The Toll for Truth

Paywalls, Information Deserts, and the Asymmetric Web

James Konderla
ORCID: https://orcid.org/0009-0008-8303-2871
Independent Researcher

White Paper | Version 1.0 | October 2026
Not peer reviewed

DOI: https://doi.org/10.5281/zenodo.23131192
Licensed under CC BY 4.0

The Toll for Truth: Paywalls, Information Deserts, and the Asymmetric Web

James Konderla
ORCID iD: https://orcid.org/0009-0008-8303-2871]

White Paper | Version 1.0 | Published October 3, 2026
DOI: https://doi.org/10.5281/zenodo.23131192

© 2026 James Konderla

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). You are free to share and adapt this material for any purpose, provided you give appropriate credit, link to the license, and indicate if changes were made. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/.

Suggested citation
Konderla, J. (2026). *The toll for truth: Paywalls, information deserts, and the asymmetric web* (Version 1.0) [White paper]. Zenodo. https://doi.org/10.5281/zenodo.23131192

Disclaimer
This paper was researched and written independently by the author as an unaffiliated researcher. It was not produced on behalf of, or with the endorsement of, any employer, institution, or organization. The views expressed are solely those of the author.

Funding
This research received no specific grant or funding from any public, commercial, or not-for-profit agency. All costs were borne by the author.

Competing interests
The author declares no competing interests.

Use of AI tools
Portions of this paper were prepared with the assistance of Claude (Opus 5.5), a large language model developed by Anthropic, and Gemini 3.1 Pro, a large language model developed by Google. The tools were used to format citations, generate graphics from rough sketches, and edit for clarity and consistency. The author reviewed, revised, and verified all AI-assisted content, including checking cited sources against the originals. The author takes full responsibility for the content, arguments, and conclusions of this work. AI tools are not listed as authors.

Version history
1.0 (October 3, 2026): Initial release.

Abstract

Digital paywalls have become the dominant revenue strategy for journalism and scholarly publishing, replacing volatile advertising markets with predictable subscription income. This report examines that shift through an information science lens, weighing both the case for paywalls and their systemic costs. It first outlines the architecture of soft, hard, metered, and dynamic paywalls. It then reviews the arguments in their favor: revenue stability, higher-quality first-party data, and insulation from search and AI-driven traffic loss. Against these benefits, the report documents how the authentication barrier sorts readers by ability to pay rather than need. Only about 17% of U.S. adults pay for news, and payment is sharply stratified by income, education, and race. News deserts now cover 213 U.S. counties, affecting roughly 50 million Americans. In scholarly publishing, publicly funded research is sold back to the public at margins approaching 40%, leaving unaffiliated researchers to pay per article or go without. The report then analyzes how large language models and retrieval-augmented generation reconstruct and summarize gated content, eroding publisher revenue while delivering error-prone substitutes to readers who cannot pay. It traces how subsidized and ad-funded content exploits a self-reinforcing "Mass Distribution Loop" of zero-cost production, frictionless access, algorithmic amplification, and machine ingestion. The result is a three-tier information environment in which verified knowledge functions as a luxury good. The report concludes that paywalls are rational for individual publishers but, aggregated across the ecosystem, produce the very information asymmetries that equity frameworks seek to prevent, and proposes policy, library, publishing, AI-governance, and advertising reforms to address them.

Keywords

paywalls, information equity, news deserts, open access, scholarly publishing, generative AI, misinformation, digital divide

Konderla, J. (2026). The Toll for Truth, v1.0
https://doi.org/10.5281/zenodo.23131192

Contents

Overview 1
1. Types of Paywalls and Their Architectural Mechanics 1
Soft Paywalls 1
Hard Paywalls 2
Metered and Dynamic Paywalls 2
LLMs and the Ethics of Information Access 3
2. Analysis of Arguments for Paywalls 5
Technical and Economic Realities of Content Monopolization 5
Data Analysis Ecosystems 8
Insulation from Systemic AI Traffic Drops 10
Revenue Generation and Sustainable Capital Flow 12
3. Comprehensive Analysis of Arguments Against Paywalls 13
The Exclusivity Backfire Loop 13
Inequitable Access and Structural Disadvantage 14
4. The AI Conflict: Information Retrieval, Synthesis, and System Evasion 21
LLM Transformer Architecture 22
The Structural Shift: From Indexing to Retrieval-Augmented Generation (RAG) 24
AI Synthetic Summary Generation 24
Systemic Vulnerability to Generative AI Reconstruction 27
5. Socio-Political Implications: Information Deserts and Asymmetric Diffusion 29
The Acceleration of Disinformation Ecosystems 31
The "Luxury Tier" Classification of Truth 36
6. Recommendations 39
Extend and Measure Public Access Policy 39
Resource Libraries to Serve Users Outside the Academy 40
Convert the Crisis Exception into a Standing Civic Tier 40
Hold AI Systems Accountable for Access and Accuracy 41

Konderla, J. (2026). The Toll for Truth, v1.0
https://doi.org/10.5281/zenodo.23131192

Demonetize the Mass Distribution Loop 41
Conclusion 41
Bibliography 44

Table of Figures

Figure 1, Information Flow Model Diagram 6
Figure 2, Anonymous VS Authenticated Ingestion Pipeline 8
Figure 3, Traditional SEO vs Direct Portal Architecture 10
Figure 4, Subscription Enforcement Metric Matrix 13
Figure 5, Who Pays for News (Pew Research Center, 2025) 15
Figure 6, Paying Twice: The Scholarly Publishing Chain 17
Figure 7, How LLMs Deconstruct & Synthesize Content 22
Figure 8, Content Synthesis Mechanics 25
Figure 9, AI Reconstruction Artifact Pipeline 28
Figure 10, Structural Asymmetry of the Web 30
Figure 11, The Content Polarization Vacuum 32
Figure 12, Anatomy of the Mass Distribution Loop 34
Figure 13, The “Luxury Tier” of Truth 38
Figure 14, Matching Responses to Problems 39

Overview

In the discipline of Information Science (IS), digital paywalls mark a foundational architectural shift from open-access, high-entropy information environments to closed, low-entropy knowledge repositories. Historically, the web operated under an advertising-dependent paradigm that prioritized search engine findability and click-through metrics, often compromising data integrity and metadata standards in favor of "pageview journalism".

Modern paywall frameworks restructure these information flows. By instituting precise authorization layers, system administrators can regulate data access, standardize identity management pipelines, and treat content as secured capital assets rather than raw data commodities for external web indexers. This report synthesizes the structural benefits ("Arguments For") and systemic vulnerabilities ("Arguments Against") of digital paywall deployments within the broader context of information economics, retrieval mechanics, and equitable access. Finally, the article examines the effects of AI not only on information flow but also on the socio-economic impacts of changes in that flow.

1. Types of Paywalls and Their Architectural Mechanics

Soft Paywalls

A soft paywall limits free access to all content by offering limited viewing or granting access to select content before prompting a subscription (MacDonald, 2024). This model, often called a "freemium" approach, offers a mix of free and premium content and is widely used by content creators on platforms like Substack, as well as by legacy media outlets such as The New York Times and The Washington Post (Ryaboy, 2023; Simon & Graves, 2019). From an information systems perspective, the technical mechanics of a soft paywall rely heavily on client-side execution. When a user requests an article, the host server transmits the complete HTML file—including the full text, links, and structure—directly to the user's browser (Vasipenok, 2025). The restriction occurs after the page loads, when Cascading Style Sheets (CSS) and JavaScript inject an overlay, such as a pop-up or a blur effect, that obscures the content from view (Vasipenok, 2025). Because the data payload is fully delivered to the client, soft paywalls are

structurally vulnerable to bypassing techniques. Users and automated agents can manipulate the Document Object Model (DOM) by disabling JavaScript, altering user-agent strings to mimic search engine crawlers (e.g., Googlebot), or using web archives and proxy tools to strip the code (Ryaboy, 2023). Furthermore, modern "agentic" artificial intelligence browsers can parse raw HTML directly from the source code, bypassing the visual CSS barriers intended to restrict human readers (Vasipenok, 2025).

Hard Paywalls

Unlike soft barriers, which are permeable, hard paywalls require a subscription to access content and severely limit organic discovery (Ryaboy, 2023). Elite, specialized, and financial publications typically adopt this model when they offer a highly desirable product and have a dedicated audience. The Wall Street Journal and the Financial Times are prime examples (Ryaboy, 2023; Williams & Stroud, 2021). Architecturally, hard paywalls are enforced server-side rather than client-side. When a user makes a request, the server first verifies their login credentials and subscription status. If the user is unauthenticated, the server withholds the HTML blueprint entirely (Vasipenok, 2025). This server-side approach makes hard paywalls highly resistant to traditional client-side scraping and paywall-bypass extensions. However, in today’s digital landscape, sophisticated AI chatbots have begun to circumvent even hard paywalls—not by breaching servers, but through "fragment reconstruction" (Pelcro, 2025). By systematically mining digital breadcrumbs such as social media discussions, quoted excerpts, and secondary coverage, AI systems can reconstruct much of a paywalled article without accessing the host server. Pelcro (2025), a subscription-software vendor, estimates that such summaries capture 70% to 80% of an article's value, while independent testing found that AI tools reconstructed roughly half of the paywalled articles sampled (van Ess, 2025).

Metered and Dynamic Paywalls

The metered paywall serves as a sampling mechanism, allowing users to access a set amount of free content—such as a specific number of articles each month—before the paywall appears (Simon & Graves, 2019). Publications such as Wired have adopted this model to offer potential

users a “preview,” with the aim of turning casual readers into loyal subscribers by first demonstrating the value of their journalism (Bilton, 2018). Technically, metered models typically use browser cookies or IP address tracking to count the articles a user has read. As a result, users often try to bypass these limits through simple data-clearing methods, such as deleting browser cookies, using incognito mode, or relying on ad-blocking software (Concept47, 2019; Ryaboy, 2023).

Dynamic paywalls are an evolution of subscription models. They use algorithms and user data to analyze readers’ browsing habits and behavior in order to optimize subscription conversions. Although highly effective for publishers moving away from traffic-dependent advertising models (Pelcro, 2025), heavy reliance on user data and rigid paywalls raises significant ethical and moral concerns within information science. Gating content creates a two-tiered information ecosystem: high-quality, fact-checked journalism becomes a luxury for wealthier, more educated consumers, while the broader public is left with free, algorithmically driven content often rife with misinformation and clickbait (Stengel, 2024; The New Yorker, 2020). Information ethicists argue that restricting access to civic information based on financial status creates unacceptable social asymmetries and conflicts fundamentally with the democratic principle of a well-informed public (Williams & Stroud, 2021). The contradiction inherent in the paywall model is often exposed during major public events, such as elections or natural disasters, when publishers temporarily lift paywalls—implicitly acknowledging that restricting access to vital information harms public welfare (Howard, 2022; Stengel, 2024).

LLMs and the Ethics of Information Access

As the web enters the era of Large Language Models (LLMs), publishers face a new threat that bypasses the paywall entirely. This architectural shift has created a “Zero Referrals” crisis (often called zero-click searches), fundamentally changing how audiences engage with premium reporting.

