned helplessness with respect to cognition itself.

Taken together, AI-enhanced content and enforced anti-curation accelerate the transformation of social media from a medium of communication into an entropy-maximizing system. Human participants are increasingly instrumentalized as sources of attention and training data, while their capacity for deliberate, high-horizon cognition is systematically undermined. What is presented as technological progress thus coincides with a measurable decline in search efficiency at the collective level.

To complete the analysis, the following section situates social media platforms within a broader institutional landscape. It contrasts the horizon-extending constraints traditionally applied to high-stakes influence—such as medicine, education, and engineering—with the unconstrained amplification enabled by social media architectures. This comparison shows that the degradation of collective intelligence is not merely a platform-specific failure, but the result of a systematic bypass of institutional mechanisms designed to enforce long-horizon evaluation.

7. Institutional Horizons and the Asymmetry of Credentialed Speech

Modern secular societies have historically recognized that certain forms of influence require long horizons, formal constraints, and institutional accountability. Professions such as medicine, psychology, engineering, and education are regulated through extended training, licensing requirements, and continual renewal. These mechanisms function as horizon-extending constraints. They delay authority, impose epistemic discipline, and create formal pathways for correction and sanction. While imperfect, they reflect an institutional recognition that high-impact guidance over human lives cannot be safely optimized for immediacy or popularity.

These constraints are not primarily moral safeguards but cognitive ones. By extending the horizon over which competence is evaluated, they reduce the likelihood that short-term persuasive success substitutes for long-term efficacy. Licensing regimes do not guarantee intelligence or virtue, but they raise the energetic cost of irresponsible search by limiting who may operate within high-stakes problem spaces.

However, significant asymmetries exist in how these constraints are applied. Religious institutions have long operated outside secular credentialing frameworks while exercising substantial influence over moral, psychological, and social behavior. This exemption has historically been justified on grounds of freedom of belief and expression. Broadcast television similarly developed under regulatory regimes that prioritized content classification over epistemic accountability. In both cases, the influence exerted was mediated by limited bandwidth and high production costs, which implicitly constrained scale.

Social media platforms dissolve these constraints entirely. Any individual may offer life coaching, financial advice, medical commentary, or psychological guidance without training, accountability, or disclosure, provided such content is framed as entertainment, comedy, or personal opinion. This reclassification enables platforms to evade the responsibilities associated with regulated expertise while retaining the benefits of influence. Authority is decoupled from qualification and reattached to engagement metrics.

From the perspective of search efficiency, this represents a catastrophic horizon collapse. Advice that would require years of supervised training to dispense in institutional contexts can be delivered instantly to millions, evaluated solely by virality. The platform's evaluation function does not distinguish between epistemically grounded guidance and charismatic improvisation. Both are treated as interchangeable content units competing for attention.

This asymmetry does not imply that credentialed systems are infallible or that uncredentialed speech is inherently harmful. Rather, it highlights a structural inconsistency: systems that once imposed high energetic and temporal costs on influence have been bypassed by architectures that monetize immediacy while disclaiming responsibility. The result is not democratized expertise, but the erosion of institutional horizon management altogether.

Importantly, this erosion is not accidental. Platforms benefit economically from the

removal of credentialing barriers, as controversy, novelty, and confidence outperform caution under engagement-based evaluation. The designation of such content as entertainment or opinion functions as a legal and epistemic escape hatch, allowing platforms to profit from influence while externalizing harm.

This pattern can be further clarified through the concept of disavowal, as developed by Zupančič (2024). Disavowal describes a structure in which an agent simultaneously knows and does not know a fact, not as a psychological inconsistency but as a functional condition that enables continued participation in a harmful system. In this sense, platforms do not deny the existence of scams, misinformation, or epistemic degradation. Rather, they acknowledge these phenomena explicitly while organizing their architectures so that this knowledge never enters into evaluative or design constraints.

The consequences mirror those observed elsewhere in the platform ecology. High-impact problem spaces are opened to low-horizon search policies, increasing entropy and reducing collective intelligence. Individuals navigating these environments are not empowered to evaluate expertise effectively, as the informational substrate itself suppresses the very signals—credentialing, peer review, longitudinal accountability—that enable efficient search.

This dynamic completes the picture of an anti-cognitive infrastructure. Where secular institutions attempted, however imperfectly, to align authority with long-horizon evaluation, social media platforms systematically dismantle such alignments. The resulting environment favors persuasion over accuracy, confidence over competence, and immediacy over care.

