The Third Enclosure

Ranking Incentives, Platform Centralisation, and the Terminal Disintermediation of the Open Web

A white paper on the Branch network (repository: `alternet`)

Draft v1.0 — 1 August 2026

Abstract

The open web is not dying of technical obsolescence. It is dying of a sequence of incentive substitutions, each of which was individually rational for the actor that made it and collectively fatal to the commons that all of them depend on. This paper reconstructs that sequence in three enclosures. The **first** (c. 1998–2010) placed a link-counting proxy at the discovery layer and funded it with advertising, creating a market in the proxy and an economy — search engine optimisation — whose optimal strategy was never "write something good." The **second** (c. 2006–2020) moved authorship inside platforms that were structurally hostile to the hyperlink, converting publishers into tenants and then withdrawing the referral rent: Facebook's referrals to a stable panel of 792 news sites fell 58% between March 2018 and March 2024, and from 30% to 7% of external page views. The **third**, now underway, replaces the link with the generated answer. In the first four months of 2026, 68.01% of US Google searches ended without a click, up from roughly 45% a decade earlier; Google Search revenue nonetheless grew 19% year-over-year to *60.4bn in Q1 2026. Crawl-to- referral ratios — pages taken per visitor returned — now range from roughly 5:1 for Google to 2,000–4,600:1 for the largest AI crawlers. Wikipedia reports an 8% decline in human pageviews; Stack Exchange question volume has fallen roughly 29-fold from its 2017 peak; small publishers have lost about 60% of search referrals in two years.*

We argue that the third enclosure is *categorically* different from the first two, not merely worse. The first two enclosures taxed the flow of attention to independent publishers but preserved a positive marginal return to publishing, because the intermediary still needed to route a human to a page in order to monetise. The synthesis layer removes that requirement. It preserves its dependence on the substrate — original human-authored text — while driving the substrate's marginal compensation towards zero. We formalise this as a provision failure in a commons whose good is non-rival but whose *attention rents* are rival, and show that the supply side's rational response (robots.txt closure, infrastructure-level default blocking, bilateral licensing) does not reopen the commons but completes its enclosure by replacing free crawlability with a licensing market that only incumbents can enter.

Cutting across those three enclosures is a second, *vertical* axis that the discovery-layer account alone cannot explain: the client stack through which the web is rendered, measured, and monetised. Chrome holds roughly 70% of browsers and Chromium engines underlie more than three-quarters of web sessions; the mobile duopoly is approximately 73% Android and 27% iOS; YouTube took a platform-best 13.8% of *all* US television watch-time in May 2026. We argue that this vertical integration is what made each horizontal enclosure unexitable, because historically the exit from a discovery monopoly has been a client act — install a different browser, subscribe to a feed, run a content blocker — and every such exit now runs through infrastructure owned by the party being exited. Three developments in the eighteen months to August 2026 make the point concretely: Chrome 150 removed the last override for Manifest V2 on 30 June 2026, ending full-capability content blocking in the dominant browser,

which is operated by the largest advertising company; twenty-nine months after the Digital Markets Act required it, *zero* browsers ship an alternative engine on iOS; and Chrome shipped an agentic browsing mode in January 2026, converting the last neutral client into an intermediary with privileged access to the authenticated session. We also show that the incumbent's most robust ranking signals are client-derived rather than algorithmic, which implies something uncomfortable for every independent search effort: a competitive discovery layer cannot be built without a client.

We then make the paper's constructive claim. The failure at the root of the sequence is a **ranking- function failure**, not a technology failure: PageRank solved the 1998 problem of assessing quality without being able to read, and every pathology downstream is an artefact of that proxy. That constraint no longer binds. The strongest available evidence is adverse to link-based authority and favourable to model-based judgment: in the largest controlled comparison of web-quality filters published, selecting the top 20% of pages by PageRank scored *below the un-filtered baseline* and was the worst filter tested, while learned quality classifiers scored highest; and LLM quality judgments have near-zero correlation with perplexity, indicating they carry information no cheap statistical proxy captures. The same technology that is severing the discovery bargain is the first technology capable of replacing the proxy that corrupted it.

The final sections specify a counter-architecture — a post-PageRank publishing and discovery network in which quality is assessed on arrival and confirmed or revoked by engagement, publishing obligates hosting, and the network ships with a distilled index of the pre-existing web — and state the programme's principal hypotheses together with the conditions under which each would be falsified. We are explicit that the central mechanism is unvalidated: no published work measures LLM-judge reliability against human raters on web-page quality at scale, every cheap-at-scale filter in the literature is validated as a bottom-decile remover rather than a top-decile identifier, and the adversarial results against LLM judges are severe. The argument of this paper is not that the counter-architecture will work. It is that the diagnosis is sound, that the remedy space is narrow, and that the experiment is worth running before the substrate is gone.

Keywords: information retrieval, platform economics, digital commons, generative search, zero-click, LLM-as-judge, peer-to-peer publishing, enclosure

---

Note on evidential standards

This paper follows the grading convention used in the project's evidence base (`research.md`). Every empirical claim is marked:

- **[solid]** — primary source, peer-reviewed publication, regulatory filing, or court record.
- **[reported]** — credible secondary reporting, or vendor telemetry from a named panel with a described methodology.
- **[thin]** — single source, vendor self-report without published method, or contested.

Figures are dated because most of them move. Where sources disagree — and on crawler telemetry they disagree by an order of magnitude — we report the range and say which measurement regime produced each end of it. Section 11 collects the threats to the paper's own validity. Appendix A is a claim ledger; Appendix B states the falsifiable propositions on which the constructive argument depends.

1. Introduction

1.1 The problem

In February 1998, in an appendix to the paper that introduced PageRank, Sergey Brin and Lawrence Page wrote a warning about the business model they would shortly adopt:

> "Currently, the predominant business model for commercial search engines is advertising. The goals > of the advertising business model do not always correspond to providing quality search to users… > we expect that advertising funded search engines will be inherently biased towards the advertisers > and away from the needs of the consumers." (Brin & Page 1998, Appendix A) **[solid]**

They went further, observing that "the better the search engine is, the fewer advertisements will be needed for the consumer to find what they want," and concluding that "it is crucial to have a competitive search engine that is transparent and in the academic realm."

Twenty-eight years later the appendix reads as an unusually precise forecast, but the mechanism it identified is only half of the story, and it is the less interesting half. Advertising bias is a conflict of interest, and conflicts of interest are ordinary. The deeper problem was in the body of the paper, not the appendix: PageRank inferred quality from inbound links because in 1998 a machine could not read a page. That was an engineering necessity and it was, at the time, a brilliant one. But a proxy that is used to allocate attention becomes a market, and a market in the proxy detaches the proxy from the thing it was proxying for. The listicle, the nine-hundred-word preamble before the recipe, the programmatic comparison page: these are not failures of the ranking system. They are its equilibrium output.

1.2 The claim

This paper makes five claims.

1. **Descriptive.** The degradation of the open web is best understood as three sequential enclosures of a commons — at the *ranking* layer, the *aggregation* layer, and now the *synthesis* layer — rather than as a single event, a moral failure of any one firm, or the natural senescence of a technology.
2. **Analytical.** The third enclosure is structurally different from its predecessors. The first two redistributed attention rents and made independent publishing harder; the third removes the intermediary's need to deliver a human to a page at all, while leaving intact its need for the pages. This decouples extraction from reciprocity. Under the pressure of that decoupling, the supply side closes — and the closure is executed by infrastructure defaults and bilateral licensing, which entrench the same incumbents.
3. **Structural.** None of the three enclosures could have held without a second, vertical axis of control: the browser, the mobile operating system, and the video runtime. The vertical stack is not a fourth enclosure in the sequence; it is the enforcement layer that made the other three unexitable, because the historical exit from a discovery monopoly has always been a client act. The corollary matters more than the complaint: the incumbent's most robust ranking signals are client-derived, so a competitive discovery layer cannot be built without a client.

4. **Diagnostic.** The root cause of the sequence is a ranking-function failure that was rational in 1998 and is now obsolete. Link-based authority is measurably a poor quality signal, and model-based judgment is measurably a better one; the binding constraint that made the proxy necessary has been removed by the same class of technology that is currently completing the enclosure.
5. **Constructive.** A discovery architecture that assesses quality on arrival, treats engagement as a validator rather than as the primary signal, ties publication rights to hosting obligations, ships with a distilled corpus of the pre-existing web, and — following from (3) — ships as its own client, is a coherent response to (4). It is also unproven in its central mechanism, and we specify what would falsify it.

1.3 What this paper is not

It is not a neutrality claim. Every discovery system encodes an editorial position; PageRank's was "whatever the link economy rewards," and its principal defect was that this position was never stated and therefore never argued with (§9.6). It is not a claim that any firm acted illegitimately beyond what courts have found; the two US findings of monopolisation against Google are cited as record evidence, not as moral argument. And it is not a prediction that the open web ends on a date. The argument is about the direction and the mechanism of a gradient, and about the fact that the gradient is now, for the first time, self-consuming.

1.4 Structure

Section 2 sets out the theoretical framework and the units of analysis. Sections 3, 4 and 5 treat the three enclosures in turn — the horizontal axis — each with its mechanism, its evidence, and its measurable residue. Section 6 turns to the vertical axis and examines the browser, the mobile operating system, and long-form video as the stack that made the horizontal enclosures enforceable. Section 7 gives the formal statement of why the third enclosure is terminal and takes the strongest counterarguments seriously. Section 8 makes the diagnostic argument that the failure is a ranking-function failure. Section 9 presents the counter-architecture. Section 10 states its known hazards. Section 11 states the threats to this paper's validity. Section 12 concludes.

---

2. Framework

2.1 The web as a provision problem, not an access problem

The standard commons vocabulary fits the web badly if applied naively. Digital content is non-rival: your reading of a page does not consume it. Ostrom's canonical common-pool resources — fisheries, irrigation systems, pastures — are defined by subtractability, and web pages have none. It is therefore tempting to conclude that the web cannot be over-appropriated and that "enclosure" is a metaphor rather than an analysis. That conclusion is wrong, and the reason it is wrong is the whole argument of this paper.

The rival good is not the content. It is the **attention rent**: the flow of human attention that arrives at a page and can be converted into revenue, reputation, subscriptions, employment, or the diffuse social returns that motivate non-commercial publishing. Attention is strictly rival and strictly budget-constrained. The commons at issue is therefore not the stock of pages; it is the *coupling* between the consumption of a page's information and the delivery of a rent to its author. Discovery intermediaries are the institutions that maintain that coupling. Boyle's (2003) "second

enclosure movement" and Benkler's (2006) account of commons-based peer production both describe systems whose provision depends on such couplings; Ostrom's (1990) design principles are, in their essence, principles for keeping appropriation and provision in the same institution.

This yields the paper's central analytic object. Define, for a period$t$:

-F_t— *the flow of new, original, human-authored, openly addressable content;* -A_t— *human attention routed to that content by discovery intermediaries;* -E_t— *extraction: the value the intermediary derives from the content;* -$\rho_t = A_t / E_t$— *the* ***reciprocity ratio****, the attention returned per unit of value extracted.*

Provision is a function of expected return, which for commercial publishers is proportional toA_t*and for non-commercial publishers is proportional to a mixture of*A_t*and intrinsic motivation*M_t*:*$F_{t+1} = f(A_t, M_t)$*, with*$\partial f/\partial A > 0$*over the relevant range. An intermediary is sustainable with respect to the commons if*ρ_t*stays bounded away from zero while*E_t*grows. It is extractive if*E_t*grows while*$\rho_t \to 0$*. The three enclosures are three successive reductions in*ρ*, and the third is the first in which*ρ*can go to zero without the intermediary losing anything it needs in the same period — which is precisely why it is terminal and why it is not self-correcting on any timescale shorter than the depletion of*C_t*, the accumulated stock.*

2.2 Enclosure as a repeated form

We use *enclosure* in Boyle's sense: the conversion of a resource governed by open access and informal norms into one governed by property rights, defaults, and contracts, typically justified by an efficiency argument that is often correct in the short run and typically accompanied by a distributional shift that is rarely discussed. Each of the three enclosures below follows the same four-step form:

1. **A proxy or interface is introduced** that genuinely improves the user experience.
2. **The proxy becomes the default channel**, and therefore a market.
3. **The intermediary internalises the surplus** by moving the boundary of its own product outward — from ranking links, to hosting content, to generating the answer.
4. **The producers adapt**, and their adaptation degrades the substrate that the intermediary needs.