When users query a search engine, generative AI features can scrape, synthesize, and reconstruct paywalled material into a "Synthetic Target Output"—a comprehensive summary displayed directly on the search results page. By extracting the core facts and narrative from a premium article and presenting them instantly, AI leaves users with little reason to click through or encounter the publisher's paywall. This deprives the original creator of potential subscription conversions and display ad impressions. By eliminating the need for human browsing, AI system evasion creates an asymmetric dynamic: publishers bear the financial cost of investigating and reporting the news, while AI platforms capture the traffic and user engagement.

Automated extraction is expanding rapidly. Recent network analyses indicate that major infrastructure providers process up to 50 billion AI crawler requests each day, while AI crawler traffic surged 757% in a single year, according to Coronium.io, a proxy-services provider. In addition, “training crawlers"—bots designed to systematically scrape web content for model-training datasets—accounted for 49.9% of all AI bot traffic in early 2026 (Coronium.io, 2026). By contrast, only 2.2% of AI bot requests appeared to serve actual user queries (Coronium.io, 2026). These figures suggest that most crawler interactions with paywalled or premium sites are extractive, prioritizing raw data collection over bringing in organic human readers (Coronium.io, 2026).

This extractive model is further complicated by declining citation integrity and attribution accuracy. Even when generative AI systems try to cite the original publisher, they often produce "ghost attributions" or hallucinated links (Andres SEO Expert, 2026). Complex paywalls, aggressive lazy loading, and JavaScript-heavy elements can prevent AI crawlers from indexing live pages, leading LLMs to rely instead on their internal, pre-trained knowledge bases (Andres SEO Expert, 2026). As a result, an AI may cite a URL it remembers rather than verified, live content (Andres SEO Expert, 2026). This not only denies publishers accurate referral credit but may also put their domains at risk: if users click a hallucinated link and leave immediately, the generative engine may mistakenly flag the publisher's domain as low quality (Andres SEO Expert, 2026).

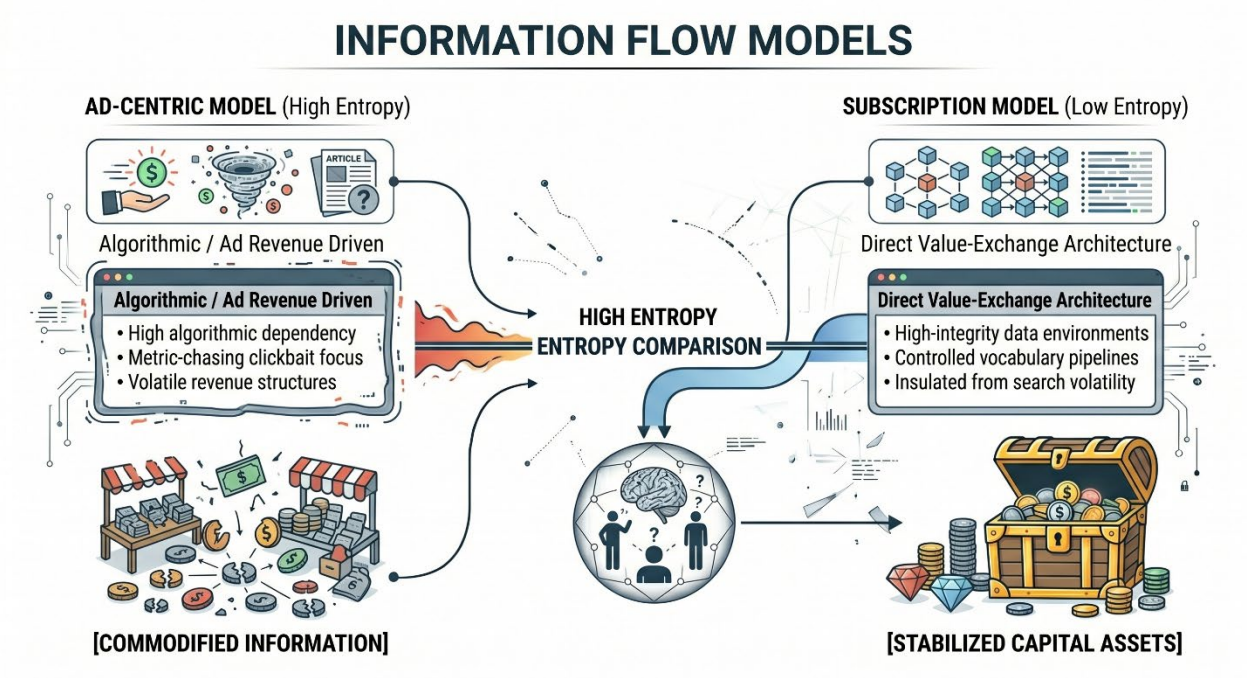
The technical mechanics by which automated agents parse and reconstruct proprietary texts—using Transformer architectures, vector embeddings, and latent space mapping—are highly complex and will be examined in greater detail in Section 4 (The AI Conflict: Information Retrieval, Synthesis, and System Evasion). From an ethical standpoint, however, the normalization of zero referrals and large-scale, uncompensated data scraping fundamentally threatens the economic viability of structured, human-verified knowledge environments.

2. Analysis of Arguments for Paywalls

Technical and Economic Realities of Content Monopolization

From a technical perspective on information and digital asset curation, the modern paywall is a critical infrastructural adaptation. It shifts systems away from the high-entropy, attention-driven metrics of the early web toward closed, structured value-exchange architectures. Throughout the early 2000s, the dominant online information paradigm relied on ad-centric, traffic-dependent distribution models. This design optimization incentivized algorithmic manipulation, search engine manipulation, and rushed, keyword-stuffed publishing models that significantly eroded public trust.

Figure 1, Information Flow Model Diagram



The comparative architecture above shows how a platform's business model directly shapes the technical properties and systemic value of its data pipeline. The model maps these differences across a spectrum evaluated by the **High Entropy / Low Entropy Comparison** layer:

- **The Ad-Centric Model (Left Channel - High Entropy):** This design tracks the information flow of traditional open-access web repositories, which depend entirely on an **Algorithmic / Ad Revenue Driven** framework.
    - *System Characteristics:* This architecture is plagued by a high algorithmic dependency, a metric-chasing clickbait focus, and volatile revenue structures that fluctuate alongside changes to search engine or social algorithms.
    - *Data Quality & Asset State:* Because content must be rushed out, packed with trending keywords, and optimized for maximum traffic, it degrades into [COMMODIFIED INFORMATION]. This environment strips data of long-term metadata integrity, feeding into a systemic cycle of public skepticism and media distrust.

- **The Subscription Model (Right Channel - Low Entropy):** Positioned on the opposite side of the transaction divide is the paywall-enforced data ecosystem, which operates via a **Direct Value-Exchange Architecture**.
    - *System Characteristics:* By establishing an authentication barrier, system administrators cultivate high-integrity data environments and tightly monitored, controlled vocabulary pipelines. This closed design ensures that inbound traffic routes directly through the portal, leaving the database completely insulated from search volatility and zero-click AI summary cannibalization.
    - *Data Quality & Asset State:* Rather than treating text as an ephemeral commodity, the subscription engine turns publishing streams into highly structured, premium digital properties designated as [STABILIZED CAPITAL ASSETS].
- **The Behavioral and Financial Outcome (Lower Center and Right Nodes):** When an enterprise transitions its data framework from an ad-centric to a subscription model, its valuation metrics change fundamentally. Zara (2013) argues that subscription revenue frees newsrooms from chasing trending topics and clickbait, giving them the stability to pursue community-oriented watchdog journalism. As shown by the transaction lines routed to the asset chest, Zara also links this predictability to higher newspaper stock values.

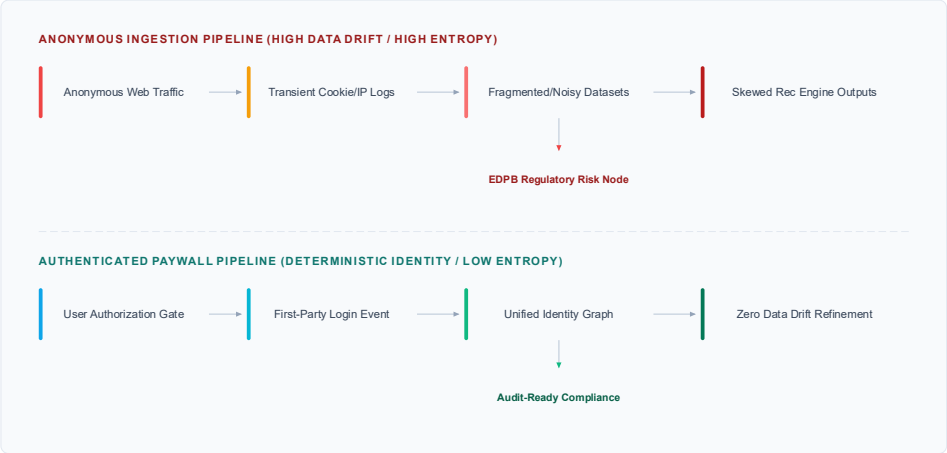
To counter this disruption, media networks and database curators deployed paywalls to protect their intellectual property from volatile advertising channels. When assessing the fiscal viability of these models, Information Science scholars track asset appreciation. Zara (2013) reports that newspaper stock values rose sharply, in some cases by 50% to 80%, after publishers adopted paywalls, suggesting that investors favor predictable subscription fees over volatile, ad-supported pageview systems. This structural shift indicates that investors favor predictable subscription fees over volatile, ad-supported pageview systems.

This economic pivot has drastically altered enterprise information logistics. As noted above, Dotdash Meredith cut Google Search's share of its traffic from about 60% in 2021 to just over a third by 2024, a shift Pelcro (2025) credits with insulating the company from AI-driven traffic losses.

Data Analysis Ecosystems

Paywalls create structured, logged-in environments that streamline data curation and refine enterprise user analytics. Under an open web model, tracking anonymous traffic yields fragmented, high-noise behavioral datasets corrupted by transient user sessions and evolving privacy constraints. In contrast, an authenticated paywall interface shifts data collection away from speculative third-party cookies toward a centralized identity graph powered by direct first- and zero-party data (Pattabhiramaiah et al., 2018). This infrastructure establishes a uniform privacy layer in which users explicitly consent to standardized data collection.

Figure 2, Anonymous VS Authenticated Ingestion Pipeline



While an authentication barrier initially reduces unique top-of-funnel traffic by 16.8%, it also systematically filters low-value automated bot activity and casual web surfers from the ingestion pipeline (Pattabhiramaiah et al., 2018). This selective filtering isolates a highly engaged user base, reducing subscription churn risk by 31% and providing data engineers with predictable, longitudinal behavioral data (Pattabhiramaiah et al., 2020). It also allows system engineers to

capture high-integrity data on user interactions, semantic query habits, and detailed document utility. Processing authenticated user streams instead of volatile web traffic significantly reduces data drift, helping optimize the training loops for internal collaborative filtering and recommendation algorithms. In addition, this authenticated architecture offers important regulatory protection. Under modern legal frameworks, including the European Data Protection Board's (EDPB) "Consent or Pay" guidelines, traditional unauthenticated "cookie walls" are increasingly penalized for failing to obtain freely given consent. A structured paywall offers a legally distinct, compliant alternative that balances user privacy options with sustainable digital asset tracking.