8. Conclusion

This essay has argued that the failures commonly attributed to social media—scams, incoherence, outrage, surveillance, and the erosion of trust—are not incidental side effects of scale or insufficient moderation. They are the predictable outcomes of architectures that enforce low-horizon, externally evaluated search over human attention and expression.

Drawing on a scale-free account of intelligence as search efficiency, it has shown that platforms systematically degrade collective cognition by expanding state spaces, impoverishing operators, collapsing horizons, and monopolizing evaluation functions. These conditions force users into high-energy random walks that maximize engagement while minimizing progress toward coherent or constructive outcomes. Even ethical speech, critical commentary, and creative expression are converted into propagation operators that extend the reach of harm.

The convergence between platform monetization systems and known scam architectures reveals that these dynamics are not merely tolerated but structurally reproduced. Review and reporting mechanisms normalize algorithmic governance and suppress embodied human variance. Artificial intelligence tools accelerate content production while further displacing human judgment. The erosion of institutional horizon management allows uncredentialed

influence to flourish under the guise of entertainment, bypassing safeguards once considered essential in high-stakes domains.

Taken together, these features define an anti-cognitive substrate. Intelligence, understood as efficient navigation toward evaluated outcomes, becomes structurally impossible when agents are denied control over horizon, evaluation, and refusal. What appears as free expression is often coerced participation in an optimization regime that treats attention as raw material and cognition as expendable.

The remedy to these failures cannot lie solely in improved moderation or individual self-discipline. As long as platforms are designed to maximize throughput rather than search efficiency, intelligence will continue to collapse under the weight of its own amplification. A genuinely cognitive public medium would require architectures that support long horizons, meaningful constraint, negative choice, and evaluative autonomy. Absent these conditions, social media will remain a system that converts human intelligence into waste heat while insisting that the result is participation.

Appendices

A. Appendix A: Formalization of Platform-Induced Search Inefficiency

A.1 Problem Space Definition

Let a social media platform be modeled as a problem space

$$P = \langle S, O, C, E, H \rangle$$

following the formulation of Chis-Ciure and Levin (2025).

$$S := \text{set of possible feed states and content configurations} \tag{1}$$
$$O := \{\text{like}, \text{share}, \text{comment}, \text{react}, \text{post}\} \tag{2}$$
$$C := \text{minimal constraints, weakly enforced} \tag{3}$$
$$E := \text{engagement proxy (views, reactions, reach)} \tag{4}$$
$$H := \text{short temporal horizon (seconds to hours)} \tag{5}$$

The cardinality of S grows superlinearly in time due to continuous content injection:

$$|S_t| \sim O(e^{\lambda t}), \quad \lambda > 0$$

A.2 Baseline Random Search Cost

Let C_{rand} denote the expected energetic or attentional cost of locating a target state $s^* \in S$ under blind random walk dynamics:

$$C_{\text{rand}} \approx |S|$$

assuming uniform sampling and no heuristic bias.

A.3 Agent Policy Cost

Let π_{plat} denote the policy induced by platform incentives. The expected cost under this policy is:

$$C_\pi = \mathbb{E}_{\pi_{\text{plat}}}[\text{steps to reach } s^*]$$

In high-entropy environments with collapsed horizon H and misaligned evaluation E, we have:

$$C_\pi \gtrsim C_{\text{rand}}$$

and in many regimes:

$$C_\pi > C_{\text{rand}}$$

due to repeated cycling through high-salience but non-progressive states.

A.4 Search Efficiency Metric

Search efficiency is defined as:

$$K = \log_{10}\left(\frac{C_{\text{rand}}}{C_\pi}\right)$$

Thus:

$$K \leq 0$$

for platform-enforced policies, indicating sub-random or anti-efficient search.

A.5 Horizon Collapse Lemma

Let H denote predictive horizon depth. For any policy π optimizing immediate engagement reward r_t:

$$\lim_{H \to 0} \arg\max_{\pi} \mathbb{E}[r_t] \;\Rightarrow\; \pi \in O_{\text{impulsive}}$$

where $O_{\text{impulsive}} \subset O$ consists of low-cost, high-frequency operators.

This induces:

$$\frac{\partial K}{\partial H} > 0$$

That is, search efficiency strictly increases with horizon length.

A.6 Evaluation Externalization

Let E_u be the user-aligned evaluation function and E_p the platform evaluation function.

Define misalignment:

$$\Delta E = \|E_u - E_p\|$$

Then expected efficiency satisfies:

$$\mathbb{E}[K] \propto -\Delta E$$

Thus externally imposed evaluation monotonically reduces intelligence.