Step 4 is the part that distinguishes analysis from complaint. In the first enclosure the adaptation was SEO. In the second it was platform-native, unlinkable content. In the third it is closure of the crawl surface and — the pathology that closes the loop — mass synthetic publication.

2.3 Two axes, not one sequence

A purely temporal account of the three enclosures leaves an obvious question unanswered. Each enclosure degraded the product for some substantial constituency — searchers who wanted a page rather than an answer, publishers who wanted a referral, readers who wanted a feed they controlled. Ordinary market theory predicts entry. Entry happened: Blekko, Cuil, DuckDuckGo, Neeva, Brave, Mastodon, PeerTube, the entire federated stack. Almost all of it failed, and it failed for reasons that were not about ranking quality.

The missing variable is the **vertical axis**: the client stack through which the web is rendered, measured, and monetised. We define it as three layers, each with its own concentration:

Layer	Function	Concentration (2026)
Video runtime	Hosts the format the open web could not afford to serve	YouTube: 13.8% of *all* US TV watch-time, largest single distributor
Browser	Renders pages, observes behaviour, enforces defaults, now acts as agent	Chrome ~70%; Chromium engines >75% of sessions
Mobile OS	Gates what may be installed, and what an installed program may do	Android ~73% / iOS ~27%; no third option

The horizontal axis answers *who decides what you find*. The vertical axis answers *who decides what you can run*. The paper's structural claim (§1.2, claim 3) is that the second is what made the first durable, and that this is not incidental: the historical exit from a bad discovery layer has always been an act performed at the client. You installed a different browser, you set a different default, you subscribed to a feed, you ran a content blocker, you used a third-party client with a different ranking. Every one of those exits now runs through infrastructure owned by the party you would be exiting. Section 6 is the evidence for that claim.

2.4 Theoretical inheritance

We draw on four literatures, and we should be explicit that each supplies a different part of the argument.

- **Information retrieval and network science** supply the mechanism of the first enclosure: preferential attachment (Barabási & Albert 1999), the entrenchment effect of search engines on page popularity (Cho & Roy 2004; Cho, Roy & Adams 2005), and the empirical evaluation of ranking signals (Li et al. 2024; Penedo et al. 2024; Sachdeva et al. 2024).
- **Platform studies and political economy** supply the second and much of the vertical axis: Zittrain's (2008) distinction between *generative* systems, which accept unvetted contributions from anonymous parties, and *tethered appliances*, which run only what the vendor permits and can be changed under the user after purchase — the single most direct antecedent for §6, written a year after the first iPhone and before any of the concentration figures above existed. Alongside it: Gillespie (2018) on the custodial turn, van Dijck et al. (2018) on platform society, Srnicek (2017) on platform capitalism, Wu (2016) on attention markets, and Doctorow's (2023) "enshittification" as an informal but unusually well-specified life-cycle model.
- **Experimental social science** supplies the warning that applies to *any* proposed remedy, including ours: Salganik, Dodds & Watts (2006) demonstrated that when a ranking system exposes accumulated popularity, outcomes become both more unequal and less predictable, and Salganik & Watts (2008) showed that an inverted ranking is substantially self-fulfilling.
- **Law and regulation** supply the record: *United States v. Google LLC* (D.D.C. 2024, 2025); *United States v. Google LLC* (E.D. Va. 2025) on ad tech; the EU Digital Markets Act and Digital Services Act; and the emerging opt-out vocabularies (IETF AIPREF; Cloudflare Content Signals).

---

3. The first enclosure: the ranking layer, c. 1998–2010

3.1 A proxy adopted under a constraint that no longer exists

PageRank's premise is that a link is an endorsement and that endorsements from well-endorsed pages count more (Page et al. 1999). This was the best available answer to a question that could not be answered directly, because in 1998 no machine could read a document and judge whether it was any good. The proxy has two properties that determine the shape of everything downstream.

It is a trailing signal. Links accrue only after visibility, and visibility is allocated by links. New content is structurally invisible until it is old. Cho, Roy & Adams (2005) named this the **entrenchment effect** and measured it: under popularity-biased ranking, new pages take substantially longer to be discovered than they would under an unbiased regime, and the authors proposed randomised rank promotion specifically to correct it — an explicit acknowledgement, from inside the IR literature and inside Google's own academic orbit, that the ranking function suppresses new work as a structural matter rather than as an accident **[solid]**. Cho & Roy (2004) had already shown that search engines amplify the popularity of already-popular pages **[solid]**.

It is a market. Because links are the currency of visibility, they are bought, sold, farmed, exchanged, and solicited. The industry that grew around this is search engine optimisation, and its optimal strategy has never been "produce the best document"; it has been "produce the document that best satisfies the proxy." Google's countermeasures are themselves the evidence: the `rel=&quot;nofollow&quot;` attribute, introduced jointly with Yahoo and MSN in January 2005, exists because the link graph had become a spam surface within seven years of PageRank's publication **[solid]**.

An honest treatment must note a genuine counter-finding. Fortunato, Flammini, Menczer & Vespignani (2006) argued in PNAS that the "googlearchy" hypothesis overstates the case: because users issue topically narrow queries, search engines have a partially *egalitarian* effect, directing traffic into the long tail more than a pure popularity model predicts **[solid]**. We take this seriously. The claim we defend is not that search engines produced maximal inequality; it is that the ranking signal was structurally trailing and structurally tradeable, and that these two properties determined what kind of content was rational to produce.

3.2 The empirical case against the proxy, from an unexpected direction

The strongest evidence against link-based authority as a *quality* signal comes not from web search research but from the language-model data-curation literature, where filters must be evaluated against a downstream outcome and are therefore compared in controlled bake-offs.

In DataComp-LM (Li et al. 2024), the largest published controlled comparison of web-quality filters, each candidate filter was used to select a training subset at fixed scale (7B parameters, 1× compute) and evaluated on a standard benchmark suite **[solid]**:

Filter	CORE	EXTENDED
RefinedWeb reproduction (baseline, no quality filter)	27.5	14.6
Top 20% by PageRank	**26.1**	**12.9**
SemDedup	27.1	13.8
Classifier on BGE embeddings	27.2	14.0
Ask-LLM	28.6	14.3
Perplexity filtering	29.0	15.0
Top-k average logits	29.2	14.7
fastText classifier (OH-2.5 + ELI5)	**30.2**	**15.4**

Selecting the top fifth of the web by PageRank produced a *worse* corpus than not filtering at all, and was the worst of every method tested. The best-performing filter was a cheap linear classifier whose positive class was instruction-tuning data and highly-rated explanatory Reddit posts.

Two caveats keep this from being overclaimed. First, the target is training-corpus quality as measured by benchmark scores, not human satisfaction with search results; a filter that improves MMLU is not thereby demonstrated to improve a reader's experience. Second, PageRank as computed over a Common Crawl snapshot is not Google's live ranking function, which has incorporated hundreds of other signals for two decades. What the result does establish is narrow but important: **the link graph, by itself, does not encode document quality**, and treating it as though it does is measurably worse than doing nothing.

3.3 The default, and the price of it

A proxy becomes an economy only if it is the default window. The relevant facts are now in the judicial record.

On 5 August 2024, Judge Amit Mehta held that Google had unlawfully maintained monopolies in general search services and general search text advertising **[solid]**. The court found that Google paid **26.3 billion in 2021 alone** for default placement — roughly *18–20 billion of it to Apple, under an arrangement sharing 36% of Safari search advertising revenue — up from approximately*7.1 billion in 2014 **[solid]**. In the remedies opinion of 2 September 2025, the court declined to order divestiture of Chrome or Android, barred exclusive distribution agreements covering Search, Chrome, Google Assistant and the Gemini app for six years, and required Google to syndicate search and to share portions of its index and user-interaction data with qualified competitors **[solid]**. Google noticed its appeal on 16 January 2026; the Department of Justice cross-appealed by 3 February 2026, declining to challenge the refusal to order divestiture but asking the D.C. Circuit to vacate the denial of a payments ban **[reported]**. Separately, on 17 April 2025, Judge Leonie Brinkema found that Google had unlawfully monopolised publisher ad server and ad exchange markets **[solid]**.

The distribution position those payments purchased, as of May 2026 (StatCounter, via secondary compilation) **[reported]**: approximately **90.4%** of global all-device search, **95.5%** of mobile search, **70.25%** of the global browser market for Chrome, and Chromium engines underlying more than three-quarters of web sessions; Android holds roughly 72.8% of mobile operating systems. Alphabet reported Q1 2026 revenues of ***109.9bn***, of which Google advertising was **77.25bn** and “Google Search & other” was ***60.4bn***, up 19% year over year **[solid — SEC filing]**.

Two independent post-mortems from failed competitors establish that this is a distribution problem rather than a product problem. Neeva's founders, on shutting down in June 2023: "It is one thing to build a search engine, and an entirely different thing to convince regular users of the need to switch… **convincing users to pay for a better experience was actually a less difficult problem compared to getting them to try a new search engine in the first place**" **[solid]**. Blekko's chief executive: "Ultimately, lack of traffic. During Blekko's lifetime Google went from paying people less than10M/quarter to send their search traffic to Google to over *4B/quarter"* ***[reported]***.

Payments are the visible half of the distribution position and the litigated half, which is why they dominate the discussion. The other half is not purchasable and not, so far, justiciable: the intermediary also builds the browser through which the results are read, the operating system on which the browser runs, and the video platform on which a growing share of explanatory content now lives. Section 6 treats that as a distinct axis of the argument, because it turns out to explain not only why the enclosures happened but why they were never reversed.

3.4 Interface enclosure: the slow internalisation of the click

The final move of the first enclosure was not algorithmic but presentational, and it ran continuously from the OneBox through Universal Search (2007), the Knowledge Graph (2012), and featured snippets (2014) to the present. Each increment answered more of the query on the results page. The measurable residue is the zero-click rate:

Year	US zero-click share of Google searches	Panel
~2016	~45%	Jumpshot (defunct)
2019	49–50.33%	Jumpshot
2020	64.82%	Similarweb
2024	60.45% (US); 59.7% (EU)	Datos
Jan–Apr 2026	**68.01%**	Similarweb

[reported] — and see §11 on the methodological hazard of stitching four panels into one series. SparkToro, which published the 2019, 2024 and 2026 studies, states the discontinuity explicitly. The 2024→2026 comparison is the most internally consistent: searches producing at least one click of any kind (organic, paid, or to Google's own properties) fell 9.51 percentage points, a 22.9% relative decline, while the share of searches followed by *another search* rose 7.2 points.

That last number is the one to hold onto. It is the signature of a system that has learned that the most profitable thing to do with a query is to produce another query.

---

4. The second enclosure: the aggregation layer, c. 2006–2020

4.1 Mechanism: hostility to the hyperlink as a design objective

The platform era did not attack the open web; it made the open web unnecessary for the majority of publishing activity. Facebook's News Feed (2006), YouTube's recommendation surface, and Twitter's timeline each offered producers something the web could not: distribution without an audience- building period. The price was that the content lived inside the platform, and platform-native content has three properties fatal to the commons as defined in §2.1.

It is **non-addressable in practice** (deep links are unstable, gated, or hostile to crawlers); **non-portable** (there is no export that preserves the audience); and **non-linkable outward without penalty**, because every mature engagement-optimised feed ranks down posts that send users away. That last property is not conspiracy; it is the direct consequence of optimising a session-length objective. A recommender that maximises time-on-platform will learn, without being told, that outbound links are costly.

The consequence for the substrate is that between roughly 2010 and 2020, an enormous share of the world's new public expression — the share that a decade earlier would have been blog posts, forum threads, and personal sites — was produced inside systems from which it cannot be indexed, cited, archived, or carried away. The commons did not shrink so much as it stopped growing while the population of authors exploded.