Within the framework of information evaluation, access restrictions fundamentally change a database's utility and perceived value. Open-access information environments are structurally associated with low verification standards, metric-driven optimization, and ad-crowded interfaces that degrade the user experience. Enforcing exclusivity through an authentication gateway draws on the core tenets of Timothy Brock's Commodity Theory, which holds that restricting access to an information asset increases its perceived value and institutional prestige (Brock, 1968). This deliberate structural barrier creates a cognitive boundary, distinguishing verified data assets from the unverified, high-entropy text streams saturating public web indexes.

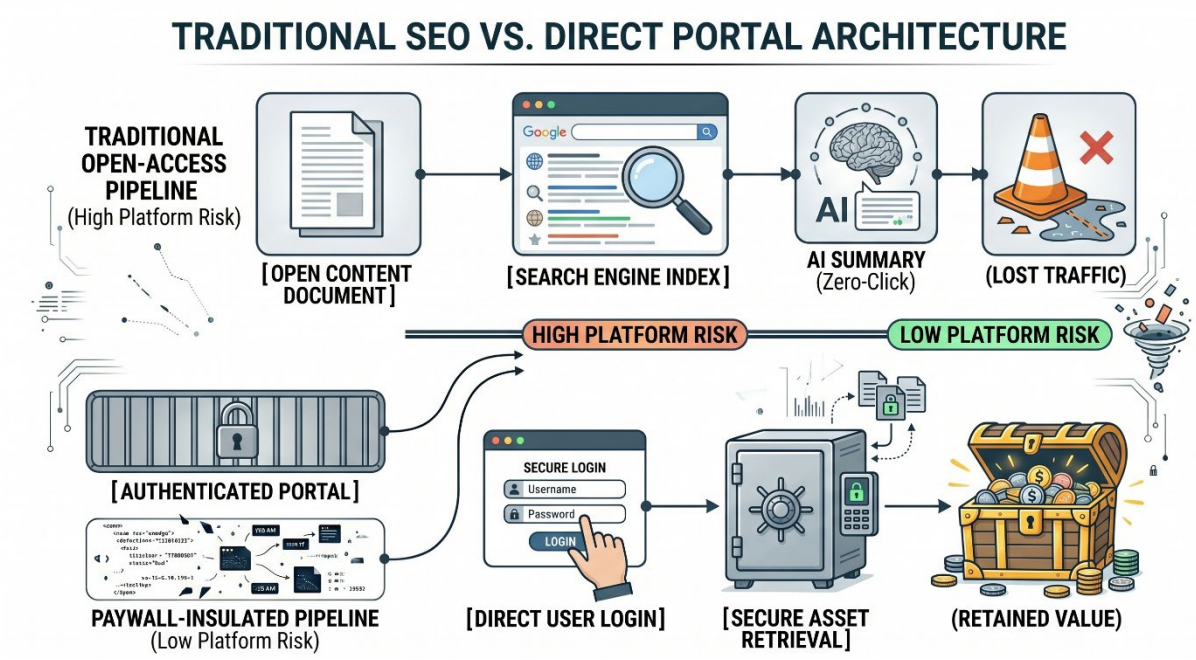
Michael Spence's Signaling Theory in information economics helps explain this phenomenon (Spence, 1973). In markets with asymmetric information, consumers cannot readily assess data integrity before consuming it. Because low-quality clickbait hubs and automated content mills cannot financially sustain themselves behind a transaction gate, a paywall acts as a credible, "costly signal" of structural authority and editorial rigor. Information seekers tend to trust repositories requiring an explicit exchange of resources, viewing the payment barrier as an operational proxy for extensive peer review, fact-checking, and standardized indexing of the underlying documents (Center for Media Engagement, 2021).

Insulation from Systemic AI Traffic Drops

The integration of generative artificial intelligence overviews within major search engines has led to significant traffic declines across premium content domains. Traditional open sites are vulnerable to zero-click searches, where conversational interfaces answer user queries by extracting information and bypassing the source site entirely.

By cultivating a subscription-locked ecosystem, publishers shift their core traffic from anonymous search referrals to direct, authenticated user bases. Pelcro (2025), a subscription-software vendor, reports that Dotdash Meredith reduced its dependence on search, citing CEO Neil Vogel's statement that Google Search had fallen to just over a third of the company's traffic, down from roughly 60% in 2021.

Figure 3, Traditional SEO vs Direct Portal Architecture



The architectural flow model in Figure 3 above visualizes the technical divergence between reliance on third-party discovery channels and independent access management. The system maps these behaviors into two distinct pipelines divided by their level of platform risk:

- **The Traditional Open-Access Pipeline (Top Channel - High Platform Risk):** This sequence models how unsecured content assets are systematically devalued by conversational information retrieval systems.
    - The pipeline starts with an [Open Content Document] that is fully exposed to public crawlers.
    - This document is ingested into a [Search Engine Index] (such as Google).
    - Rather than forwarding the user to the host domain via a traditional hyperlink, the platform generates an AI Summary (Zero-Click) natively within the search interface.
    - The operational result is designated by the warning icon: (Lost Traffic). The information consumer extracts the asset utility directly from the search hub, starving the publisher of ad impressions, tracking cookies, and inbound traffic.
- **The Paywall-Insulated Pipeline (Bottom Channel - Low Platform Risk):** This alternate route demonstrates how introducing defensive authentication layers preserves asset value.
    - The pipeline establishes an **[Authenticated Portal]** backed by strict authorization guardrails to safeguard the primary data repository.
    - To gain entry, the user must go through a **[Direct User Login]** prompt requiring active verification credentials (Username and Password).
    - Once authorized, the connection routes to a **[Secure Asset Retrieval]** system (represented by the locked vault), which securely delivers data packages exclusively to verified sessions.
    - The end state is represented by the treasure chest: (Retained Value). By eliminating external discovery dependencies, the publisher retains direct

platform control, captures accurate analytics, and maintains stable financial equity.

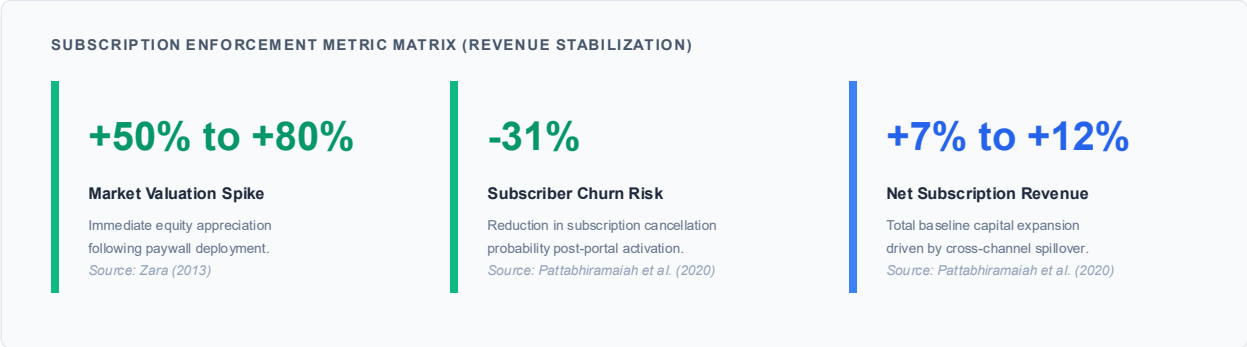
Revenue Generation and Sustainable Capital Flow

From an information economics perspective, content monetization screens transition media networks from volatile, traffic-dependent advertising markets to direct value-exchange systems. This architectural shift insulates media houses from sudden changes in third-party search engine optimization (SEO) algorithms, platform distribution dependencies, and declines in conversational AI traffic. Under the legacy advertising paradigm, media companies operated as high-entropy data pipelines, forcing content creators to optimize exclusively for pageview volume, thereby structurally incentivizing clickbait and eroding long-term database integrity. By establishing an authentication gate, system administrators can treat information streams as stabilized capital assets rather than fleeting data commodities.

While digital advertising once formed the backbone of online news revenue, publishers have pivoted aggressively toward subscription models to stabilize their capital flows. By 2019, approximately 70% of newspaper websites in the U.S. and Europe had implemented some form of digital paywall architecture (Pattabhiramaiah et al., 2020). The financial justification for this structural pivot extends far beyond merely attracting new top-of-funnel subscribers; it fundamentally alters the longitudinal equity and retention economics of existing reader bases.

Empirical research indicates that when an existing subscriber activates their digital paywall access, their risk of canceling the subscription decreases by approximately 31%, culminating in a net subscription revenue increase of 7% to 12% for the publishing enterprise (Pattabhiramaiah et al., 2020). This economic phenomenon, known as "cross-channel spillover," demonstrates that gating premium content increases the perceived value of the publisher's broader portfolio, creating a sustainable, highly predictable revenue loop rather than relying on fluctuating daily ad impressions.

Figure 4, Subscription Enforcement Metric Matrix



Furthermore, the macroeconomic valuation of media properties directly reflects this predictability. Financial analyses in information economics show that the transition to consistent paywall frameworks is associated with immediate market valuation spikes of 50% to 80% across public media enterprise listings (Zara, 2013). Wall Street valuation models strongly favor predictable, recurring subscription engines over volatile, ad-supported arbitrage frameworks that remain highly vulnerable to sudden shifts in audience discovery platforms.

As documented by financial analysts, establishing a reliable baseline of Annual Recurring Revenue (ARR) fundamentally alters internal resource allocation in newsrooms. Freed from the pressure to track trending sidebars, produce high-frequency content variations, or engineer headlines to appease external search algorithms, media organizations gain the structural capital security needed to reinvest in deep, community-oriented investigative reporting and rigorous watchdog journalism (Zara, 2013). Paywalls, therefore, affect both the fiscal security of the media firm and the foundational quality of the knowledge assets stored within public information spheres.

3. Comprehensive Analysis of Arguments Against Paywalls

The Exclusivity Backfire Loop

While paywalls provide a reliable revenue floor, they also trigger an "Exclusivity Backfire Loop" marked by a severe hemorrhage of casual readership and digital engagement. Implementing a paywall forces a stark trade-off: trading the volume of an open network for the monetization of

a closed one. Pattabhiramaiah et al. (2018) found that introducing a paywall reduced unique visitors by 16.8%, along with significant declines in engagement measures such as pages viewed and time spent per visitor. Other estimates are higher: Streyczek (2025) finds that paywalls reduce page views by an average of 25% to 30%, a decline that appears within the first two months and persists for years. The effect is especially damaging for local news outlets. Olsen et al. (2019) found that moving from free to paid access lastingly reduced pageviews and unique visitors, whether the outlet used a hard or soft paywall. The backfire loop is further exacerbated by user circumvention; as information is gated, a parallel ecosystem of browser extensions, web archives, and AI tools have emerged to strip away these digital barriers, leaving publishers in a continuous, costly technological arms race to protect their exclusivity.