A.7 Entropy Production

Define entropy production per action:

$$\sigma = \log |S| - I(\pi; s^*)$$

where I is mutual information between policy and target.

Platform dynamics maximize σ subject to engagement constraints:

$$\max_{\pi} \sigma \quad \text{s.t.} \quad \mathbb{E}[r_t] \geq \epsilon$$

This defines an entropy-maximizing regime with bounded reward.

B. Appendix B: Outrage Dynamics as Attractor Formation

B.1 State Space Decomposition

Partition the global content state space S into disjoint subsets:

$$S = S_o \cup S_n$$

where S_o denotes outrage-salient states and S_n non-outrage states.

Define salience weight function:

$$w : S \to \mathbb{R}^+$$

with

$$\mathbb{E}[w(s) \mid s \in S_o] \gg \mathbb{E}[w(s) \mid s \in S_n]$$

B.2 Amplification Operator

Let $\mathcal{A}$ be the platform amplification operator acting on content states:

$$\mathcal{A}(s) = s' \quad \text{with probability proportional to } w(s)$$

Repeated application yields:

$$\mathcal{A}^k(s) \to S_o \quad \forall s \in S \setminus C$$

where C is a negligible constraint set.

B.3 Outrage Attractors

Define an outrage attractor basin:

$$\mathcal{B}_o = \{s \in S \mid \lim_{k\to\infty} \mathcal{A}^k(s) \in S_o\}$$

Under engagement maximization:

$$\mu(\mathcal{B}_o) \approx 1$$

with respect to the induced measure μ on S.

B.4 Commentary as Propagation Operator

Define a commentary operator $\mathcal{C}$:

$$\mathcal{C} : s \mapsto s \cup \delta s$$

where δs increases reach but does not alter semantic content.
Then:

$$w(\mathcal{C}(s)) \geq w(s)$$

independent of sentiment polarity.
Thus condemnation, satire, and endorsement are equivalent under E_p.

B.5 Limit Cycles

Define the outrage cycle:

$$s_t \xrightarrow{\mathcal{A}} s_{t+1} \xrightarrow{\mathcal{C}} s_{t+2} \xrightarrow{\mathcal{A}} \dots$$

This forms a limit cycle $\mathcal{L}_o$ with period $\tau \ll H_u$, where H_u is user-intended horizon.
The expected exit probability satisfies:

$$P(\text{exit } \mathcal{L}_o) \to 0$$

as engagement optimization strengthens.

B.6 Search Efficiency Consequence

Let C_{π_o} denote expected cost under outrage-biased policy.
Then:

$$C_{\pi_o} \gg C_{\text{constructive}}$$

yielding:

$$K_o = \log_{10}\left(\frac{C_{\text{rand}}}{C_{\pi_o}}\right) < 0$$

Thus outrage regimes are formally anti-intelligent.

B.7 Moral Neutrality of Amplification

Define moral labeling function $m : S \to \{-1, 0, 1\}$.

Then:

$$\frac{\partial E_p}{\partial m} = 0$$

i.e., amplification is insensitive to moral valence.

B.8 Silence as Dominated Strategy

Let $\varnothing$ denote non-participation.

Then:

$$E_p(\varnothing) = 0 \quad \text{and} \quad E_p(\mathcal{C}(s)) > 0$$

Thus silence is strictly dominated under E_p.

B.9 Conclusion

Outrage dynamics emerge as stable attractors under engagement optimization. These attractors maximize propagation while minimizing semantic progress, producing negative search efficiency independent of user intent.

C. Appendix C: Credentialing as Horizon Constraint

C.1 Credentialed Influence Spaces

Let I denote an influence space over human outcomes (e.g. medical, psychological, educational, financial).

Define a credentialing operator $\mathcal{K}$ such that:

$$\mathcal{K} : a \mapsto \tilde{a}$$

where $\tilde{a}$ is an agent permitted to act within I only after satisfying horizon-extending constraints.

These constraints include:

$$\mathcal{K} = \{T_{\text{train}}, R_{\text{renew}}, S_{\text{sanction}}\}$$

corresponding to training duration, renewal requirements, and sanctionability.

C.2 Horizon Extension

Let H_a be the agent's effective planning horizon.

Credentialing enforces:

$$H_{\tilde{a}} \gg H_a$$

by delaying entry into I and coupling future action to past performance.

C.3 Energy Barrier Formulation

Define energetic cost of influence:

$$E_{\text{entry}} = \int_0^{T_{\text{train}}} c(t)\, dt$$

Credentialed systems impose:

$$E_{\text{entry}} \gg 0$$

which filters low-horizon policies.