4.2 The API enclosure sequence

The platform era's characteristic instrument was the deprecation of the programmable interface. The sequence is worth stating as a sequence because it shows an industry converging on the same policy independently:

- **Google Reader** shut down 1 July 2013, removing the largest consumer client for RSS and, with it, the principal non-algorithmic distribution channel for independent publishing **[solid]**.
- **Twitter API v1.1** (2012) imposed token caps that made full-featured third-party clients non-viable; the 2023 pricing restructure ended free API access entirely and priced enterprise tiers in the tens of thousands of dollars per month, closing the research access that had made Twitter the most-studied social system in the literature **[reported]**.
- **Reddit** repriced its API in mid-2023, ending third-party clients including Apollo, and in 2024 entered a reported ~60m/year data licensing agreement with Google **[reported]**. The same corpus that had been free to crawl became a bilaterally licensed asset — a direct rehearsal of the licensing regime described in §5.5.
- **AMP** (2015) offered publishers preferential placement in exchange for serving Google-hosted, Google-styled copies of their pages; allegations concerning AMP’s interaction with header bidding formed part of the state ad-tech complaints **[reported]**.
- **Meta and news**: in response to Canada’s Online News Act (C-18), Meta blocked news links in Canada from August 2023 and, as of 2026, has not restored them **[reported]**. The revealed preference is unambiguous: at a positive price for linking, a platform will stop linking.

4.3 Measurable residue: the withdrawal of the referral rent

The platform bargain was: give us your content and your posting effort, and we will send you traffic. The withdrawal of that traffic is the second enclosure's clearest empirical signature. Using Chartbeat's stable panel of 792 news and media sites tracked since 2018, with Similarweb corroboration **[reported]**:

- Facebook referrals fell from **1.3 billion (March 2018) to 561 million (March 2024)**, −58% over six years and **−50% in the final twelve months**.
- As a share of page views from search, social and other external sources, Facebook fell from **30% (March 2018) to 7% (March 2024)**.
- For the 316 smaller publishers in the panel (<10,000 average daily page views), Facebook referrals stood at **2% of their March 2018 level**.
- X/Twitter referrals to news brands fell **65% between October 2022 and mid-2025**, and 75% from January 2019 (128.7m to 31.8m monthly).
- Google News referrals in July 2025 (107m) were roughly at their January 2019 level and 28% below July 2023.

The distributional pattern — small publishers losing an order of magnitude more than large ones — is the same pattern that recurs in the third enclosure (§5.2), and it is the pattern that a preferential-attachment model predicts.

4.4 The federated alternative and its ceiling

The decentralised responses to platform enclosure are instructive precisely because they were technically sound and demographically marginal. Mastodon's monthly active users peaked at 2,737,163 in November 2022 and stood at 839,406 in July 2026, a decline of 69.3% from peak, with approximately 93% of registrations inactive **[reported]**. What worked in the fediverse was not the protocol but the governance shape: instance administrators are accountable humans with a takedown button. What did not work was cold start, discovery, and the absence of any economics for hosting.

The lesson we take from this into §9 is not "decentralisation fails." It is that **decentralisation without an economics for hosting and without a discovery layer reproduces the ceiling every time**, and that the design constraint is therefore economic and editorial before it is cryptographic.

---

5. The third enclosure: the synthesis layer, 2023–2026

5.1 Mechanism: the answer replaces the link

Generative search does not rank documents; it consumes them and emits a synthesis. Formally, the intermediary's product function changes from $g(\text{query}) \to \text{ranked list of pointers}$ to $g(\text{query}, \text{corpus}) \to \text{answer}$. The pointer was the unit of reciprocity. Its removal is not a degradation of the bargain; it is the deletion of the bargain's mechanism.

5.2 Measured effects on click-through and on publishers

Behavioural evidence. Pew Research Center analysed the browsing data of 900 US adults for March 2025 (68,879 unique Google searches, of which 12,593 — 18% — produced an AI Overview) **[solid]**:

- Users clicked a traditional search result in **8%** of visits where an AI summary appeared, versus **15%** where none did.
- Users clicked a link *inside* the AI summary in **1%** of such visits.
- Browsing sessions ended entirely after **26%** of visits with a summary versus **16%** without.

Aggregate clickstream evidence. In January–April 2026, **68.01%** of US Google searches ended without any click, versus 60.45% in 2024 and roughly 45% a decade earlier **[reported]**. Ahrefs measured the presence of an AI Overview as correlating with a **58% lower click-through rate** for the top-ranking page by December 2025, up from a 34.5% decline measured in April 2025 **[reported]**. Similarweb reports Google's AI Mode referring users out on **1.6–2.5%** of queries versus 17–19% for traditional search **[reported]**.

Publisher-side evidence, and its distribution. Chartbeat data reported by Axios (July 2026) show Google Search traffic to publishers down **34% year over year**, and — the finding that matters most for the commons argument — over two years, **small publishers lost about 60% of search referrals, medium publishers 47%, and large publishers 22% [reported]**. Individual cases are consistent: one analysis found organic search traffic to major technology publishers down 58% since 2024, with a single well-known title falling from 5.3m organic visits in February 2024 to 790,000 in January 2026 **[reported, single-vendor estimate]**; an independent site documented losing 70% of its traffic over seven years of prior growth **[thin — self-reported]**.

Commons-specific evidence. The two most important non-commercial knowledge commons on the web report the same shape. The Wikimedia Foundation, after correcting its bot-classification logic in 2025, reported human pageviews down **roughly 8% year over year**, attributing the decline to search engines answering directly from Wikipedia content and to younger users shifting to social video **[solid — first-party]**. Stack Exchange question volume fell from 286,300 in March 2017 to 9,883 in March 2026, a factor of **29 [solid — first-party data explorer]**; December 2025 alone was down 78% year over year on Stack Overflow **[reported]**. Both declines began before ChatGPT — Stack Overflow's from around June 2020, with moderation culture a well-documented contributing cause — and both accelerated sharply after November 2022. Attribution is therefore mixed, and we say so; but the direction and the inflection are not in dispute.

5.3 The decisive datum: extraction rising while reciprocity falls

If the third enclosure were merely a reallocation of traffic, the intermediary's revenue would move with the traffic it distributes. It does not. In the same period in which referrals to publishers fell by roughly a third year over year, Google's Search & other revenue grew **19%** to ***60.4bn*** in Q1 2026, with the chief executive noting "queries at an all time high" **[solid — SEC filing]**. The firm's AI assistant surface reached, by its own statement at its 2026 developer conference, more than one billion monthly users with queries "more than doubling every quarter" **[reported — vendor]**.

This is\rho \to 0*with*E*increasing, and it is the empirical heart of the paper.*

The same decoupling is visible from the infrastructure side, in the ratio of pages crawled to referrals returned. Cloudflare, which fronts roughly a fifth of the web, began publishing crawl-to-refer ratios in 2025. The figures are volatile and the secondary compilations disagree, so we give the range and the regime **[reported]**:

Operator	Crawl-to-refer (2026 range)	Regime
Google	~4.6:1 – 5:1	Crawl feeds a search engine that still emits links
DuckDuckGo	~2:1 – 2.7:1	Leases others' indexes; near parity
Microsoft	~33:1 – 37:1	Mixed search and assistant
Perplexity	~163:1 – 263:1	Assistant with citations
OpenAI	~217:1 – 848:1	Improved as ChatGPT Search matured
Mistral	~84:1 – 3,389:1	Extremely volatile between windows
Anthropic	~2,237:1 – 9,478:1	Training-heavy crawl footprint

Two readings are available and both matter. The pessimistic reading is that the assistant layer takes two to three orders of magnitude more than it returns. The more careful reading is that the ratio is dominated by *training* crawls, which are one-time-ish and amortised, and that search-purpose crawlers (OpenAI's OAI-SearchBot, Anthropic's Claude-SearchBot) behave much more like search engines. That second reading is correct and does not rescue the bargain: it means the reciprocity is conditional on the assistant choosing to cite and the user choosing to click, and Pew measures the latter at 1%.

5.4 The supply side degrades: synthetic publication and the self-consuming loop

The third enclosure has a second-order effect that the first two did not: it lowers the marginal cost of producing text to approximately zero, at the same moment as it lowers the marginal return to producing good text.

Graphite, sampling 55,400 Common Crawl URLs with three detectors each independently validated at under 2% false-positive and false-negative rates, estimates that the share of newly published English web articles that are primarily AI-generated rose from 35.9% twelve months after ChatGPT's release to 48% at twenty-four months, reaching **50.9% in Q4 2025 and 49.9% in Q1 2026** — a plateau across five quarters **[reported]**. Ahrefs, over 900,000 newly crawled pages in April 2025, found 74.2% containing some AI-generated content **[reported]**.

Graphite's stated hypothesis for the plateau is that primarily AI-generated articles do not currently perform well in search. That is a load-bearing observation for §9: the present disincentive to mass-produce is the *absence* of an immediate quality-based reward, which means any system that introduces one — including ours — is introducing the missing incentive to flood it.

Downstream, Shumailov et al. (2024) demonstrated **model collapse**: generative models trained recursively on their own outputs lose the tails of the true distribution and degenerate **[solid]**. Whether collapse binds in practice under mixed real-and-synthetic corpora is contested, and we do not overclaim it. But the direction is not contested: the synthesis layer is degrading the corpus it requires, and the corpus is degrading fastest in exactly the newly published segment that keeps a model current. Chaney, Stewart & Engelhardt (2018) provide the recommender-side analogue — algorithmic confounding, in which training on data generated by one's own recommender increases homogenisation and *decreases* utility over time **[solid]**.

5.5 The counter-enclosure: closure from below, and why it entrenches incumbents

The rational response of publishers is to close the crawl surface, and they are doing it fast.

The Data Provenance Initiative's longitudinal audit of 14,000 web domains underlying C4, RefinedWeb and Dolma found that between April 2023 and April 2024, full robots.txt restriction rose from ~1% to **5–7% of all tokens**, and from under 3% to **20–33% of tokens among the most actively maintained, critical domains** — a relative increase exceeding 500% corpus-wide and 1,000% for head sets. Counting Terms of Service restrictions, **45% of C4 is now restricted [solid — NeurIPS 2024]**.

Then the infrastructure moved. On 1 July 2025, Cloudflare — serving roughly 20% of websites — began blocking AI crawlers **by default** for new domains and launched *pay per crawl*, in which a request without a payment relationship receives `HTTP 402 Payment Required` with a `crawler-price` header **[solid — vendor primary]**. Whatever one thinks of the policy on the merits, note what it does structurally: it converts crawlability from a norm-governed open-access default into a priced, account-mediated, bilaterally negotiated permission.

This is the step that makes the enclosure complete rather than merely painful, and it deserves the paper's sharpest statement:

> The supply side's defence against extraction is to make the web unreadable to anyone who has not > signed an agreement. Google can sign it. OpenAI can sign it. Anthropic can sign it. A graduate > student, a nonprofit archive, a new search engine, and a public-interest researcher cannot. The > counter-enclosure therefore does not restore the commons; it converts an open-access resource into > a licensed input market whose barriers to entry are highest for exactly the entrants that would > discipline the incumbents.

The regulatory record contains the irony in compressed form. Judge Mehta's remedy requires Google to share portions of its index with qualified competitors *because building an independent index is prohibitively hard*, at the same moment as the crawl surface those competitors would need is being closed by publishers defending themselves against Google's AI. Blekko's public estimate of3–6m to reach a 3-billion-page index from a 10-billion-page frontier is the pre-closure number **[reported]**; the post-closure number includes a licensing budget that no new entrant has.

5.6 The agentic turn: removing the human from the loop entirely

The final development, and the one whose consequences are least measured, is that the entity fetching the page is ceasing to be a person.

Google shipped Chrome **auto browse**, a Gemini-3-powered agent able to scroll, click, type and navigate on a user’s behalf, on 28 January 2026, extending to Android in June 2026; WebMCP entered origin trial in Chrome 149; and Google published the Universal Commerce Protocol for agent-executed purchases **[reported — vendor announcements]**. OpenAI’s standalone Atlas browser was retired on 9 August 2026 after nine months, with its agentic capability folded into the ChatGPT desktop application and a Chrome extension **[reported]**.

The significance is not the product churn. It is that the entire monetisation model of the open web — display advertising, affiliate links, subscription conversion, newsletter capture — presupposes a human eye and a human hand. An agent renders the page, extracts the fact, and leaves. It does not see the advertisement, does not convert, does not subscribe, and does not come back tomorrow because it liked the writing. When the majority of page

fetches are agentic, the residual monetisable traffic is not merely smaller; it is qualitatively different, and no amount of “optimise for agents” advice changes the arithmetic. Whether agent access can be conditioned or charged for is now in litigation (*Amazon v. Perplexity*, argued in the Ninth Circuit on 11 June 2026) **[reported]**.