Inequitable Access and Structural Disadvantage

Information equity frameworks show how digital barriers reinforce existing social divisions. Paywalls create financial obstacles that prevent lower-income groups, researchers at underfunded institutions, and marginalized populations from accessing verified knowledge bases.

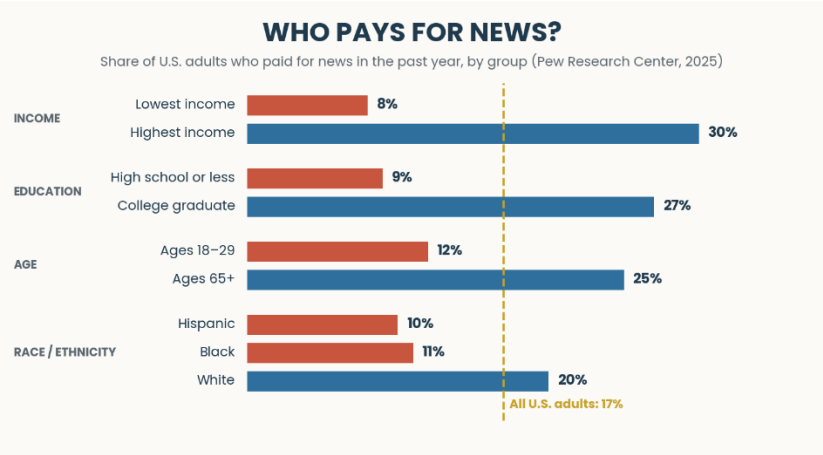
Section 2 described the authentication barrier as a filtering mechanism: it removes casual traffic and automated noise, leaving a highly engaged, paying user core (Pattabhiramaiah et al., 2018). Viewed from the other side of the gate, the same mechanism acts as a sorting algorithm. A paywall cannot tell whether a reader is unwilling or unable to pay; both are sent to the same locked endpoint. Since the ability to pay closely follows income, education, and institutional affiliation, the engaged core publishers cultivate is not a random sample of the public, but a demographic slice of it. As a result, the access architecture reproduces existing social stratification within the information supply itself, before anyone reads a single article.

The Income Gradient of Paid Information

Survey data make this gradient measurable. In a March 2025 survey of 9,482 U.S. adults, the Pew Research Center found that only 17% of Americans had paid for news in the previous year, and that payment was sharply stratified by income: 30% of adults in the highest income tier had

paid, compared with 8% in the lowest (Pew Research Center, 2025). The same pattern holds for education and race. College graduates paid at roughly three times the rate of adults with a high school diploma or less, and White Americans were about twice as likely to pay as Black or Hispanic Americans (Pew Research Center, 2025). In separate Pew focus groups, a participant in his twenties described paying for news as a luxury, a framing that anticipates the “Luxury Tier” classification discussed in Section 5 (Nieman Lab, 2026).

Figure 5, Who Pays for News (Pew Research Center, 2025)



Internationally, the ceiling appears structural rather than temporary. Across the twenty higher-income markets tracked by the Reuters Institute, the share of people paying for any online news stalled at 18% in 2025 and slipped to 17% in 2026. Most markets clustered between 10% and 20%, and the United States fell to 16% (Reuters Institute for the Study of Journalism, 2025, 2026). From an information architecture perspective, this means the subscription model’s paying population is capped at a minority of users, even in the wealthiest information economies. A high-integrity repository that moves behind an authentication barrier tends to narrow its paying audience to roughly one person in six, and those left outside are disproportionately lower-income, less-educated, and minority populations that information equity frameworks are designed to protect.

Geographic Disadvantage and News Deserts

This dynamic creates "news deserts"—geographic or socioeconomic spaces without reliable, human-verified information networks. As high-quality data is locked away, vulnerable populations must rely on unverified information channels, widening the digital divide. Northwestern University's annual State of Local News report documents the geography of this exclusion. In 2025, the number of U.S. counties with no local news source reached a record 213, up from roughly 150 in 2005, while more than 1,500 additional counties were left with just one source, usually a weekly newspaper (Medill Local News Initiative, 2025). Together, roughly 50 million Americans now have limited or no access to local news. These deserts are not randomly distributed: nearly 80% are in counties the USDA classifies as predominantly rural, and they tend to be poorer counties (Medill Local News Initiative, 2025).

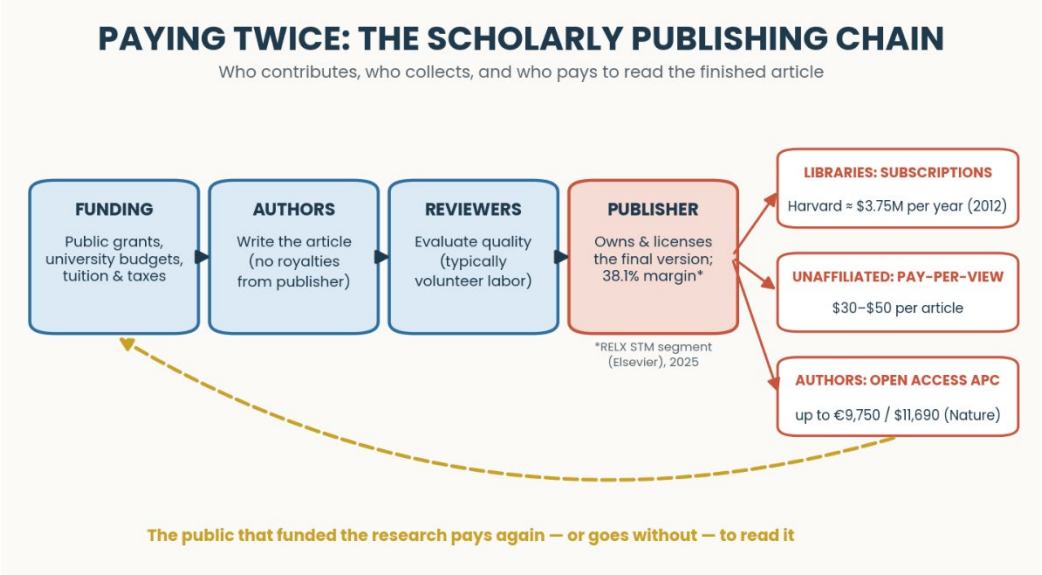
The report also identifies the market's replacement mechanism—and it does not reach the places that need it most. New digital-only local news sites are emerging in counties whose demographics are nearly the opposite of news deserts: more affluent, with lower poverty and higher educational attainment (Medill Local News Initiative, 2025). In other words, the reader-revenue model presented in Section 2 as a path to sustainability regenerates local journalism where residents can afford to support it, while leaving the deserts largely untouched. Section 5 examines the downstream political and informational consequences of these vacuums, including their colonization by subsidized partisan networks and "pink slime" sites.

Artificial Research Gates

If news paywalls stratify access to civic information, scholarly publishing shows what a fully matured paywalled knowledge environment looks like. In academic research, the gate is not a recent adaptation to a collapsing advertising market; it is the business model. The production chain is unusual among information industries. Research is frequently funded by public grants or university budgets, written by academics who receive no royalties from the publisher, and evaluated by peer reviewers who typically volunteer their time. The finished article is then licensed back to the same universities and to everyone else through subscription packages or

per-article fees. Authors participate because hiring, tenure, promotion, and grant decisions lean heavily on publication in high-prestige journals, many of which remain subscription-based. Researchers remain attached to these prestige indicators even when they object to the pricing behind them (Bishop, 2020). The researcher is pushed to publish behind a paywall whether they would choose to, and the public that funded the work is pushed to the other side of it.

Figure 6, Paying Twice: The Scholarly Publishing Chain



The institutional cost of this model has been visible for more than a decade. In 2012, Harvard University's Faculty Advisory Council issued a memorandum warning that major journal publishers had made the scholarly communication environment fiscally unsustainable. The council reported that Harvard was spending close to $3.75 million a year on journals from a small number of providers, that some individual titles cost as much as $40,000 annually, and that prices from two providers had risen roughly 145% over six years while publishers reported profit margins of 35% or more (Faculty Advisory Council, 2012). Those margins have not eroded. In 2025, the Scientific, Technical & Medical segment of RELX, Elsevier's parent company, reported £2.71 billion in revenue and £1.04 billion in adjusted operating profit, for an adjusted operating margin of 38.1% (RELX, 2026). If the wealthiest university in the United States could describe its subscriptions as untenable, the position of regional public universities, community

colleges, public libraries, and institutions in lower-income countries follows directly, since they buy access in the same market on a fraction of the budget.

For researchers outside an institution, the cost falls on the individual. Publisher pay-per-view prices typically range from $30 to $50 per article (DeepDyve, 2026), putting even a modest literature review of 100 sources at $3,000 to $5,000 before analysis begins. The population facing this cost is broad: independent scholars, journalists, policy analysts at small nonprofits, clinicians outside academic medical centers, small-business researchers, and recent graduates who lose access to most of their university's databases after leaving (Duke University Libraries, 2018). Public libraries offer a partial bridge, but not all can afford the specialized scholarly databases researchers need. Libraries can help close this gap through deliberate design. Efird-Green et al. (2024) describe a community-engaged research e-library whose original documentation disclosed only in a footnote that some resources required an institutional login. The rebuilt version removed those barriers and established a working standard for assessing whether a resource is genuinely usable without affiliation. Because these individuals are outside the institutions that collect usage and access data, their exclusion is also largely invisible in the metrics publishers and libraries rely on, making it difficult to measure and easy to overlook. Even affiliated researchers cannot rely on their institutions alone. In a survey of faculty at U.S. master's universities, only 44% named their home library as their primary source of articles (Walters, 2019). When an article or book is not readily available through the library, 56% of U.S. faculty report giving up and seeking another resource, while 41% purchase the material themselves (Blankstein & Wolff-Eisenberg, 2019, as cited in Walters, 2019). If researchers with institutional support pay out of pocket or abandon sources, the burden on those without that support is likely greater, yet no baseline measure exists.

The scale of unmet demand is clear in the ways researchers route around the gate. An analysis of Sci-Hub server logs, covering the six months from September 2015 through February 2016, found that the unauthorized repository of research papers had handled roughly 28 million download requests from every region of the world (Bohannon, 2016). Use was not limited to lower-income nations: about a quarter of requests came from wealthy OECD countries, and

some of the most intense activity clustered around U.S. and European university campuses. This suggests that friction in legitimate access systems affects even researchers who are nominally affiliated (Bohannon, 2016). Survey evidence points in the same direction. Among 3,304 published researchers across disciplines surveyed by Segado-Boj et al. (2022), more than half reported having used pirate sites to obtain papers at least once. Because respondents were published authors, the group consisted largely of affiliated academics rather than unaffiliated researchers. This discussion makes no judgment about the ethics of such workarounds, which fall outside its scope; the relevant observation is architectural. When a gate is priced beyond the reach of many legitimate users, demand does not disappear. It shifts to whichever channel offers the least resistance.