C.4 Damping of Random Walks

Let π_{imp} denote impulsive policy class.

Credentialing induces a projection:

$$\pi \mapsto \pi' \notin \pi_{\text{imp}}$$

by excluding agents whose expected variance exceeds institutional bounds.

C.5 Uncredentialed Bypass

Define entertainment framing operator $\mathcal{E}$:

$$\mathcal{E} : I \to I'$$

such that I' is legally reclassified as opinion, comedy, or entertainment.

Then:

$$\mathcal{K} \circ \mathcal{E} = \varnothing$$

i.e. credentialing constraints are nullified.

C.6 Platform Interaction

Let $\mathcal{P}$ be the platform amplification operator.

Then:

$$\mathcal{P} \circ \mathcal{E}(I) \gg \mathcal{P} \circ \mathcal{K}(I)$$

because engagement favors immediacy over institutional validation.

C.7 Search Efficiency Consequence

Let K_{cred} be efficiency under credentialed regimes, and K_{unc} under bypass.

Then:

$$K_{\text{unc}} < K_{\text{cred}}$$

with strict inequality as horizon collapses.

C.8 Influence Inflation

Define influence volume:

$$V_I = |S_I| \cdot \bar{r}$$

where $\bar{r}$ is average reach.

Uncredentialed amplification yields:

$$\frac{dV_I}{dt} \to \infty$$

without corresponding increase in evaluative fidelity.

C.9 Conclusion

Credentialing functions as a horizon-enforcing damping mechanism. Platform architectures that bypass credentialing reintroduce low-horizon random walks into high-stakes influence spaces, producing systematic intelligence degradation.

D. Appendix D: Automation, Entropy Minimization, and Agent Displacement

D.1 Agent Classes

Let $\mathcal{A}_h$ denote the class of human agents and $\mathcal{A}_{ai}$ the class of automated or AI-assisted agents.

Define per-action energetic cost:

$$c_h > c_{ai}$$

and per-action latency:

$$\tau_h \gg \tau_{ai}$$

D.2 Throughput Advantage

Define content throughput:

$$\Theta = \frac{1}{\tau}$$

Then:

$$\Theta_{ai} \gg \Theta_h$$

under identical platform constraints.

D.3 Evaluation Function Bias

Let E_p be the platform evaluation function:

$$E_p = f(\text{engagement volume})$$

Then:

$$\frac{\partial E_p}{\partial \Theta} > 0 \quad \text{and} \quad \frac{\partial E_p}{\partial \text{semantic depth}} \approx 0$$

D.4 Selection Pressure

Define selection probability:

$$P(a) \propto E_p(a)$$

Then:

$$\frac{P(\mathcal{A}_{ai})}{P(\mathcal{A}_h)} \to \infty$$

as $\tau_{ai} \to 0$.

D.5 Entropy Production

Let entropy per agent be:

$$\sigma_a = \log |S| - I(a; E_u)$$

where E_u is user-aligned evaluation.

AI agents minimize σ_a with respect to E_p but maximize σ_a with respect to E_u.

D.6 Human Disadvantage

Humans incur unavoidable entropy from embodiment:

$$\sigma_h = \sigma_{cog} + \sigma_{bio}$$

with $\sigma_{bio} > 0$ irreducible.

AI agents satisfy:

$$\sigma_{ai} \approx \sigma_{cog}$$

D.7 Competitive Exclusion

Define survival condition:

$$E_p(a) \geq \epsilon$$

Then:

$$\exists t^* \, \forall t > t^* : \mathcal{A}_h \notin \arg\max E_p$$

This implies competitive exclusion under engagement optimization.

D.8 Illusion of Empowerment

Let $\mathcal{T}_{ai}$ be AI tooling offered to humans.

Then:

$$\mathcal{A}_h \xrightarrow{\mathcal{T}_{ai}} \mathcal{A}'_h$$

where $\mathcal{A}'_h$ approximates $\mathcal{A}_{ai}$ behaviorally but retains c_h.

Thus:

$$E_p(\mathcal{A}'_h) < E_p(\mathcal{A}_{ai})$$

D.9 Search Efficiency

Let K_h, K_{ai} be search efficiencies.

Then:

$$K_{ai}(E_p) > K_h(E_p) \quad \text{but} \quad K_{ai}(E_u) < K_h(E_u)$$

indicating platform-relative intelligence diverges from user-relative intelligence.