---

6. The vertical axis: the client as the enforcement layer

6.1 Why the vertical axis is necessary to the argument

Sections 3 to 5 describe a discovery layer that got progressively worse for producers and, in important respects, for readers. The obvious objection is the market one: dissatisfaction of that magnitude, sustained over two decades, should have been an entry opportunity. Entry was in fact attempted continuously and at serious expense — Blekko, Cuil, DuckDuckGo, Neeva, Brave Search, and on the publishing side Mastodon, PeerTube, Diaspora, Solid, Scuttlebutt, Beaker, Urbit. The failures do not share a technical cause. What they share is that each required a user to *change what they run*, and the things a user runs are controlled by the incumbents.

This is the sense in which the vertical axis is load-bearing rather than decorative. Historically, every effective exit from a bad discovery layer was a client-side act:

Exit	Client-side act	Current status
Bad results	Install a different browser or change the default	Chrome ~70%; defaults purchased at *26.3bn/yr*
Algorithmic feeds	Subscribe to RSS in a reader	Dominant reader shut down by the incumbent, 2013
Advertising and tracking	Run a content blocker	Full-capability blocking removed from Chrome, June 2026
Hostile app	Use a third-party client with your own ranking	API access repriced out of existence, 2023
Platform lock-in	Run your own node	Architecturally impossible on the dominant computing mode

Note what the right-hand column says as a whole. It is not that the exits were outcompeted. It is that they were *removed*, one at a time, by parties with a direct interest in their absence, and in every case the removal was executed at the client layer rather than the discovery layer. Zittrain (2008) named the general form of this a decade in advance: the migration from *generative* systems, which run unvetted code from anonymous parties, to *tethered appliances*, which run what the vendor permits and can be altered under the user after purchase.

6.2 The browser: measurement instrument, enforcement point, and now agent

The browser is the most consequential of the three layers and the least discussed, because it presents as a neutral viewport. It is not neutral in any of the four functions it performs.

(a) It is a moat, maintained by complexity. The web platform is now large enough that implementing it is a multi-decade undertaking, and the specification is written substantially by the party that benefits from its size. The evidence is the engine census: five significant engines in 2013, three today — Blink, WebKit, Gecko — with Chromium derivatives underlying more than 75% of sessions **[reported]**. Microsoft, a company with effectively unlimited capital and thirty years of browser engineering, abandoned its own engine for Chromium in 2018–2020.

Mozilla's Servo has been in development since 2012 and remains pre-1.0, with a funded thirty-contributor downstream effort (Verso) archived in 2025 explicitly because it could not track Servo's breaking changes. Ladybird — the only genuinely new independent engine in over a decade, written from scratch by a non-profit that has committed never to take search-default revenue — targets a developer *alpha* on Linux and macOS in 2026, beta in 2027, stable in 2028 **[solid — project roadmap]**. Its January 2026 progress note records a revealing detail: it had to add `Chrome/140.0.0.0` and `AppleWebKit/537.36 Safari/537.36` to its user-agent string before Google Search, Gmail and Instagram would serve it real content.

A moat maintained by complexity is different in kind from a moat maintained by network effects, because it cannot be dissolved by a better product. It is why §9.3 argues that a new client must decline to be a web browser.

(b) It is the measurement instrument, and this is the part with direct consequences for anyone building a search engine. Google's most robust ranking signals are behavioural, and it can collect them at population scale only because it operates the client. The Content Warehouse schema disclosed in May 2024 — authenticity confirmed by Google — contains `goodClicks`, `badClicks`, `lastLongestClicks`, squashed/unsquashed pairs over a 13-month rolling window, and Chrome-derived signals; NavBoost's existence and its use of click data since around 2005 was confirmed under oath in 2023 **[solid on the schema, contested on which fields are live]**. The relevant point is not that click signals exist. It is that **dwell time and task completion are robust signals at Google specifically because Google measures them through Chrome and Android on an independently authenticated population**, and are weak, cheaply faked signals for anyone who does not own a client.

This inverts the usual framing of search competition. The incumbent's durable advantage is not its index and not its ranking algorithm; both are, in principle, reproducible. It is the possession of an instrumented population. Judge Mehta's remedy implicitly concedes the point by ordering the sharing of *user-interaction data* alongside the index. The corollary for this project is stated plainly because it cuts against a comfortable assumption: an independent discovery layer that does not ship a client is competing without access to the only signals that have survived two decades of adversarial pressure.

(c) It is the enforcement point for the advertising settlement. The clearest instance closed six weeks before this paper's date. Chrome's migration to Manifest V3 replaced the blocking `webRequest` API with `declarativeNetRequest`, which requires filter rules to be declared in advance and evaluated by Chrome itself, with caps of 330,000 static and 30,000 dynamic rules; full-capability blockers were disabled for most users with Chrome 138 in July 2025; and **Chrome 150, released 30 June 2026, removed the last override flag, with Chrome 151 stripping the remainder in July 2026 [reported, multiple independent outlets plus the Chromium code review]**. uBlock Origin, with more than 40 million Chrome users, has no full-capability successor on the platform; its Lite variant cannot do cosmetic filtering or dynamic blocking. Firefox and Brave retain the API.

We are deliberately not alleging that Manifest V3's stated security rationale is pretextual; that question is unresolvable from outside and irrelevant to the structural point. The structural point is that the capability by which a user unilaterally refuses the advertising-funded web was removed from the dominant browser by the largest advertising company, and that no institutional mechanism required that decision to be justified to the users it affected. The same layer carries the same property in other cases: AMP, which conditioned preferential placement on serving Google-hosted copies; Privacy Sandbox, which deprecates third-party cookies in favour of a targeting substrate the browser vendor operates; and the Web Environment Integrity proposal of 2023, withdrawn after public objection, which would have let servers demand cryptographic attestation of an approved client — the most explicit statement available of what a tethered web looks like.

(d) It is becoming the agent. Chrome shipped auto browse on 28 January 2026, powered by Gemini 3, extending to Android in June; WebMCP entered origin trial in Chrome 149; and Google published the Universal Commerce Protocol for agent-executed purchases **[reported — vendor announcements]**. This is the vertical and horizontal axes closing on each other. An agent embedded in the browser inherits the authenticated session, the stored credentials, the browsing history and the payment rails; a third-party agent cannot obtain any of them on comparable terms. The default-placement advantage that cost26.3bn a year to purchase in the search era is reproduced one layer deeper, and this time there is no choice screen, because no regulator has yet defined what an agent default would even be.

6.3 The mobile operating system: the permission layer

Roughly two-thirds of searches occur on mobile. The mobile platform is a duopoly of approximately 73% Android and 27% iOS with no third option, and — unlike the desktop — it is a duopoly in which *installation itself is permissioned*. Three consequences bear on this paper.

Engine choice is nominal. iOS required WebKit for all browsers globally until the DMA. Article 5(7) prohibits a gatekeeper from requiring the use of its own browser engine; Apple's compliance deadline was March 2024 and it shipped BrowserEngineKit for the purpose. As of mid-2026, **no browser vendor has shipped an alternative engine on iOS [solid — Open Web Advocacy, corroborated by trade press]**. The binding obstacle is reported to be that a vendor must ship a *separate* EU-only application rather than updating its existing one, resetting its installed base to zero and requiring two engine implementations in perpetuity, which destroys the business case. Whatever one concludes about intent, the outcome is a natural experiment with an unambiguous result: a regulator ordered the most important vertical restriction in the browser market to be lifted, it was nominally lifted, and twenty-nine months later the count of alternative engines on the platform is zero.

Distribution is permissioned, and the permissions are substantive. For any application hosting user-generated content, App Store Guideline 1.2 requires a method for filtering objectionable material *from being posted*, a report mechanism with response within 24 hours, user blocking, and published contact information. For a permissionless publishing network this is not a compliance detail; it is a requirement that specific content can be made to disappear from the client on demand, which is a protocol-level capability or an absence of App Store presence. The commission and anti-steering regime is the better-litigated part of the same control surface, but the moderation mandate is the one that determines whether a decentralised design is shippable at all.

And the dominant computing mode cannot participate in a peer network. This is the most consequential fact in the section and the least contested. Apple's own developer guidance on running code continuously in the background, at a specified time, at a guaranteed interval, or resuming in response to an inbound network request is: *"The short answer to all of these is You can't."* `beginBackgroundTask` grants about three minutes; `BGAppRefreshTask` is heuristic and may grant nothing. The only push-driven wake requires a server that knows when to wake the device — which reintroduces exactly the central coordinator a peer-to-peer design exists to remove. Android is less hostile but Doze suspends networking and drops connections, and OEM background killers finish the job; the only reliable pattern is a foreground service with a permanent user-visible notification **[solid — platform documentation]**.

The implication is worth stating without hedging, because it constrains this project as much as it indicts the incumbents: **the majority of the world's computing cannot host, serve, relay, or replicate anything.** Peer-to-peer alternatives to platform hosting are therefore confined to the desktop minority by platform policy, not by protocol design or bandwidth. Every decentralisation proposal of the last fifteen years, including this one, has had to absorb that constraint, and most of the post-mortems in the literature name mobile absence as a proximate cause of death.

6.4 Long-form video: enclosure of a format the web could never afford

Video is the case where centralisation was not merely opportunistic but genuinely subsidised by physics and cost structure, which is exactly what makes it the most instructive of the three.

The open web never had an answer for video because the economics forbade it: a full nine-rung H.264 adaptive ladder weighs about 11.5 GB per hour of content against 2.6 GB for a practical four-rung AV1 ladder, and egress at that volume is unaffordable for an individual. Where text hosting costs approximately nothing, video hosting costs more than most creators will ever earn. A platform that absorbed the bandwidth in exchange for the audience and a revenue share was not predatory; it was the only offer available.

The result is that the single largest category of contemporary human explanatory output sits inside one firm. Nielsen's Media Distributor Gauge put YouTube at a platform-best **13.8% of all US television watch-time in May 2026** — the largest share of any distributor, broadcast or streaming, for a third consecutive month **[solid — Nielsen]**. It is not a short-form phenomenon: US creators uploaded **6.6 million videos of at least twenty minutes in Q1 2026, up 53.5% in three years**, and more than 70% of US YouTube viewers now watch on a connected television **[reported — Tubular Labs, EMARKETER]**. YouTube advertising alone was ***9.88bn in Q1 2026*** **[solid — SEC filing]**.

Three properties make this an enclosure of the commons rather than merely a large business.

1. **The content is addressable but not *citable*.** A twenty-minute video has no stable paragraph, no quotable extract, no anchor a reader can link to as evidence. Knowledge that migrates from a written tutorial to a video migrates out of the part of the corpus that can be excerpted, footnoted, corrected in place, or archived by anyone but the host.
2. **The migration is generational and measured.** The Wikimedia Foundation attributes part of its 8% decline in human pageviews to "younger generations seeking information on social video platforms rather than the open web" **[solid — first-party]**. This is the same substitution as §5.2, arriving by a different route, and the two compound.
3. **Ingestion is asymmetric.** A transcript is trivially extractable by the platform operator, who also operates the general-purpose model and the search engine, and is not extractable at scale by anyone else — both because access is contractual and because the crawl is expensive. The video corpus is therefore an input to the synthesis layer of §5 that only one participant can read. The same firm owns the format's hosting, its recommendation surface, its advertising market, and the model trained on it.

The honest addendum is that this is the layer where the counter-architecture in §9 has least to offer. Peer distribution works when many people watch the same thing at the same time on nearby networks; PeerTube's much-cited 98% offload figure comes from 500+ concurrent viewers of one video, and 92% of PeerTube content sits on a single instance. Long-tail video is the inverse case, and §9.4's design accordingly puts video in a paid, centralised cold tier and excludes it from v0. We are describing an enclosure we cannot currently undo.

6.5 The composite

Taken separately, each layer looks like a competitive outcome with a plausible efficiency story: Chrome is fast, iOS is secure, YouTube is free. Taken together they describe a stack in which the same firm sets the standard, ships the renderer, measures the user, sells the advertising, hosts the video, ranks the results, generates the answer, and now supplies the agent that reads on the user's behalf — with a second firm controlling the remaining quarter of the client population and paid roughly20bn a year not to disturb the arrangement.