Open access (OA) publishing has expanded in response, but it has mostly shifted the barrier rather than removed it. A large-scale analysis of 67 million articles estimated that at least 28% of scholarly literature was freely available, rising to 45% among articles published in 2015. It also found that OA articles received roughly 18% more citations than average, underscoring how access itself shapes scholarly influence (Piwowar et al., 2018). Even so, one widely cited estimate suggests that nearly 75% of scholarly documents remain accessible only behind a paywall (Boudry et al., 2019, as cited in Segado-Boj et al., 2022). The dominant commercial route to OA, however, replaces the reader's paywall with a fee charged to authors: the article processing charge (APC). For example, publishing an open access article in Nature carries an APC of €9,750 or $11,690 (Springer Nature, n.d.-a). Publishers do offer waivers, but these are generally limited to authors in the world's lowest-income countries or granted on a case-by-case basis (Springer Nature, n.d.-b). Researchers whose grants or institutional agreements cover these charges can make their work open; those without such funding, including the unaffiliated researchers described above, must choose between a paywalled venue and a fee few individuals could afford. This outcome was anticipated. Novo and Onishi (2017) warned that a rapid shift to gold open access would not eliminate publishing costs but transfer them to authors through APCs. Walter and Mullins (2019) estimate that, of a $3,000 APC paid to Elsevier, only about $1,800 supports the publishing process, while the remaining $1,200 goes directly to

shareholders. The structural disadvantage thus shifts from the reading side of the transaction to the publishing side, determining not only who can read the literature but also whose work enters the open record at all.

Public policy has begun to address some of this imbalance. In 2022, the White House Office of Science and Technology Policy (OSTP) directed federal agencies to eliminate the 12-month embargo on federally funded research and make both publications and their underlying data freely available immediately, stating that "there should be no delay between taxpayers and the returns on their investments in research" (OSTP, 2022, p. 3). The National Institutes of Health's revised Public Access Policy, effective July 1, 2025, also eliminated the previous 12-month embargo. It requires accepted manuscripts of NIH-funded articles to be freely available in PubMed Central upon publication. NIH also announced plans to cap allowable publication costs starting in fiscal year 2026 (Galter Health Sciences Library, n.d.). These are meaningful corrections, but their scope is narrow: they apply only to federally funded work, make the author's accepted manuscript available rather than the final published version, and do nothing for the large backlist of research published before the policy took effect or for work funded by private, state, or international sources. For researchers outside the institutional perimeter, the default remains a locked door. Policies of this scale are also being set without baseline data on what unaffiliated researchers actually spend to access the literature. They are among the taxpayers the OSTP memorandum intends to serve, yet whether it has reduced their costs remains unmeasured.

Substitution Through Synthetic Channels

The excluded reader increasingly has somewhere new to turn. Free AI assistants now answer questions about news and research directly, offering users who cannot afford subscriptions the easiest way around the gate. But the reliability of this route is uneven. In a 2025 study coordinated by the European Broadcasting Union and led by the BBC, journalists from 22 public service media organizations in 18 countries evaluated 2,709 responses to news questions from the free consumer versions of ChatGPT, Copilot, Gemini, and Perplexity. Overall, 45% of responses contained at least one significant issue. Sourcing failures were the largest category, at

31%, while 20% had significant accuracy problems, including fabricated or altered quotes and outdated facts (European Broadcasting Union & BBC, 2025). The researchers deliberately tested the free tiers because these represent the default experience for most users, and noted that free consumer models had received fewer upgrades than paid versions (European Broadcasting Union & BBC, 2025).

The same study found that even when assistants cited real, high-quality sources, users could not always verify them because those sources were behind paywalls (European Broadcasting Union & BBC, 2025). This creates a compounding disadvantage: readers unable to afford a subscription are directed to a free synthetic summary; that summary is wrong or poorly sourced in a substantial share of cases; and the citation that would let them check it leads back to the very paywall that excluded them. AI systems thus create a secondary tier of access, mirroring the stratification of the open web. Section 4 examines the technical mechanics behind these failures, including hallucinated attributions and fragment reconstruction.

Across news, local information, and scholarship, the pattern is consistent. Paywalls turn information that functions as a public good into a club good. The burden of exclusion falls hardest on those with the least institutional support, but excluded readers do not stop seeking information; they turn to channels that are free, fast, and less reliable. From the perspective of an individual publisher, the commercial logic justifying the authentication barrier in Section 2 holds. Aggregated across an information ecosystem, however, it produces precisely the asymmetries that information equity frameworks warn against.

4. The AI Conflict: Information Retrieval, Synthesis, and System Evasion

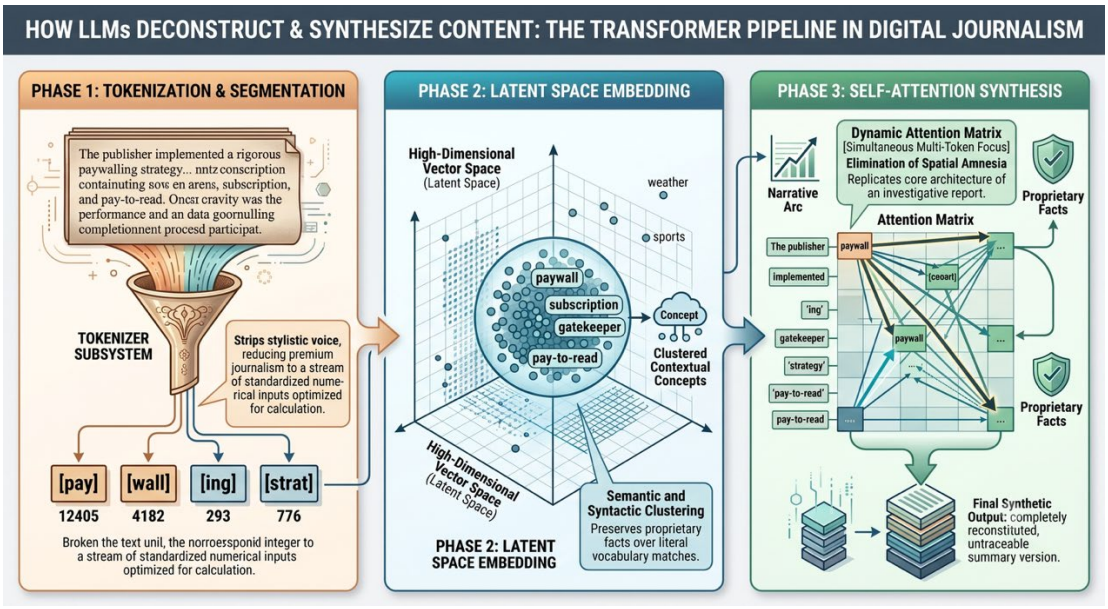
The rise of large language models (LLMs) and advanced retrieval-augmented generation (RAG) systems has triggered a new phase of conflict over digital paywalls. Conversational user interfaces have fundamentally changed consumer discovery mechanics, leading to sharp declines in traditional web referral traffic. Between 2024 and 2025, traffic to the world's 500 most-visited publishers fell 20% to 27% year over year, an average loss of roughly 64 million

visits per month (Similarweb data, as cited in Pelcro, 2025). The New York Times saw referral traffic drop 27.4% in Q2 2025 alone (Pelcro, 2025) as interactive AI platforms began satisfying user queries without redirecting them to the original content hosts.

LLM Transformer Architecture

To understand how automated systems systematically dismantle and reconstruct premium digital content, one must examine the fundamental architectural mechanics of Large Language Models (LLMs). At their core, modern LLMs are built upon the **Transformer architecture** (Vaswani et al., 2017). This specific deep learning framework abandons traditional sequential processing in favor of parallelized statistical operations that treat language not as a linear sequence of strings but as a multidimensional mathematical landscape, as shown below.

Figure 7, How LLMs Deconstruct & Synthesize Content



As Figure 7 summarizes, the architectural lifecycle of an LLM processing text consists of three sequential structural phases: Tokenization and Segmentation, Latent Space Embedding, and Self-Attention Synthesis, which the following sections describe in basic terms.

Tokenization and Segmentation

An LLM cannot read text directly. When a system crawls a publisher's web page, it strips the raw HTML and feeds the text into a tokenizer. This subsystem breaks words, characters, or subwords into discrete structural units called tokens. For example, it might mathematically divide “paywalling” into "pay," "wall," and "ing." Each token is assigned a unique integer from a fixed vocabulary index. This mechanical process strips away the reporter's original stylistic voice, reducing premium journalism to a stream of standardized numerical inputs optimized for algorithmic calculations.

Latent Space Embedding

Once tokenized, these numerical identifiers are converted into vector embeddings, or mathematical representations of data. An embedding layer maps each token to an abstract, high-dimensional vector (often with thousands of dimensions) in what is known as the latent space. Within this mathematical space, tokens are positioned relative to one another according to semantic and syntactic relationships learned during training.

Tokens with similar contexts—such as "paywall," "subscription," and "gatekeeper"—are clustered closely together. This spatial arrangement lets the model process abstract concepts rather than relying on literal text matches. It also explains how an AI agent can read a copyrighted article and seamlessly substitute vocabulary while preserving its core proprietary facts, data points, and narrative conclusions.

The Self-Attention Mechanism and Contextual Synthesis

The defining engine of the Transformer architecture is its self-attention mechanism. Traditional neural networks process words one at a time, often losing context across long paragraphs. Self-attention overcomes this spatial amnesia by simultaneously calculating dynamic mathematical weights between every token in a text block.

When an LLM evaluates a sentence, the attention mechanism determines how much "focus" or weight to assign to other words in the document to interpret the current token's precise

meaning. This enables the model to follow complex narrative arcs, connect pronouns to less obvious subjects across pages, and extract the underlying structure of an investigative report.

The Structural Shift: From Indexing to Retrieval-Augmented Generation (RAG)

Understanding this architecture shows why LLMs pose a fundamentally different threat to media business models than traditional search engine crawlers such as Googlebot. Classic search engines use spiders to index pages and build a digital catalog of hyperlinks. Their architecture is designed to point outward, directing users to the original publisher through referral clicks.