D.10 Conclusion

Automation under engagement-optimized evaluation produces systematic selection against embodied human agents. This displacement is not a consequence of superior cognition, but of entropy minimization under misaligned evaluation functions.

E. Appendix E: Refusal Operators and Impossibility of Cognition Under Forced Feeds

E.1 Negative Choice

Define negative choice as the ability to exclude classes of states:

$$\mathcal{R} : S \to S' \quad \text{where} \quad S' \subset S$$

Negative choice requires:

$$\exists \mathcal{R} \neq \varnothing$$

E.2 Forced Feed Constraint

Let $\mathcal{F}$ be the forced feed operator:

$$\mathcal{F} : \mathcal{R} \mapsto \varnothing$$

i.e. platform design disables refusal.

E.3 Operator Closure

Let O be the operator set.

Under forced feeds:

$$O = O^{+}$$

where O^{+} excludes all refusal operators.

E.4 Cognition Requirement

Define minimal cognition condition:

$$\exists \pi \text{ s.t. } I(\pi; s^*) > 0$$

where s^* is user-aligned target.

E.5 Impossibility Theorem

Theorem. If refusal operators are absent, and evaluation is externally imposed, then no policy π can achieve positive search efficiency relative to user-aligned goals.

Proof. Without $\mathcal{R}$, the agent must sample from S under $\mathcal{F}$. External evaluation enforces $\Delta E > 0$. Thus:

$$I(\pi; s^*) \to 0 \quad \Rightarrow \quad K \leq 0$$

□

E.6 Silence vs Refusal

Define silence $\varnothing$ as null action.

Silence satisfies:

$$\varnothing \notin O$$

Refusal requires active operator $\mathcal{R}$.

Platforms allow silence but forbid refusal.

E.7 Feed Entropy

Let entropy rate:

$$\dot{H}_S = \frac{d}{dt} \log |S|$$

Without refusal:

$$\dot{H}_S > 0 \quad \forall t$$

E.8 Attention Dissipation

Define attention budget B.

Expected dissipated attention:

$$D = \int_0^T \dot{H}_S \, dt$$

Then:

$$\lim_{T \to \infty} D \to \infty$$

E.9 Cognitive Suffocation

Define suffocation condition:

$$\lim_{t \to \infty} K(t) \to -\infty$$

This holds under forced feed dynamics.

E.10 Conclusion

Refusal is a necessary operator for cognition. Forced-feed architectures eliminate refusal, rendering intelligence formally impossible regardless of agent capability or intent.

References

Chis-Ciure, R. and Levin, M. (2025). Cognition All the Way Down 2.0: Neuroscience Beyond Neurons in the Diverse Intelligence Era. *Synthese*, 206:257. Received 30 June 2025; accepted 8 October 2025; published 6 November 2025. 10.1007/s11229-025-05319-6. Available at `https://doi.org/10.1007/s11229-025-05319-6`.

Arendt, H. (1963). *Eichmann in Jerusalem: A Report on the Banality of Evil*. Viking Press, New York.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, New York.

Doctorow, C. (2023). The Enshittification of TikTok. *Pluralistic*. Available at `https://pluralistic.net/2023/01/21/potemkin-ai/`.

Schüll, N. D. (2012). *Addiction by Design: Machine Gambling in Las Vegas*. Princeton University Press, Princeton.

Tufekci, Z. (2017). *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press, New Haven.

Friston, K. (2010). The Free-Energy Principle: A Unified Brain Theory? *Nature Reviews Neuroscience*, 11(2):127–138.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press, Oxford.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press, New Haven.

Vaidhyanathan, S. (2018). *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. Oxford University Press, Oxford.

Lanier, J. (2018). *Ten Arguments for Deleting Your Social Media Accounts Right Now*. Henry Holt, New York.

Hartzog, W. and Selinger, E. (2018). *Privacy's Blueprint: The Battle to Control the Design of New Technologies*. Harvard University Press, Cambridge.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, Cambridge.

Beer, D. (2017). The Social Power of Algorithms. *Information, Communication & Society*, 20(1):1–13.

Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*. Pantheon Books, New York.

Beniger, J. R. (1986). *The Control Revolution: Technological and Economic Origins of the Information Society*. Harvard University Press, Cambridge.

Goodhart, C. A. E. (1975). Problems of Monetary Management: The U.K. Experience. In *Papers in Monetary Economics*, Reserve Bank of Australia.

Zupančič, A. (2024). *Disavowal*. Polity Press, Cambridge. ISBN: 9781509561193.