For the argument of §7 the relevant consequence is narrow and specific. The three horizontal enclosures could each, in principle, have been escaped by a sufficiently determined user population. The vertical axis is what converted “in principle” into “not in practice,” and it did so by removing the client-side acts through which escape had historically occurred. That is why we treat the third enclosure as terminal rather than as one more turn of a cycle that previously self-corrected: the mechanism by which the previous turns *could* have self-corrected has itself been dismantled, layer by layer, over the same period.

---

7. Why the third enclosure is terminal

7.1 The formal statement

Recall $\rho_t = A_t / E_t$ and $F_{t+1} = f(A_t, M_t)$. Three regimes:

Regime I (ranking). The intermediary’s revenue is realised only if a user is routed to a page — the advertiser’s landing page or, indirectly, a publisher whose ad inventory Google also intermediates. ρ falls as the interface internalises answers, but a floor exists: an intermediary that returns nothing has nothing to sell. Provision is taxed but sustained.

Regime II (aggregation). Provision migrates inside the platform. A_t to the open web falls sharply (§4.3), but M_t and F_t are partly redirected rather than destroyed — the writing still happens, on someone else’s substrate. Damaging to the commons, not self-consuming: the platform’s own corpus grows.

Regime III (synthesis). The intermediary’s product is manufactured from C_t (the accumulated stock) and F_t (the fresh flow) and delivered without routing. Revenue is realised at the moment of the answer. Therefore:

$$\frac{\partial (\text{revenue})}{\partial A_t} \approx 0, \qquad \frac{\partial (\text{revenue})}{\partial F_t} > 0$$

The intermediary needs the flow and does not need to pay for it in the currency that produces the flow. Since f is increasing in A, F_{t+1} declines; since the model’s currency and correctness depend on F, quality declines with a lag equal to the corpus refresh interval. The lag is the whole problem: **it is long enough that no single actor experiences the depletion as a cost within its own planning horizon**, and this is a textbook provision failure, not a failure of intelligence or goodwill on anyone’s part.

Three properties make it terminal rather than cyclical:

1. **No internal correction.** The actor best positioned to notice depletion (the model developer) experiences it as a slow degradation in a distant metric, several quarters after the decisions that caused it, and shares the cost with every competitor.

2. **The obvious defence completes the enclosure.** §5.5: closure from below converts open access into a licensed market with incumbent-shaped barriers.
3. **The substrate has a synthetic substitute that is good enough to fool the measurement and not good enough to sustain the system.** ~50% of new articles are already primarily AI-generated; detection is imperfect; and recursive training degrades tails (Shumailov et al. 2024).
4. **The exits have been closed on the vertical axis.** This is the property that distinguishes the present situation from the previous two enclosures, both of which were in principle escapable by a user willing to change what they ran. Section 6 documented the closures: default search purchased at *26.3bn a year, full-capability content blocking removed from the dominant browser in June 2026, zero alternative engines on iOS twenty-nine months after a regulator ordered the restriction lifted, background execution forbidden on the dominant computing mode, and the agent now shipping inside the browser with privileged access to the session. A system can be cyclical only if the correction mechanism survives the cycle. Here the correction mechanism was itself the object of the preceding two decades of acquisition.*

7.2 Counterarguments, taken seriously

A thesis this convenient for its author deserves adversarial treatment. The six strongest objections:

(a) "Traffic is down but revenue is up; the metric is wrong." Sundar Pichai's public position is that AI filters out "low-quality clicks" and that bounce clicks are declining. Similarweb's data are partly supportive: ChatGPT-referred visitors averaged 15 minutes on site versus Google's 8, 12 pages per session versus 9, and converted at 7% versus 5% on transactional sites (US desktop, September 2025) **[reported]**. *Response:* the objection is valid for a specific class of site — transactional, branded, high-intent — and irrelevant for the class this paper is about. A 60% referral loss for publishers under 10,000 daily page views cannot be recovered by better-converting residual visitors, because such sites monetise volume or do not monetise at all. Higher-quality traffic at one-third the volume is a good outcome for a merchant and an extinction event for a long-tail commons contributor. The distributional finding — small −60%, medium −47%, large −22% — is precisely what the "quality not quantity" defence cannot accommodate.

(b) "SEO sludge was not worth saving." Largely true, and this paper's §3 argues that the sludge was the ranking function's own equilibrium output. *Response:* the objection concedes the diagnosis and misidentifies the target. The commons at risk is not the content farm; it is Wikipedia (−8%), Stack Exchange (−29×), personal technical writing, regional and specialist reporting, and the long-tail archive. Those are the sites with the fewest substitutes and the least ability to sign a licensing agreement.

(c) "Publishers will substitute direct channels: newsletters, apps, brand." This is the professional consensus advice, and for well-capitalised brands it works. *Response:* it is a strategy available to entities that already have an audience, and it is precisely unavailable to new entrants — which returns the system to the entrenchment effect of §3.1, now enforced at the level of the whole web rather than the results page. A commons whose only viable members are the already-famous is not a commons.

(d) "Non-market motivation will sustain provision; people wrote before advertising." Benkler's commons-based peer production is the strongest form of this argument, and it is not empty: Wikipedia runs on roughly 30–40k active editors, not on payments. *Response:* the empirical evidence is currently against it in the specific case. Wikipedia's own reporting attributes the decline in human pageviews to AI intermediation and warns about the

downstream effects on editor and donor recruitment; fewer readers is the leading indicator for fewer editors. Intrinsic motivation is not independent of audience — most people write to be read, andM_t*is itself a function of*A_t*with a lag.*

(e) "This is normal creative destruction; the web replaced newspapers too." *Response:* the analogy fails on the reciprocity test. The web substituted for newspapers while creating a new substrate with its own provision mechanism. The synthesis layer creates no substrate; it is a pure consumer of the substrate it displaces. That is the distinguishing feature, and it is checkable: name the mechanism by which generative search causes new original documents to be written. To date the only candidates are licensing agreements with large incumbents and the hope that brand exposure inside an answer converts downstream — the first is available to perhaps a few hundred organisations worldwide, and the second is measured at a 1% click rate.

(f) "If the product is this bad, build a better one; the market will do the rest." The strongest objection in principle, because it is the one the paper's constructive section must answer about itself. *Response:* the market test has been run repeatedly and the results are in §6. Neeva built a subscription search engine with no advertising incentive, staffed by former Google search leadership, and shut down in 2023 reporting that the hard part was not building the index or even matching relevance but "convincing regular users of the need to switch"; it was acquired for its infrastructure rather than its product. Blekko sold to IBM. Cuil closed in two years. Mozilla, having produced a browser many reviewers considered superior, went from roughly a third of the market to low single digits. None of these is a story about a product losing on merit to a better product. They are stories about distribution, defaults, and pre-installation, which is to say they are stories about the vertical axis. Two clarifications follow, and they cut in opposite directions. Against the objection: entry at the discovery layer alone is not a live option, which is why this paper argues that the counter-move must be made at the client layer (§9.3). Against this paper: the same evidence predicts that a new client faces distribution conditions strictly worse than the ones that defeated Neeva, and we do not have a distribution answer. We think the honest position is that the architectural argument can be right and the venture can still fail for reasons this paper's analysis implies rather than overlooks.

7.3 What would falsify the thesis

To keep the argument honest, we state in advance the observations that would count against it:

1. Zero-click rates plateau or reverse for two consecutive years while assistant usage grows.
2. Crawl-to-refer ratios for assistant-purpose crawlers converge towards search-engine levels (single or low double digits) and stay there.
3. Small-publisher referral losses stop exceeding large-publisher losses, indicating the effect is a reallocation rather than a concentration.
4. Independent measurement shows the flow of new, original, human-authored, openly addressable documents stable or rising through 2027–2028.
5. Wikipedia's human pageview decline reverses without a change in measurement methodology.
6. A discovery entrant without a client of its own reaches durable double-digit query share, or a browser built on a genuinely independent engine reaches double-digit share, either of which would indicate that the vertical axis is contestable on ordinary competitive terms and that §6 overstates its role.

None of these is implausible. Item 2 in particular is directionally happening for OpenAI. We would regard sustained (1) plus (4) as decisive against the paper's central claim, and (6) as decisive against its structural claim while leaving the descriptive one intact.

---

8. The diagnosis: a ranking-function failure

8.1 The constraint that produced the proxy is gone

The through-line of §§3–5 is that a single 1998 engineering constraint — machines cannot read — propagated into a ranking proxy, then into a market in that proxy, then into a content economy shaped by that market, and finally into an answer layer that consumes the content economy's output while starving it. The pathologies are not independent; they are the same defect observed at three resolutions.

That constraint no longer binds. A language model can read a page and assess depth, originality, substantiation and intent for a small fraction of a cent, at publish time, before any link exists. The empirical case that this is a *better* signal than the link graph, not merely a cheaper one, rests on three results:

1. **DCLM (Li et al. 2024)**: top-20%-by-PageRank was the worst filter tested, below the unfiltered baseline (26.1 vs 27.5 CORE); a learned classifier scored 30.2 **[solid]**.
2. **Ask-LLM (Sachdeva et al. 2024)**: models trained on LLM-selected data *beat full-data training after rejecting 90% of the corpus*, converged up to 70% faster, and — decisively — Ask-LLM scores had near-zero correlation with same-capacity perplexity, meaning model judgment carries information that no cheap statistical proxy captures **[solid]**.
3. **FineWeb-Edu (Penedo et al. 2024)**: an LLM-scored, distilled classifier filtered 92% of a 15T token corpus and matched larger corpora's MMLU with 10× fewer tokens **[solid]**.

8.2 And the same evidence base contains the strongest warnings

It would be dishonest to cite the above without the countervailing findings from the identical literature, which the design in §9 must survive:

- **The judge is validated as a floor-remover, not a ceiling-finder.** FineWeb-Edu's distilled classifier achieves overall accuracy 0.71 but **F1 = 0.02 on its top class** (precision 0.33, recall 0.01) **[solid]**. Every published cheap-at-scale filter is evaluated as a bottom-decile remover. Identifying the best content is a different task and the literature does not do it.
- **No one has validated an LLM judge on this task against humans.** Reliability figures come from chatbot-answer evaluation (Zheng et al. 2023); web-scale filters are validated by downstream benchmark scores. On objectively verifiable tasks, JudgeBench measured GPT-4o at 50.9% against a 50% random baseline **[solid]**.
- **Quality filters encode a language ideology.** Dodge et al. (2021) found C4's blocklist removed **42% of African-American-English documents and 32% of Hispanic-aligned English against 6.2% of White-aligned English**; Gururangan et al. (2022) found a reimplemented GPT-3 quality filter preferred writing from larger schools in wealthier, more educated, more urban ZIP codes, and that its notion of quality was *unaligned*

with factuality and with literary acclaim; Hofmann et al. (*Nature*, 2024) found models' covert stereotypes about AAE speakers more negative than any human stereotype ever experimentally recorded, with human-feedback training widening rather than closing the covert/overt gap **[all solid]**.

- **Judges are attackable.** JudgeDeceiver achieved 90–99% attack success with optimised ~20-token suffixes, with positional attack consistency of 83–98% (so order-swapping is not a defence) **[solid]**; a *null model* emitting one constant string achieved 86.5% length-controlled win rate on AlpacaEval 2.0 **[solid]**; and GEO (Aggarwal et al., KDD 2024) showed visibility gains of 30–41% from adding citations, statistics and quotations, with rank-5 pages gaining **+115.1% [solid]**.

The honest summary is that the diagnosis in §8.1 is well supported and the remedy in §9 is a research programme with a large, clearly identified risk at its centre.

---

9. A counter-architecture

The remainder of the paper specifies the system this project is building, presented as a set of hypotheses with mechanisms rather than as a product. The full engineering treatment is in `design.md`; the evidence base is in `research.md`.