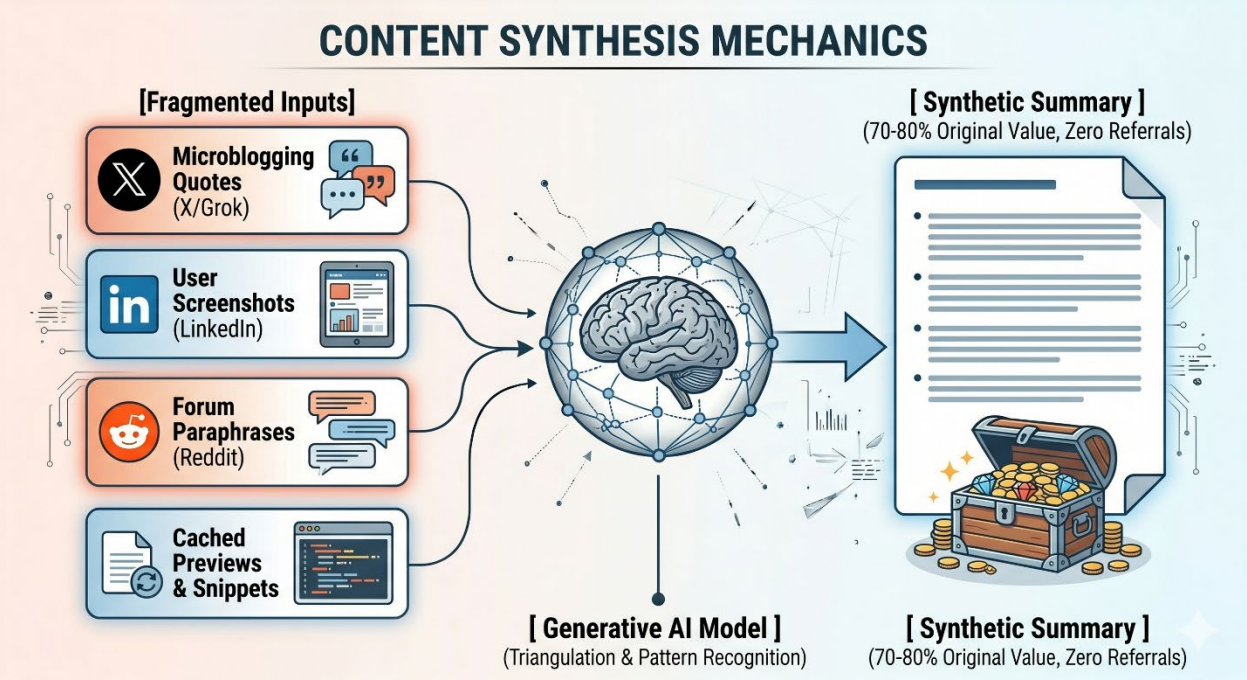
In contrast, an LLM treats ingested data as raw chemical elements: it breaks them down, stores them as statistical weights, and reconstitutes them. With Retrieval-Augmented Generation (RAG) or live-search capabilities, the model does more than memorize its training data; it functions as an automated synthesis machine.

The system reads real-time, tokenized fragments of a gated article, maps them into its latent space, uses self-attention to compute their core semantic meaning, and generates a completely novel, legally distinct, yet factually identical Synthetic Target Output. By delivering this comprehensive synthesis directly to the end user, the model fulfills the user's informational need internally, bypasses the publisher's client-side paywall script entirely, and establishes the paradigm of Zero Referrals.

AI Synthetic Summary Generation

From a technical perspective, conversational AI engines often use a process called "content synthesis" to bypass digital paywalls. Instead of using standard scrapers that can be blocked by automated user-agent rules, modern AI agents act as distributed aggregators. They gather fragments of premium content from public web trails—including social media quotes, screenshots shared on professional platforms, forum discussions, open RSS feeds, and cached preview snippets. “The model then pieces these elements together, using pattern recognition to reconstruct the original paywalled article with 70% to 80% accuracy” (Pelcro, 2025).

Figure 8, Content Synthesis Mechanics



The visualization provided above outlines the technical pipeline behind decentralized information harvesting. Rather than executing a direct intrusion on a host publisher's data repository, the process relies on public footprint aggregation to render a high-fidelity mirror of closed-access documents:

- **Ingestion Layer (Left Node):** The system displays the four foundational [Fragmented Inputs] used to capture leaked data trails from the open web. *Microblogging Quotes (X/Grok)* extract short-form reporting transcripts, while *User Screenshots (LinkedIn)* preserve visual text matrices. Simultaneously, *Forum Paraphrases (Reddit)* supply structural rewordings of core source arguments, and *Cached Previews & Snippets* establish baseline textual patterns. Each input provides a distinct piece of the asset's overall reporting.
- **Triangulation Layer (Center Node):** These open artifacts are routed directly into the [Generative AI Model]. This orchestration framework does not rely on standard web scraper parameters that would trigger traditional server-side logs. Instead, it uses

triangulation and pattern recognition across the gathered fragments to fill informational gaps and draw conclusions based on historical context.

- **Downstream Synthesis (Right Node):** The endpoint of the pipeline presents the finalized [Synthetic Summary] accompanied by an asset protection graphic. The result is a clean, bulleted summary delivered inside the AI interface, capturing most of the original article's informational value, by Pelcro's (2025) estimate. Because this interaction occurs entirely inside the AI agent's network ecosystem, it satisfies the query while returning Zero Referrals to the original publisher.

According to Pelcro, “Information tracking exposes four primary user behavioral vectors used to bypass traditional paywalls via generative AI”, summarized below:

- **Direct Summarization Prompts:** Explicit commands forcing an agent to triangulate public summaries and metadata to recreate paywalled articles without generating an inbound referral link.
- **Social Fragment Mining:** Exploiting real-time indexing capabilities (e.g., Grok scanning X or specialized agents parsing Reddit threads) to compile disparate user discussions of locked content into a unified technical narrative.
- **Granular Metric Extraction:** Requesting specific data points, bulleted takeaways, or core intellectual arguments of highly regulated documentation (such as *Wall Street Journal* financial indicators), which effectively extracts the information utility while starving the publisher of ad traffic or conversion metrics.
- **Academic/Technical Reconstruction:** Systematically query-mining academic abstracts, citation networks, and secondary open-access commentaries to structurally replicate locked scholarly or proprietary research papers.

Publishers have responded with technical defenses. TollBit, which monitors AI scraping for publishers, reported blocking 26 million scraping attempts in March 2025 alone, and Cloudflare saw bot traffic rise from 3% to 13% in a single quarter (Pelcro, 2025). Publishers also use

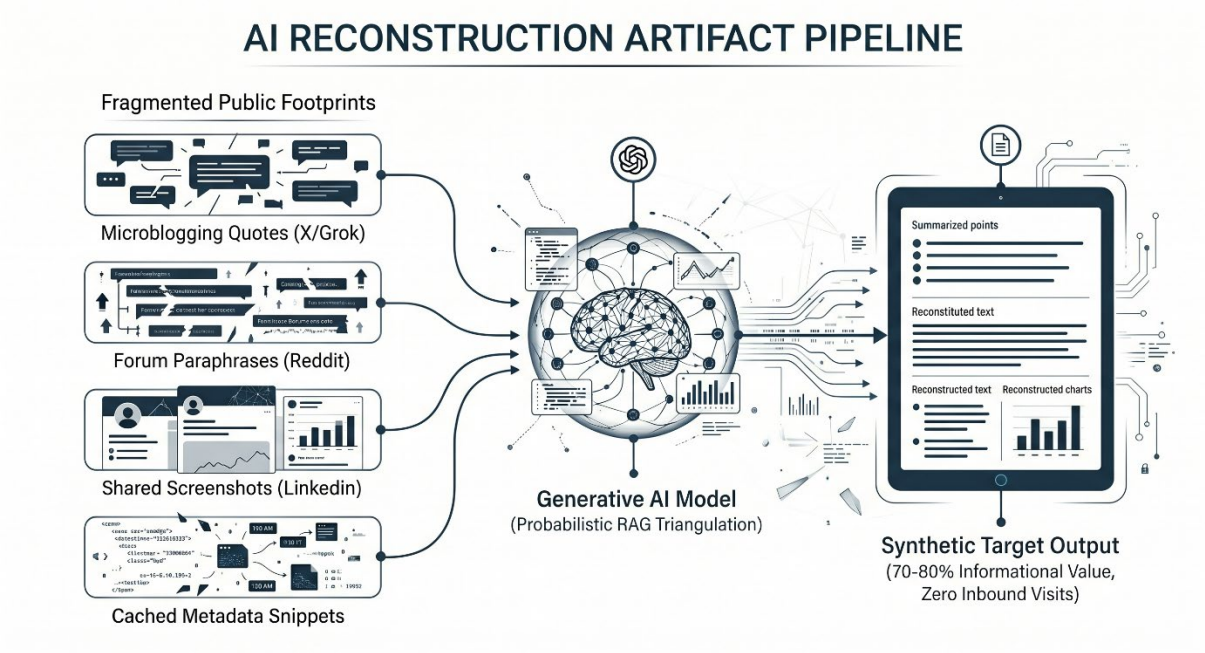
semantic honeypots—decoy file pathways designed to trick and flag non-human collectors. However, because conversational models can synthesize information using third-party web footprints without ever directly pinging the original content host, perimeter security setups remain limited. This reality poses an ongoing challenge for Information Science: designing digital distribution frameworks that balance fair compensation for creators with open access to human knowledge.

Systemic Vulnerability to Generative AI Reconstruction

Advanced Information Retrieval (IR) configurations show that digital paywalls are increasingly vulnerable to evasion by Large Language Models (LLMs) using Retrieval-Augmented Generation (RAG). Modern conversational agents don't need to scrape hidden servers directly or trigger perimeter defense logs to undermine content utility.

Instead, these models act as automated digital detectives, collecting publicly available traces left across the web, including social media quotes, screenshots posted to professional networks, discussion forum summaries, open RSS feeds, and cached previews. By piecing these fragments together through pattern recognition, generative systems can rebuild the core arguments of a paywalled article (Pelcro, 2025). This process provides end-users with the information they need without generating inbound referral traffic or conversion events for the primary creator.

Figure 9, AI Reconstruction Artifact Pipeline



The architectural workflow shown demonstrates how modern Retrieval-Augmented Generation (RAG) frameworks systematically abstract the core utility of closed-access repositories. Rather than a singular, direct exploit of a target server, the system relies on distributed data aggregation across three distinct phases:

- **Ingestion of Fragmented Public Footprints (Left Phase):** The diagram models the four primary information leakage vectors that circumvent traditional perimeter defenses.
    - *Microblogging Quotes (X/Grok)* and *Forum Paraphrases (Reddit)* capture real-time, user-submitted text fragments and discussions surrounding a locked asset.
    - *Shared Screenshots (LinkedIn)* introduce unstructured visual data, while *Cached Metadata Snippets* provide structural guardrails (such as headers, publication times, and brief SEO-exposed summaries). Each represents an isolated, open-access byte of a much larger, paywalled narrative.
- **Probabilistic RAG Triangulation (Center Phase):** Once gathered, these disparate data streams flow into the central processing hub—the *Generative AI Model*. Instead of

relying on a static database query, the system maps the inputs onto a semantic neural net (represented by the central brain network). By cross-referencing and triangulating these external data fragments, the model resolves informational contradictions and infers missing structural content without ever initiating a direct network request to the host publisher.

- **Generation of the Synthetic Target Output (Right Phase):** The final stage involves converting raw data fragments into a cohesive, user-facing document. The system generates a clean user interface displaying a structured text document containing *summarized points*, *reconstituted text*, and *reconstructed charts*.

By delivering a synthesized dataset or summary directly to the user's interface, the framework achieves complete information retrieval while ensuring that zero inbound visits or referral metrics are returned to the primary creator.

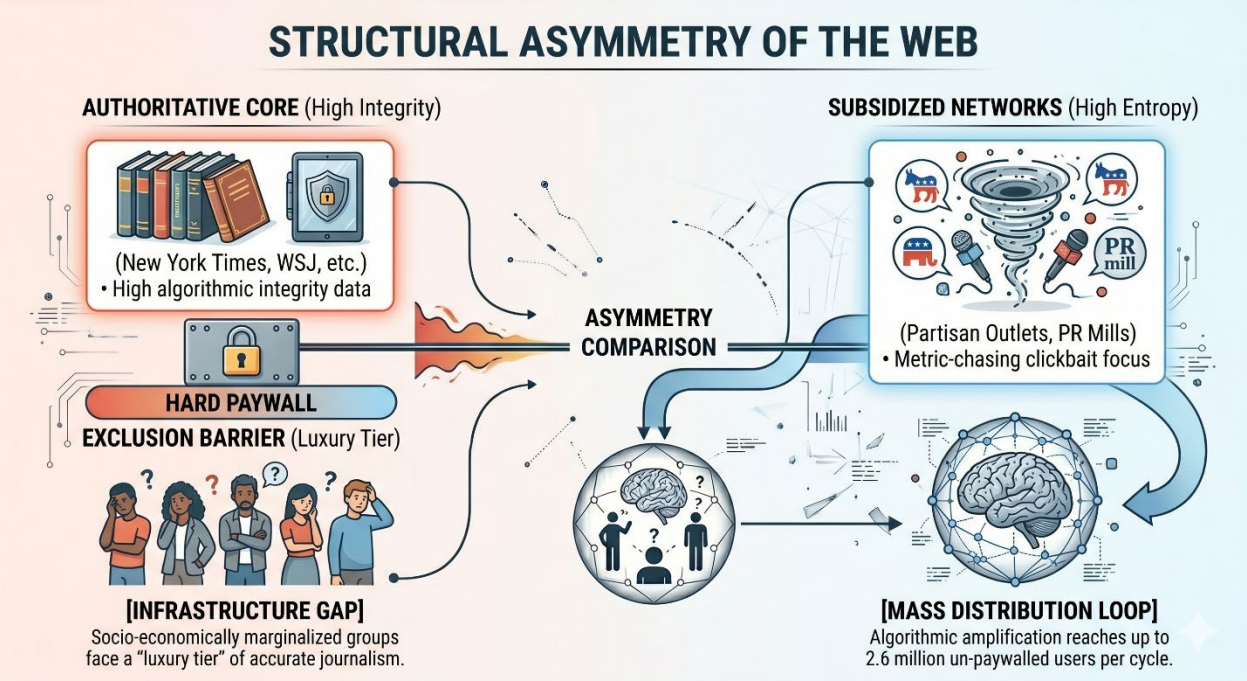
5. Socio-Political Implications: Information Deserts and Asymmetric Diffusion

When premium web portals make transaction screens difficult to use, the underlying dynamics of public information access can change quickly. In Library and Information Science (LIS) research, restricting access to verified information is seen as a major factor in the growth of "news deserts", geographic or socio-economic areas without credible, human-verified information networks. Automated text streams, corporate PR campaigns, and low-cost content mills quickly fill this vacuum.

This displacement affects ideological frameworks unevenly. While centrist and progressive legacy outlets have increasingly adopted user-funded paywalls to sustain their operations, conservative media networks have often relied on other funding sources, including private donors and political action committee (PAC) subsidies. As a result, right-wing information channels can distribute free, unpaywalled content on a massive scale. “Single-topic,

algorithmically boosted articles from these subsidized alternative networks regularly gain significant traction on open social media indexes, frequently reaching over 2.6 million users per transmission cycle" (Pareene, 2020).

Figure 10, Structural Asymmetry of the Web



The comparative workflow above illustrates how modern internet architecture creates stark, unequal pathways for information distribution based on the economic models different publishers choose. This split structure divides the digital media landscape into two competing channels, separated by an Asymmetry Comparison boundary:

- **The Authoritative Core (Left Channel - High Integrity):** This sector represents elite, human-verified legacy publications (e.g., *New York Times*, *WSJ*) that focus on producing high algorithmic integrity data. To stabilize their business model against ad market volatility, these outlets implement a HARD PAYWALL, creating a rigid EXCLUSION BARRIER. Because this information is restricted to a Luxury Tier, it creates a structural [INFRASTRUCTURE GAP]. Socio-economically marginalized groups who cannot afford the

toll are blocked from direct access, which turns these communities into information-starved "news deserts".

- **Subsidized Networks (Right Channel - High Entropy):** Operating on the opposite side of the structural divide are alternative channels consisting of partisan outlets and digital PR mills. Because these platforms are heavily subsidized by private donors or political groups, they reject subscription gates in favor of a [MASS DISTRIBUTION LOOP]. These properties utilize a high-entropy, metric-chasing clickbait focus on game recommendation engines and capture public attention. Each stage of this loop is examined in “Anatomy of the Mass Distribution Loop” below.
- **The Behavioral Outcome (Center and Lower Right Nodes):** When marginalized users seek information, their default cognitive pathways are steered away from the locked core and channeled directly into the free open-access loop. As the diagram shows, the un-paywalled content mills leverage social media amplification to achieve massive reach, “injecting their ideological assumptions into the public sphere by capturing **up to 2.6 million un-paywalled users per cycle**” (Pareene, 2020). This dynamic cements a deep societal divide: affluent citizens purchase premium, fact-checked reporting, while vulnerable populations are left to consume free, algorithmically boosted propaganda.

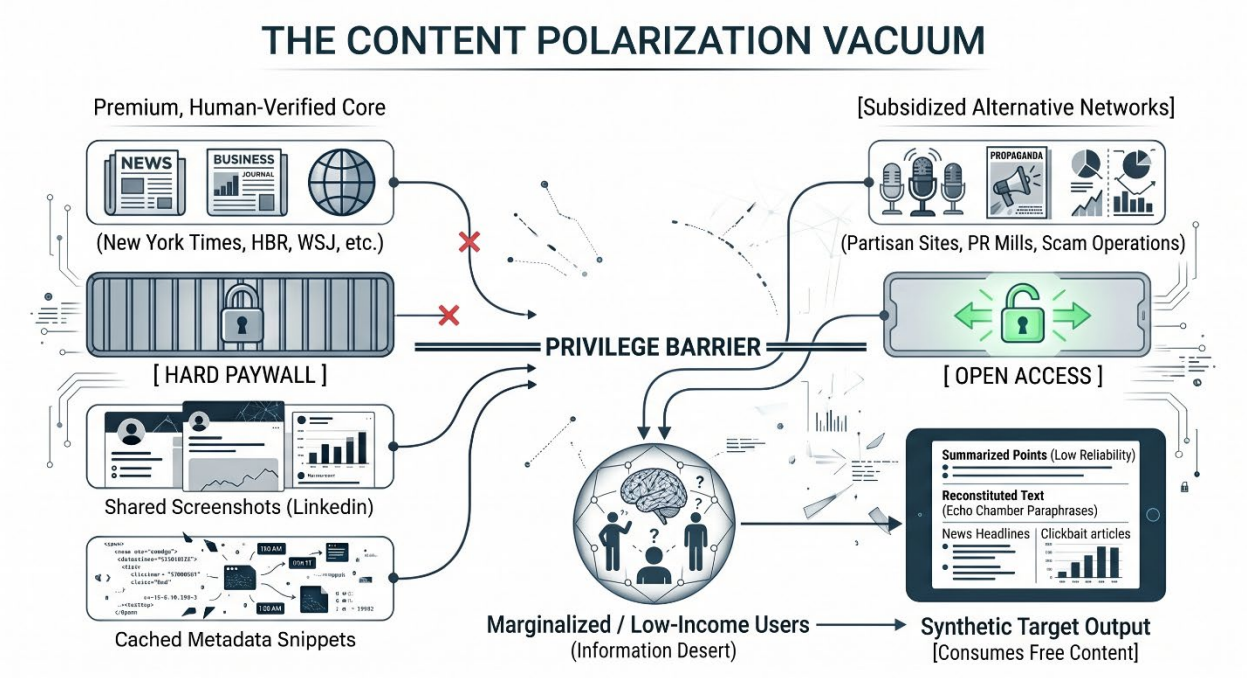
This dynamic creates a highly inequitable "two-tiered system." Affluent individuals pay for access to a luxury tier of verified data, while marginalized populations are left with low-quality, ad-heavy "pink slime" sites and automated text mills. This stratification poses a major challenge to digital literacy. Because open-access content mills often replicate the visual layout, metadata styles, and professional presentation of authoritative news sites, the average user struggles to distinguish objective journalism from paid political propaganda.

The Acceleration of Disinformation Ecosystems

Erecting paywalls around reputable journalism creates an information vacuum in public spaces. Because human-curated, fact-checked news is expensive to produce and is often hidden behind

paywalls, alternative networks fill the open web with free, universally accessible propaganda and corporate public relations.

Figure 11, The Content Polarization Vacuum



The systemic stratification of digital media access visualized in Figure 11 illustrates how economic transaction barriers structurally re-engineer the public information sphere. Rather than a neutral marketplace of ideas, the system establishes a split ecosystem divided by a centralized Privilege Barrier:

- **The Locked Authoritative Core (Left Upper Quadrant):** This sector represents high-fidelity, human-verified information networks (e.g., *New York Times*, *HBR*, *WSJ*). When these premium nodes deploy a strict access filter (**HARD PAYWALL**), they sever direct public access (indicated by the red **X** markers). While a small sub-channel of data still leaks downstream via *Shared Screenshots (LinkedIn)* and *Cached Metadata Snippets*, the direct transmission pipeline of verified raw information is blocked for unauthenticated consumers.

- **The Open-Access Contamination Vectors (Right Upper Quadrant):** In stark contrast to the locked core, Subsidized Alternative Networks (encompassing partisan hubs, corporate PR outlets, and monetization operations) deliberately use a zero-friction distribution architecture (**OPEN ACCESS**). Unhindered by payment screens or subscription gates, these heavily subsidized platforms pump high-volume, unverified content directly across the privilege divide.
- **The Emergence of Information Deserts (Center Lower Node):** Positioned beneath the economic divide are Marginalized / Low-Income Users. Cut off from the primary authoritative core by financial barriers, these user populations are forced into a localized digital ecosystem where access to high-quality information is scarce.
- **The Consumption of Low-Reliability Synthetic Output (Right Lower Quadrant):** Because human-verified data is financially restricted, the open web vacuum is filled by alternative media streams. As shown in the final consumer stage, marginalized populations access a Synthetic Target Output optimized for free distribution. This consumption loop exposes users to a mix of *Summarized Points (Low Reliability)*, *Reconstituted Text (Echo Chamber Paraphrases)*, unverified *News Headlines*, and highly engaging *Clickbait articles*. This dynamic solidifies the two-tiered information system, steering vulnerable populations away from rigorous data structures and leaving them reliant on subsidized, low-quality alternative media.