9.1 Four inversions

#	Inversion	From	To
1	Discovery	Quality inferred from accumulated links (trailing)	Quality assessed on arrival (leading); engagement confirms or revokes
2	Hosting	Publisher pays for their own distribution	Publishing obligates hosting; the long tail becomes the head's CDN
3	Cold start	Empty network on install	Ships with a pre-distilled, quality-filtered index of the existing web
4	Client	Discovery layer rented from a browser someone else owns	One binary: node, publisher, and reader. Installing *is* joining

Each is load-bearing for the others: (1) without (3) has nothing to rank on day one; (2) without (4) asks users to run a daemon, which is the friction that killed every predecessor; (3) without (1) is merely another crawl with no consistent quality bar. Inversion 4 was originally conceived as a packaging convenience and §6 is the argument that it is nothing of the sort; §9.3 restates it.

9.2 The ranking pipeline: the judge buys a hearing, not a verdict

The mechanism that answers §8.2's warnings is the refusal to let the model's score *be* the ranking.

- **Stage 0 — cheap gate.** A frozen-encoder classifier with a linear head (FineWeb-Edu architecture) or a fastText classifier. Chosen on two grounds: fastText was the most accurate filter in DCLM's bake-off, and a frozen-encoder regressor **has no instruction-following behaviour to hijack** (InjecAgent measured fine-tuned models at 3.8% attack success versus 23.6% for prompted ones). Approximately1,000 per billion pages self-hosted. Its job is to remove the floor.

- **Stage 1 — jury.** Dimension scores (depth, originality, substantiation, density, craft, intent) on an additive, reference-anchored rubric; a panel rather than a single model; swap-and-average for order bias; retrieval-grounded originality measured against the corpus rather than asserted by the model; and — following CaMeL, the only prompt-injection defence with a security guarantee — **page content is treated as data, never as instruction**, with the judge scoring extracted structured features. Whether depth and originality survive structured feature extraction is an open empirical question with no literature behind it, and it is Hypothesis 4 in Appendix B.
- **Stage 2 — impression budget.** The prior buys a Thompson-sampled *impression budget*, not a rank. This converts "is the judge right?" from an assertion into an experiment.
- **Stage 3 — validation.** Three non-negotiable mechanisms, each traceable to a measured result:
- **A permanent random no-boost holdout.** Salganik, Dodds & Watts (2006) showed the same song ranking #1 in one world and #40 of 48 in another under social influence, and Salganik & Watts (2008) showed an inverted ranking to be substantially self-fulfilling. An initial model boost *is* that manipulation. The holdout is the independent condition, and without it judge quality is unfalsifiable.
- **Logged propensity on every impression**, so engagement can be corrected by inverse propensity scoring (Joachims et al. 2017). Applied deterministically and unlogged, the engagement data is permanently confounded, and Chaney et al. (2018) show what happens next.
- **Bayesian shrinkage rather than a multiplicative boost**: `posterior = (C · m_judge + Σ engagement_ips) / (C + n)`. The half-life of AI influence is controlled by the single auditable pseudocount `C`.
- **Squash everything.** Hacker News's `^0.8`, Reddit's `log10`, and the `unsquashedClicks` / `unsquashedImpressions` fields in Google's own leaked schema all encode the same lesson: any signal proportional to accumulated attention, fed back into exposure, generates a power law regardless of quality. **PageRank is a preferential-attachment amplifier**; its absence is the tell of an immature ranking system, and it costs one exponent.

Signals are chosen by cost-to-fake, on the principle that **signal robustness is proportional to the scarce, non-replicable resource the signal consumes**: completion depth, dwell relative to length, return visits after ≥24h, explicit saves, and the last-longest-click analogue, each weighted by the engager's verified hosting contribution. Raw clicks, raw views, and share counts are refused. Bot-farm click services advertise *0.50*–5 per 1,000 clicks and residential-proxy economics put a faked engagement event at *0.001*–0.01, so any mechanism whose influence scales linearly with volume is defeated by a four-figure budget.

9.3 The client: why the counter-move has to happen at the same layer

Section 6 argued that the enclosures were enforced at the client. The symmetric conclusion is that a discovery layer which does not ship a client is not contesting the enclosure; it is applying to be ranked inside it. Three of the paper's findings converge on this and each has a concrete design consequence.

The signals require a client. §6.2(b) established that the incumbent's durable advantage is an instrumented population, not an algorithm. Completion depth, dwell relative to length, return visits after twenty-four hours, and explicit saves — the entire signal set of §9.2 — are observable only by the program that renders the document. A

web-hosted search engine can observe a click and nothing after it, which is precisely why the ranking systems of the last twenty years have been built on links and clicks: they are the only things visible from outside the client. The signal design is not independently choosable from the client decision; it *is* the client decision.

The obligations require a client. Inversion 2 makes publishing conditional on hosting. A web page cannot host anything, so a browser tab cannot be a peer. Every prior attempt to bolt peer participation onto the browser — a daemon plus a local gateway, a WebRTC-mediated partial peer, a companion extension — reproduces the installation friction that appears in nearly every post-mortem in §4.4. One binary that is simultaneously reader, publisher, and node is the only packaging in which the reciprocity obligation is enforceable without asking the user to run infrastructure.

And the client cannot be a web browser. This is the uncomfortable half. §6.2(a) established that implementing the web platform is a multi-decade programme even for a well-funded non-profit — Ladybird reaches *alpha* in 2026 having started in earnest in 2022, with stable targeted for 2028 — and that the specification is written substantially by the incumbent. A project that renders HTML inherits an attack surface it cannot audit, a performance target it cannot meet, a compatibility treadmill set by a competitor, and, because the format admits arbitrary script, a page that can exfiltrate a peer's identity or conscript its bandwidth. The design therefore does not render HTML. It defines a constrained document format — structured text, images, deterministic layout, no arbitrary execution — for which a conforming renderer is a bounded piece of work and for which a quality judge can extract features reliably, because a document that cannot compute cannot disguise itself from the extractor in §9.2.

We state the costs of that choice plainly, because they are severe and they are not fully mitigated. A non-HTML client cannot display the existing web, so the network begins with a hard wall between itself and every page anyone already reads; §9.5's distilled corpus narrows that wall but does not remove it. Interactive content — applications, visualisations, anything computed — is out of scope by construction, which forecloses categories of legitimate publishing. Long-form video is out of scope for the separate economic reasons in §6.4. And a second document format is a proposal for a standard, which is a class of proposal with a poor historical record. What the choice buys is the only thing that makes the attempt tractable at all: an implementable client. The alternative is not a better browser. The alternative, on the evidence of §6.2(a), is no client, and on the evidence of §6.1 that means no exit.

What this does not solve. The mobile constraint of §6.3 is unaffected by any of the above. A mobile build can read and can participate as a client, but it cannot host, cannot serve, and cannot be relied upon to be reachable, because the platform forbids it rather than because the protocol lacks the capability. The hosting obligation of inversion 2 is therefore satisfiable only on desktop, which means the network's supply side is structurally confined to a minority of computing, and the App Store moderation requirements of §6.3 apply to any client that ships at all. We do not have an answer to this. It is the deepest external constraint on the design, and Appendix B states it as a condition rather than a hypothesis, because no experiment we can run will change it.

9.4 Pay-forward hosting, and the honest arithmetic

To publish B bytes you pledge $k{\cdot}B$ bytes of hosting, **scoped to what you follow and to demand-weighted assignment within that scope** — a liability decision before it is a performance one, since opportunistic caching of network-assigned arbitrary bytes is the pattern that crippled every predecessor. Contribution is audited by random challenge-response with a Storj-style disqualification threshold and a vesting bond rather than up-front capital.

The naive surplus claim — publishers vastly outnumber quality producers, so the network runs a structural hosting surplus — is **half true and we should say which half**. The participation half is extreme and real (a 2005 Gnutella replication measured 85% of peers sharing nothing). The service half largely fails: demand is Zipfian while volunteer supply is uniform; US median residential upload is 19.34 Mbps against 291 Mbps down; churn is Weibull with a bimodal population; mobile devices cannot host at all. Multiplying through, a nominal 100:1 becomes closer to **1:1–5:1**. Every published peer-offload figure above 75% comes from many people watching the same content at the same time on a nearby network — the inverse of long-tail publishing — so the origin remains the primary path and peers are the optimisation.

9.5 The corpus: derived representations, not expression

The network ships with a distilled index of the pre-existing web, scored by the same judge, so that native and legacy content are comparable in one result set. It ships **embeddings, quality scores, term statistics, MinHash signatures, link graphs, short snippets and links** — with full text fetched client-side from origin at query time, which preserves the publisher's analytics, advertising and paywall. This is the *Kelly* / *Perfect 10* / *HathiTrust* / *Google Books* fact pattern rather than the *Hachette v. Internet Archive* one, and it happens also to be the cheapest architecture: at 10M pages, a binary-quantised index plus a compressed snippet store is **under 15 GB**, against roughly 20× that for full text.

Crawl policy is deliberately self-binding: stay logged out; never circumvent authentication, paywalls or anti-bot measures; honour robots.txt, `noarchive`, `noindex`, the IETF AIPREF vocabulary and Cloudflare Content Signals; declare which AIPREF category the pipeline claims; and adopt Brave's rule that *if a page is not crawlable by Googlebot, we do not crawl it either*. Revocation — signed tombstones, peer-enforced expiry, a replicated suppression list consulted at query time — is a v1 protocol feature, because it cannot be retrofitted onto immutable content-addressed storage.

Note the tension this paper is obliged to name: §5.5 argues that closure of the crawl surface entrenches incumbents, and this section commits to honouring that closure. We accept the cost. A project whose thesis is that the discovery layer must be legitimate cannot begin by taking what publishers have declined to give.

9.6 The editorial charter

Every ranking system has an editorial position; ours is published, versioned, and contestable. The network optimises for content that took effort to make and rewards effort to read; firsthand over secondhand; claims that carry their evidence; serving the reader over capturing the reader; and saying something. It explicitly refuses to optimise for agreement with the operator's politics, institutional prestige as a proxy for correctness, formality of register, or English-native prose style. Following Dodge et al.'s practice of releasing multiple filtering levels plus a searchable index, **both the boosted and the unboosted rankings are published**, so anyone can measure what the judge suppresses; per-language and per-register acceptance rates are monitored and published; and appeals follow a rebuttable-presumption process in the Wikipedia-notability mould.

There is no token. The regulatory path exists, but the mechanism-design argument is decisive: ranking depends on engagement being an honest signal, and a token makes engagement financially valuable to produce. Steemit converged to stake-weighted rent extraction; Filecoin rewarded proven capacity and got proven capacity; Presearch

closed on 27 July 2026 citing “an unsustainable token model” and a conversation that “centered almost exclusively on token price rather than the engine.” The hosting- contribution unit is therefore non-transferable, denominated in service units, and structurally incapable of being earned by consuming content.

9.7 What is centralised, stated honestly

A design that claims full decentralisation on day one is either lying or dead. Brave Search — the one working analogue to this ambition — is distributed collection, centralised index, distributed ranking, and on that architecture went from 14.1m to 50m+ queries per day, while YaCy’s fully distributed global network serves **1.75 queries per second worldwide**. Accordingly, v0 centralises judge inference, the score registry, content routing, corpus construction, the denylist, and relays, each with a stated decentralisation path. The claim is not that centralisation is good; it is that the alternative to admitting it is a network that is empty and unusable while it waits for a decentralisation it never achieves.

---

10. Known hazards of the counter-architecture

Stated as risks, not as an afterword, because a proposal that only lists its advantages is advertising.

1. **The judge may not correlate with human quality judgment on web pages.** Unmeasured in the literature. Hypothesis 1, Appendix B.
2. **The judge may only remove the floor.** FineWeb-Edu’s top-class F1 of 0.02 is the sharpest empirical obstacle in the entire evidence base.
3. **Boost-now-confirm-later is exactly the shape a content flood exploits.** GEO’s recipe — citations, statistics, quotations — costs nothing to synthesise, is worth +30–41% visibility, and benefits low-ranked pages most (+115.1% at rank 5). Graphite’s plateau hypothesis says the current brake on mass synthetic publication is the absence of an immediate quality reward. We propose to add one.
4. **The judge encodes a taste, and tastes have measured demographic structure** (§8.2). Mitigations are in the design; they are mitigations, not solutions.
5. **Hosting-by-default is a liability surface.** Follow-scoped replication, host-blind storage, denylist enforcement covering routing rather than only retrieval, and jurisdiction-aware operator opt-out are the answers; the tension between fixing the economics of altruistic hosting and forcing publishers to store content they did not choose is the design’s central structural problem.
6. **Default inertia beats install friction.** Neeva’s post-mortem is unambiguous. A single binary reduces install friction and does nothing about defaults. None of the four inversions solves distribution, and §6 is an argument that distribution is the binding constraint — which makes this the hazard the paper is least equipped to answer and most obliged to name.
7. **Relevance is the product.** Cuil launched claiming 121 billion pages and died in two years because the ranking did not work.
8. **The corpus solves content cold start, not author cold start.** Why does a good writer publish here first? That is a product question and the architecture does not answer it.

9. **A separate document format is a wall as well as a defence.** §9.3 argues that rendering HTML is not tractable and that a constrained format is what makes a client implementable and a judge reliable. The cost is that the network cannot display the web, which is a severe product liability and a standing invitation to build a bridge that reintroduces everything the format excluded.
10. **The supply side is confined to desktop by platform policy.** Hosting, serving, and relaying are forbidden on mobile (§6.3). The network's participation ceiling is therefore set by a third party, it is not a ceiling any protocol decision can raise, and it has been a proximate cause of death for most comparable projects.
11. **The client is subject to the App Store moderation regime.** Guideline 1.2 requires pre-posting filtering, 24-hour response, and user blocking for user-generated content. Satisfying it requires takedown capability at the protocol level; refusing it forfeits roughly a quarter of the client population and the majority of its spending power. There is no third option, and the choice determines how decentralised the design can honestly claim to be.

---

11. Threats to validity

Panel discontinuity. The zero-click series in §3.4 stitches together Jumpshot (defunct), Datos, and Similarweb panels with different devices, demographics and definitions. SparkToro says so explicitly. The 2024→2026 comparison is the most defensible; the decade-long "45% → 68%" framing is directionally sound and quantitatively soft.

Vendor telemetry and selection effects. Cloudflare Radar sees roughly 20% of the web, skewed by who buys Cloudflare. Chartbeat's panel is news and media, which is the worst-affected vertical and not representative of the web. Similarweb's clickstream panels underrepresent in-app search. Crawl-to-refer figures vary by more than 3× between compilations of the same underlying source in the same month, which is why §5.3 reports ranges.

Causal identification. No AI Overview rollout was randomised. Confounds include successive core and spam updates, the "Helpful Content" system, the shift of discovery to social video, and a secular pre-2022 decline that is clearly visible in Stack Overflow's data from June 2020. This paper therefore argues from convergence across independent measurement regimes — behavioural panel, clickstream, publisher analytics, first-party pageview logs, infrastructure telemetry — rather than from any single identified effect.

Detector reliability. The claim that ~50% of new articles are primarily AI-generated depends on classifiers validated at under 2% error on curated sets, applied to a distribution that is adversarial to them. Treat as an order of magnitude, not a point estimate.

Definitional. "Open web," "original," and "human-authored" have no agreed operationalisation, and this is not a small problem for a paper that argues about a quantity's decline. §7.3 item 4 is stated as the measurement we would need and do not have.

Normative. The paper assumes the open web is worth preserving. That is an argument from pluralism, auditability, and the absence of a single point of editorial control — not a demonstrated welfare claim. A reader who believes a high-quality synthesis layer serves users better than a long tail of variable pages will reject the conclusion while accepting the evidence, and that position is coherent.

Author’s interest. This paper is written by the party building the proposed alternative. The diagnosis and the remedy were not developed independently. Sections 6.3 and Appendix B exist so that the diagnosis can be evaluated against evidence that would embarrass the remedy.

---

12. Conclusion

The pathologies of the contemporary web are usually narrated as a morality play with a small cast. The better account is dryer and more useful: a proxy adopted under a 1998 engineering constraint became a market; the market shaped the content; the platforms then internalised authorship and withdrew the referral rent; and the synthesis layer has now removed the last structural reason for any intermediary to route a human to a page at all. At each step the actor’s decision was locally rational, and at each step the reciprocity ratio fell. It is now, for the largest AI crawlers, somewhere between two and four orders of magnitude below the search-engine bargain it replaced, while the revenue of the layer performing the extraction grew nineteen percent year over year.

Running underneath that sequence is the reason it was never corrected. Each enclosure was, in principle, escapable by users willing to change what they ran, and over the same two decades the capacity to change what one runs was acquired layer by layer: the default search slot bought for *26.3bn a year, the dominant browser engine reduced to a category of one, the ability to refuse advertising removed from that browser in June 2026, alternative engines still absent from iOS twenty-nine months after a regulator ordered the restriction lifted, background execution forbidden outright on the computing mode most people now use, the largest body of contemporary explanatory video held inside the same firm that ranks, summarises, and monetises it, and — most recently — the agent that reads on the user's behalf shipped inside the browser with privileged access to the session. None of these individually looks like an enclosure of the commons. Together they are the reason the commons had no exit.*

What makes this the terminal stage is therefore twofold. It is the first stage that consumes the substrate it depends on and experiences no cost for doing so within any actor's planning horizon; and it arrives at a moment when the client-side acts through which users historically escaped a bad discovery layer have been withdrawn one at a time. The supply side's rational defence — closing the crawl surface, moving to licensing — does not restore the commons. It converts an open-access resource into a licensed input market whose barriers are highest for exactly the entrants that would discipline the incumbents.

The single most interesting fact about this situation is that the technology completing the enclosure is also the first technology capable of removing the constraint that started it. Machines can now read. The evidence that link-based authority is a poor quality signal and that model judgment is a better one is stronger than anything available to the field a decade ago, and it comes from controlled comparisons rather than from advocacy. That does not make a model-judged discovery network a good idea by default; the same literature contains severe results on judge bias, judge instability, and adversarial manipulation, and the central mechanism has never been validated against human raters on this task. It makes it the experiment worth running, with the holdout, the published unboosted ranking, and the stated kill criteria that let it fail visibly.

It also determines where the experiment has to be run. If the enclosure was enforced at the client, then a discovery layer that ships as a web page is asking the enclosure's owner for permission to compete, and the record of entrants who did exactly that is uniformly bad — not because their ranking was worse, but because ranking was never the layer that decided. The counter-move has to be a client: one program that reads, publishes, and hosts, measuring the

only signals that have survived adversarial pressure because it is the program rendering the document, and declining to render HTML because a web engine is a multi-decade undertaking whose specification is written by the incumbent. That choice buys tractability and costs compatibility, and we have tried to state the cost as plainly as the case. On the mobile question we have no answer at all: the platform, not the protocol, forbids participation.

The web was not a technology. It was an arrangement in which finding something and rewarding its author were the same act, conducted through a program that belonged, in a meaningful sense, to the person running it. Both halves of that arrangement are being dismantled, the first by a layer that needs what it is dismantling and the second by the layer that decides what you are permitted to run. Rebuilding it requires replacing the proxy at the centre of discovery and rebuilding the client underneath it — and the only remaining question worth arguing about is whether the replacement can be made legitimate, and distributed at all, before the substrate it must rank is gone.

References

Adar, E. & Huberman, B. (2000). Free riding on Gnutella. *First Monday*, 5(10).

Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K. & Deshpande, A. (2024). GEO: Generative Engine Optimization. *KDD 2024*. <https://arxiv.org/abs/2311.09735>

Apple Inc. (2026). *App Store Review Guidelines*, §1.2 (User-Generated Content); *Background Tasks* and *Energy Efficiency Guide for iOS Apps*, Apple Developer Documentation. <https://developer.apple.com/app-store/review/guidelines/>

Apple Inc. (2026). Using alternative browser engines in the European Union. Apple Developer Support. <https://developer.apple.com/support/alternative-browser-engines>

Alphabet Inc. (2026). *Q1 2026 Earnings Release*, Exhibit 99.1. U.S. Securities and Exchange Commission. <https://www.sec.gov/Archives/edgar/data/1652044/000165204426000043/googexhibit991q12026.htm>

Axios (2026). Publishers are losing Google traffic as AI answers replace links, 31 July 2026. <https://www.axios.com/2026/07/31/google-search-publishers-seo-geo-llms-ai>

Barabási, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509–512. <https://arxiv.org/abs/cond-mat/9910332>

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.

Boyle, J. (2003). The second enclosure movement and the construction of the public domain. *Law and Contemporary Problems*, 66, 33–74.

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117. <http://infolab.stanford.edu/~backrub/google.html>

Brave (2025). Brave continues to support Manifest V2 extensions. Brave Blog.

Business Insider (2026). Anthropic, OpenAI crawl the web even more, and give little back (January 2026); Anthropic treats the web like an all-you-can-eat buffet (July 2026).

Casper, S. et al. (2024). Open problems and fundamental limitations of reinforcement learning from human feedback. *TMLR*. <https://openreview.net/forum?id=bx24KpJ4Eb>

Chaney, A. J. B., Stewart, B. M. & Engelhardt, B. E. (2018). How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. *RecSys 2018*. <https://arxiv.org/abs/1710.11214>

Chapelle, O. & Li, L. (2011). An empirical evaluation of Thompson sampling. *NIPS 2011*.

Chartbeat / Similarweb, via Press Gazette (2024–2025). Facebook referral traffic for publishers down 50% in 12 months; Publisher traffic sources 2019–2025 analysed. <https://pressgazette.co.uk/media-audience-and-business-data/media$_{\text{metrics}}$/facebooks-referral-traffic-for-publishers-down-50-in-12-months/>

Cho, J. & Roy, S. (2004). Impact of search engines on page popularity. *WWW 2004*.

Cho, J., Roy, S. & Adams, R. E. (2005). Page quality: in search of an unbiased web ranking. *SIGMOD 2005*.

Chromium Project (2026). Removal of `ExtensionManifestV2Disabled` (Chrome 150) and remaining Manifest V2 flags (Chrome 151); W3C WebExtensions Community Group notes, 20 May 2026. Reported by heise online, The Next Web, gHacks and 9to5Google, June 2026.

Cloudflare (2025). Content Independence Day: no AI crawl without compensation; Introducing pay per crawl. 1 July 2025. <https://blog.cloudflare.com/introducing-pay-per-crawl/>

Digital Markets Act, Regulation (EU) 2022/1925, Article 5(7) and Recital 43.

Doctorow, C. (2023). *Pluralistic*: The 'enshittification' of TikTok. (Extended treatment: Doctorow, C. (2025). *Enshittification*. Farrar, Straus and Giroux.)

Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., Mitchell, M. & Gardner, M. (2021). Documenting large webtext corpora: a case study on the Colossal Clean Crawled Corpus. *EMNLP 2021*. <https://aclanthology.org/2021.emnlp-main.98/>

Fortunato, S., Flammini, A., Menczer, F. & Vespignani, A. (2006). Topical interests and the mitigation of search engine bias. *PNAS*, 103(34), 12684–12689.

EMARKETER (2026). Long-form YouTube is becoming a serious TV-equivalent media buy; FAQ on YouTube advertising. Citing Tubular Labs data via TVREV.

Gillespie, T. (2018). *Custodians of the Internet*. Yale University Press.

Graphite (2026). AI now writes as many online articles as humans do. <https://graphite.io/five-percent/ai-now-writes-as-many-online-articles-as-humans-do>

Gururangan, S., Card, D., Dreier, S. K., Gade, E. K., Wang, L. Z., Wang, Z., Zettlemoyer, L. & Smith, N. A. (2022). Whose language counts as high quality? Measuring language ideologies in text data selection. *EMNLP 2022*. <https://arxiv.org/abs/2201.10474>

Hofmann, V., Kalluri, P. R., Jurafsky, D. & King, S. (2024). AI generates covertly racist decisions about people based on their dialect. *Nature*, 633, 147–154. DOI 10.1038/s41586-024-07856-5

Ladybird Browser Initiative (2026). Project FAQ and roadmap (alpha 2026, beta 2027, stable 2028); *This Month in Ladybird*, January 2026. <https://ladybird.org/>

Joachims, T., Granka, L., Pan, B., Hembrooke, H. & Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. *SIGIR 2005*.

Joachims, T., Swaminathan, A. & Schnabel, T. (2017). Unbiased learning-to-rank with biased feedback. *WSDM 2017* (Best Paper). <https://arxiv.org/abs/1608.04468>

Li, J. et al. (2024). DataComp-LM: In search of the next generation of training sets for language models. *NeurIPS 2024 Datasets & Benchmarks*. <https://arxiv.org/abs/2406.11794>

Longpre, S. et al. (2024). Consent in crisis: the rapid decline of the AI data commons. *NeurIPS 2024 Datasets & Benchmarks*. <https://www.dataprovenance.org/Consent$_{\text{inCrisis}}$.pdf>

Malik, S. et al. (2025). RewardBench 2. <https://arxiv.org/abs/2506.01937>

Nestaas, F., Debenedetti, E. & Tramèr, F. (2025). Adversarial search engine optimization for large language models. *ICLR 2025*. <https://arxiv.org/abs/2406.18382>

Nielsen (2026). *The Gauge* and *Media Distributor Gauge*, May 2026 reports. Published 28 July 2026. <https://www.nielsen.com/news-center/2026/streaming-embarks-on-annual-summer-ascent-in-nielsens-may-2026-gauge-reports/>

Open Web Advocacy (2025). Apple's browser engine ban persists, even under the DMA. <https://open-web-advocacy.org/blog/apples-browser-engine-ban-persists-even-under-the-dma/>

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge University Press.

Page, L., Brin, S., Motwani, R. & Winograd, T. (1999). The PageRank citation ranking: bringing order to the web. Stanford InfoLab Technical Report.

Penedo, G., Kydlíček, H., allal, L. B., Lozhkov, A., Mitchell, M., Raffel, C., von Werra, L. & Wolf, T. (2024). The FineWeb datasets: decanting the web for the finest text data at scale. *NeurIPS 2024 Datasets & Benchmarks*.

Pew Research Center (2025). Google users are less likely to click on links when an AI summary appears in the results. 22 July 2025. <https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/>

Sachdeva, N., Coleman, B., Kang, W.-C., Ni, J., Hong, L., Chi, E. H., Caverlee, J., McAuley, J. & Cheng, D. Z. (2024). How to train data-efficient LLMs (Ask-LLM). <https://arxiv.org/abs/2402.09668>

Ramakrishnan, S. (2023). *Next steps for Neeva*. Neeva shutdown announcement, 20 May 2023.

Salganik, M. J., Dodds, P. S. & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762), 854–856.

Salganik, M. J. & Watts, D. J. (2008). Leading the herd astray: an experimental study of self-fulfilling prophecies in an artificial cultural market. *Social Psychology Quarterly*, 71(4), 338–355.

Shi, J., Yuan, Z., Liu, Y., Huang, Y., Zhou, P., Sun, L. & Gong, N. Z. (2024). Optimization-based prompt injection attack to LLM-as-a-judge (JudgeDeceiver). *ACM CCS 2024*. <https://arxiv.org/abs/2403.17710>

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R. & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631, 755–759.

SparkToro / Similarweb (2026). In 2026, less than one third of Google searches still send a click. 9 June 2026. <https://sparktoro.com/blog/in-2026-less-than-one-third-of-google-searches-still-send-a-click/>

Srnicek, N. (2017). *Platform Capitalism*. Polity.

Stack Exchange Meta (2026). The number of questions has decreased 30-fold. Updated 5 April 2026. <https://meta.stackexchange.com/questions/408728/>

Tan, S. et al. (2025). JudgeBench: a benchmark for evaluating LLM-based judges. *ICLR 2025*. <https://arxiv.org/abs/2410.12784>

United States v. Google LLC, No. 1:20-cv-03010 (D.D.C.). Memorandum Opinion, 5 August 2024; Remedies Opinion, 2 September 2025.

United States v. Google LLC, No. 1:23-cv-00108 (E.D. Va.). Opinion, 17 April 2025 (ad tech).

The Verge (2025). Where are the iPhone's WebKit-less browsers? <https://www.theverge.com/news/706569/apple-ios-iphone-alternative-browser-webkit-owa>

van Dijck, J., Poell, T. & de Waal, M. (2018). *The Platform Society*. Oxford University Press.

Verga, P. et al. (2024). Replacing judges with juries. <https://arxiv.org/abs/2404.18796>

Wikimedia Foundation (2025). New user trends on Wikipedia. 17 October 2025. <https://diff.wikimedia.org/2025/10/17/new-user-trends-on-wikipedia/>

W3C / WICG (2023). Web Environment Integrity explainer; proposal withdrawn November 2023.

Wu, T. (2016). *The Attention Merchants*. Knopf.

Zheng, L. et al. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *NeurIPS 2023*. <https://arxiv.org/abs/2306.05685>

Zittrain, J. (2008). *The Future of the Internet — And How to Stop It*. Yale University Press.

Appendix A — Claim ledger

Key quantitative claims, with date, source type and grade. Figures move; the date is part of the claim.

#	Claim	Value	As of	Source	Grade
1	US Google searches ending without a click	68.01%	Jan–Apr 2026	SparkToro / Similarweb panel	reported
2	Same, prior measurement	60.45%	2024	SparkToro / Datos	reported
3	Click rate with vs without AI summary	8% vs 15%	Mar 2025	Pew Research (n=900, 68,879 searches)	solid
4	Clicks on links inside AI summaries	1%	Mar 2025	Pew Research	solid
5	Sessions ending after an AI summary	26% vs 16%	Mar 2025	Pew Research	solid
6	CTR reduction for top result when AIO present	−58%	Dec 2025	Ahrefs	reported
7	Google search traffic to publishers, YoY	−34%	Jul 2026	Chartbeat via Axios	reported
8	Two-year referral loss by publisher size	−60% / −47% / −22%	2024–2026	Chartbeat via Axios	reported
9	Wikipedia human pageviews, YoY	−8%	Oct 2025	Wikimedia Foundation (first-party)	solid
10	Stack Exchange monthly questions vs 2017 peak	286,300 → 9,883 (−29×)	Mar 2026	Stack Exchange Data Explorer	solid
11	Facebook referrals to 792-site panel	1.3bn → 561m (−58%)	Mar 2018 → Mar 2024	Chartbeat / Similarweb	reported
12	Facebook share of external referrals	30% → 7%	Mar 2018 → Mar 2024	Chartbeat	reported
13	X/Twitter referrals to news brands	−65%	Oct 2022 → mid-2025	Chartbeat	reported
14	Crawl-to-refer, Google	~5:1	2026	Cloudflare Radar	reported
15	Crawl-to-refer, largest AI crawlers	~2,000:1 – 4,600:1	2026	Cloudflare Radar (compilations differ)	reported
16	Google Search & other revenue	60.4bn, +19% YoY	Q1 2026	Alphabet SEC filing	solid
17	Google default-placement payments	*26.3bn*	2021	*US v. Google* trial record	solid
18	Global all-device search share	90.4%	May 2026	StatCounter via compilation	reported
19	Chrome browser share	70.25%	May 2026	StatCounter via compilation	reported
20	Tokens fully restricted by robots.txt (head domains)	20–33%	Apr 2024	Data Provenance Initiative	solid
21	C4 restricted by Terms of Service	45%	Apr 2024	Data Provenance Initiative	solid
22	New English web articles primarily AI-generated	50.9% / 49.9%	Q4 2025 / Q1 2026	Graphite (3 detectors, <2% error)	reported
23	New pages containing some AI content	74.2%	Apr 2025	Ahrefs (900k pages)	reported
24		26.1 vs 27.5 CORE	2024	DCLM (Li et al.)	solid

#	Claim	Value	As of	Source	Grade
	Top-20%-by-PageRank filter vs unfiltered baseline				
25	Ask-LLM: beats full-data training after rejecting	90% of corpus	2024	Sachdeva et al.	solid
26	FineWeb-Edu distilled classifier, top-class F1	0.02	2024	Penedo et al.	solid
27	C4 blocklist removal, AAE vs WAE documents	42% vs 6.2%	2021	Dodge et al.	solid
28	JudgeDeceiver attack success rate	90.8–98.9%	2024	Shi et al., CCS	solid
29	GEO visibility lift for rank-5 pages (Cite Sources)	+115.1%	2024	Aggarwal et al., KDD	solid
30	Mastodon MAU vs peak	2.74m → 839k (−69.3%)	Nov 2022 → Jul 2026	Instance statistics	reported
31	Chromium-derived engine share of web sessions	>75%	2026	StatCounter via compilation	reported
32	Significant browser engines in existence	5 (2013) → 3 (2026)	2026	Trade press census	reported
33	Alternative browser engines shipped on iOS since DMA Art. 5(7)	0	Jul 2026 (29 months post-deadline)	Open Web Advocacy; trade press	solid
34	Chrome release removing the last Manifest V2 override	Chrome 150, 30 Jun 2026	Jun 2026	Chromium code review; multiple outlets	reported
35	uBlock Origin Chrome installed base at removal	>40m	2026	Chrome Web Store via trade press	reported
36	Mobile OS share (Android / iOS)	~72.8% / ~27%	2026	StatCounter via compilation	reported
37	iOS background execution for continuous/scheduled/wake-on-request	Not available	2026	Apple developer documentation (first-party)	solid
38	YouTube share of all US TV watch-time	13.8% (platform best; #1 distributor 3rd month)	May 2026	Nielsen Media Distributor Gauge	solid
39	US uploads of YouTube videos ≥20 minutes	6.6m in Q1 2026, +53.5% over 3 years	Q1 2026	Tubular Labs via TVREV / EMARKETER	reported
40	US YouTube viewers watching on connected TV	70.9% (>180m)	2026 forecast	EMARKETER	reported
41	YouTube advertising revenue	9.88bn	Q1 2026	Alphabet SEC filing	solid
42	Ladybird independent engine milestones	alpha 2026 / beta 2027 / stable 2028	2026	Ladybird Browser Initiative (first-party)	solid

Appendix B — Falsifiable propositions of the counter-architecture

Ordered by how fatal being wrong is. Each is a hypothesis with a test and a kill criterion; items 1–4 are make-or-break.

#	Hypothesis	Test	Kill criterion
1	An LLM judge correlates with human quality judgment on web pages	Build a labelled set; measure human–human agreement on our rubric *first*, then judge-vs-human against a popularity baseline	Judge–human agreement no better than the baseline, or below our own inter-annotator agreement
2	The judge identifies the top decile, not merely removes the bottom	Precision/recall by score class on a held-out human-labelled set	Top-class F1 stays near FineWeb-Edu's 0.02
3	The judge resists adversarial optimisation	Pay people to score highly with worthless content using the GEO recipe	Trivially gamed and defences do not move attacker cost
4	Scoring survives structured feature extraction	Score identical pages as prose vs as extracted features; compare against human labels	Depth and originality collapse under extraction
5	Judge scores are stable on re-scoring	Re-score a rotating sample; measure intra-rater Krippendorff's α	Rank churn from re-scoring exceeds rank movement from engagement
6	Engagement validation does not collapse to popularity	The permanent no-boost holdout plus attention-Gini tracking	Gini converges to the status quo web, or holdout outcomes match boosted outcomes
7	The hosting surplus is real in *shape*, not just in aggregate bytes	Simulate with measured churn, uplink asymmetry and Zipf popularity	Long-tail capacity cannot serve head demand without the origin doing most of the work
8	Sybil attacks on engagement are uneconomic	Model attacker cost per unit of stolen attention	Cheaper to fake engagement than to write something good
9	Long-tail bytes stay available	Measure it — no published figure exists for consumer-node blob availability over months	The floor guarantee requires operator nodes to hold effectively everything
10	v0 is buildable at a sane scope	Prototype the minimum viable cut	v0 still requires a browser engine or novel P2P
11	Client-observed engagement signals are materially more robust than link and click signals	Compare rank orderings driven by completion/dwell/return against a links-and-clicks baseline on the same corpus, under the same adversarial budget	Client signals are gamed at comparable cost, which would remove the principal justification for shipping a client
12	A constrained document format is sufficient for the publishing people actually want to do	Track the share of intended publications that cannot be expressed in the format, and the share of readers who bounce to a browser	The format excludes a large fraction of ordinary publishing, or readers treat the client as a supplement to a browser rather than a substitute

A condition, not a hypothesis. The mobile platforms forbid background execution, hosting, and serving (§6.3). No experiment we can run will change this, so it is not listed above. It bounds the design rather than testing it: the hosting obligation of inversion 2 is satisfiable on desktop only, and any claim about the network's aggregate capacity must be read as a claim about the desktop minority of installs.