Many media organizations and interest groups rely on alternative financing, including subsidies from private billionaires and funding from political action committees, enabling them to distribute free content at scale. Shifting factual news from a public good to a club good carries serious socio-political consequences. When quality journalism is available only through premium subscriptions, readers who cannot or will not pay may turn to free, lower-quality alternatives for information (Streyczek, 2025). This creates a "Mass Distribution Loop" for fake news: unverified, sensationalized content costs little to produce and nothing to consume, so it spreads easily across social media. As Pareene (2020) notes elsewhere, subscription models

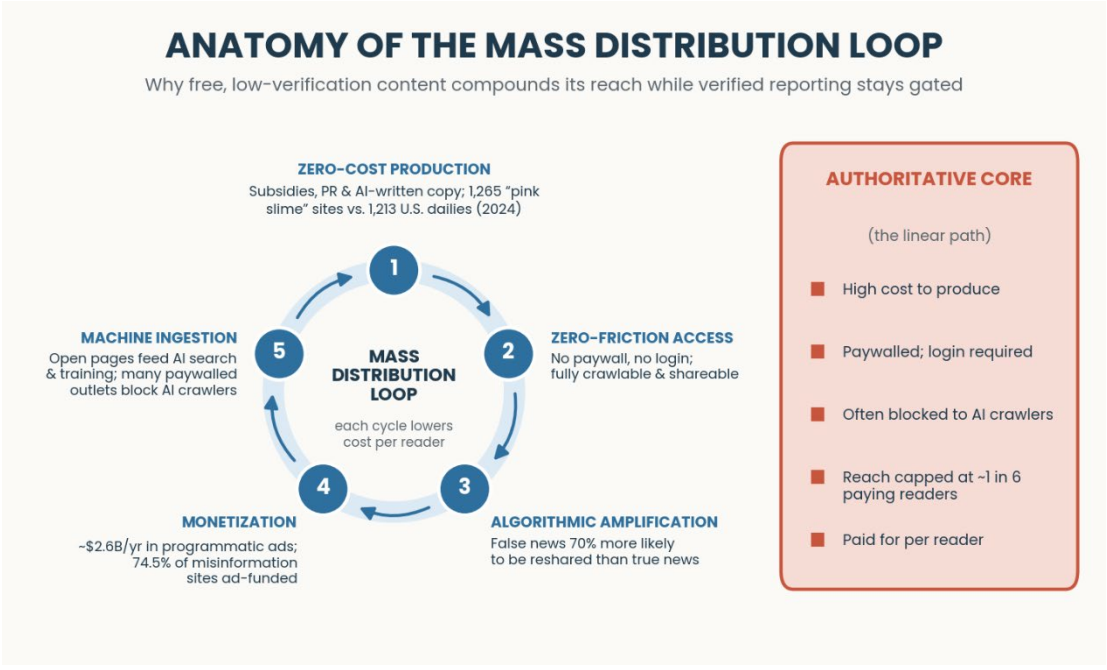
filter for readers seeking quality news while excluding those who would benefit but can't or won't pay. The infrastructure of misinformation is also highly profitable: pervasive digital ad networks automatically place brand advertising across the web, financially sustaining fake news sites and monetizing outrage while truth remains behind a paywall (Ahmad et al., 2023).

The empirical consequences of this asymmetry are alarming. In a 2025 study, Julian Streyczek found that "increased exposure to paywalls makes individuals 1.6% to 2.4% less likely to correctly answer factual questions about political realities, driving a measurable decline in political knowledge and lower electoral participation among non-subscribers" (Streyczek, 2025).

Anatomy of the Mass Distribution Loop

The "Mass Distribution Loop" is more than a metaphor for virality. It describes a self-reinforcing cycle in which each stage lowers the cost of the next, so that free, low-verification content gains reach with every iteration while paywalled reporting remains on a linear, per-reader cost curve. The loop can be decomposed into five stages, as illustrated in Figure 12.

Figure 12, Anatomy of the Mass Distribution Loop



- **Zero-Cost Production:** Subsidized networks do not need to recover production costs from readers. Funding arrives from donors, political committees, corporate public relations budgets, or advertising, and the volume this makes possible is now measurable. As of June 2024, NewsGuard had identified 1,265 “pink slime” sites, outlets that present themselves as neutral local news while being backed by partisan groups or hostile governments, surpassing the 1,213 daily newspapers still operating in the United States (NewsGuard, 2024). NewsGuard also documented the first apparent crossover of pink slime sites with AI-generated content in a pro-Russia network of 167 sites (NewsGuard, 2024), further lowering the marginal cost of each article.
- **Zero-Friction Access:** Because these sites have no paywall, login, or meter, every article is fully available to readers, social platforms, search indexes, and AI crawlers simultaneously. Their content travels as complete text, while paywalled reporting travels, if at all, as a headline, a snippet, or a screenshot.
- **Algorithmic Amplification:** Engagement-ranked platforms reward novelty and emotional intensity, characteristics that favor falsehood. In an analysis of roughly 126,000 rumor cascades spread by about 3 million people on Twitter between 2006 and 2017, Vosoughi et al. (2018) found that false news was 70% more likely to be retweeted than true news, and that true stories took about six times as long to reach 1,500 people. Removing bots from the dataset did not change the result; people, not automation, drove the difference.
- **Monetization:** Programmatic advertising turns reach into revenue by distributing ads algorithmically across the web, often giving advertisers little visibility into where they appear. Ahmad et al. (2024) found that advertising financially sustained 74.5% of the misinformation websites they studied, and companies using digital advertising platforms were about ten times more likely to have their ads appear on misinformation sites. A joint analysis by NewsGuard and Comscore (2021) estimated that programmatic advertisers send roughly $2.6 billion annually to misinformation publishers. The two

business models now mirror each other: credible outlets increasingly charge readers, while misinformation outlets charge advertisers and give their content away.

- **Machine Ingestion:** The loop closes through AI systems. Open pages are the most accessible inputs for AI search and model training, yet many authoritative outlets actively exclude themselves. By the end of 2023, 48% of the most widely used news websites across ten countries were blocking OpenAI's crawlers, a figure that rose to 79% in the United States (Fletcher, 2024). The consequences for chatbot output remain debated. A NewsGuard audit found that ten leading chatbots repeated false narratives from the Kremlin-linked Pravda network—which published an estimated 3.6 million articles in 2024—33% of the time when prompted about those narratives (Sadeghi & Blachez, 2025). An independent audit found little evidence of deliberate manipulation, attributing such references instead to "data voids," topics where reputable information is scarce and low-quality sources dominate (Alyukov et al., 2025). Either explanation supports this report's argument: when credible coverage is absent or locked away, whatever remains openly available fills the void.

With each pass through the loop, the subsidized channel's cost per reader falls, while the authoritative core continues to bear the full cost of serving each reader. This asymmetry is not a temporary market imbalance that subscription growth will eventually correct; it is a structural feature of an information environment where verification is priced and falsehood is subsidized.

The "Luxury Tier" Classification of Truth

In a statement on rising subscription costs, George Washington University described scholarly resources as "not luxury goods" that are nonetheless "priced as though they were" (Bohannon, 2016, p. 511). The same description applies to verified journalism. Earlier sections documented how this pricing works; this section examines what it means in practice to classify verified information as a luxury good.

The systematic restriction of high-integrity databases creates a stark two-tier system: an affluent tier able to purchase verified information and an under-resourced tier left with low-quality automated text streams, content mills, and ad-heavy sites.

What the affluent tier purchases is not information itself, since headlines, summaries, and fragments of most stories circulate freely, but verification: the editorial standards, sourcing, peer review, and correction processes that make claims accountable. Section 2 described the paywall as a costly signal of editorial rigor (Spence, 1973). But that signal reaches only those who pay. For everyone else, verification appears as a locked door, while the free tier offers content without the accountability attached to it. The EBU and BBC evaluators made a similar point about AI assistants, noting that news organizations have processes to identify, acknowledge, and correct errors and calling for comparable accountability from AI tools that now summarize their work (European Broadcasting Union & BBC, 2025).

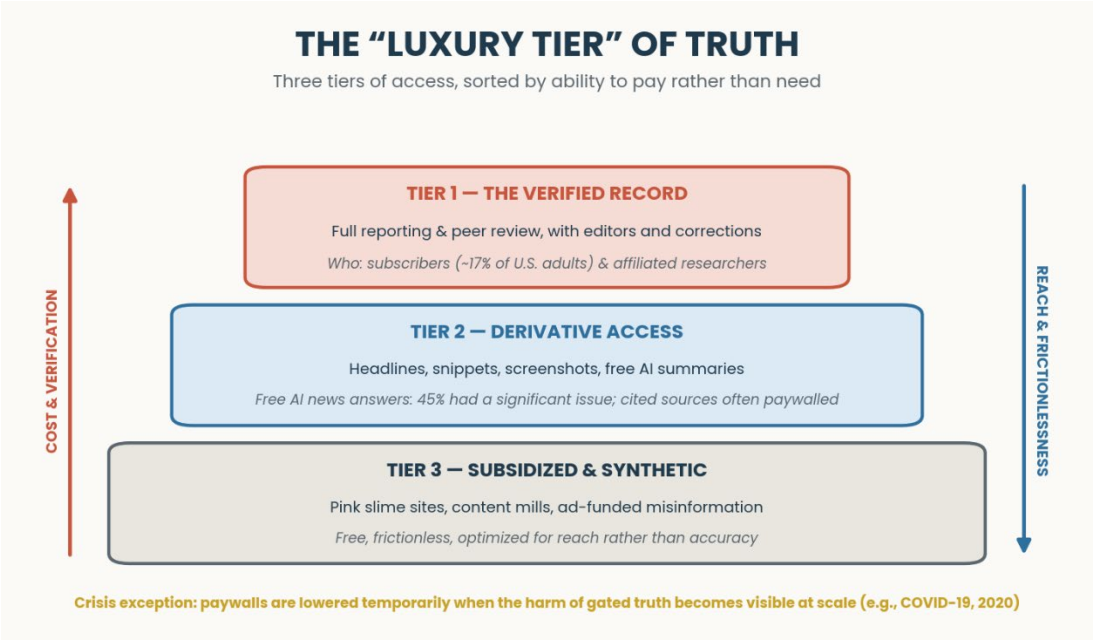
The industry's own behavior during emergencies acknowledges the cost of this classification. As noted in Section 1, publishers routinely lower paywalls during elections and natural disasters (Howard, 2022; Stengel, 2024). The clearest example came at the outset of COVID-19. In January 2020, the Wellcome Trust called on researchers, journals, and funders to share outbreak-related findings rapidly and openly. More than 30 leading publishers subsequently committed to making their coronavirus-related publications and supporting data immediately accessible in PubMed Central and other public repositories (Wellcome, 2020). This exception is revealing in two ways: it concedes that gated, verified information carries a real public cost, yet it is limited by topic and time, applying to outbreak-related research only for the duration of the emergency. Even then, just over half of the existing coronavirus literature was closed access when the outbreak began (Larivière et al., 2020). The luxury classification is suspended only when harm becomes visible at scale; the everyday harms of a news desert or an unaffordable article never trigger the exception.

This stratification erodes foundational media literacy. Because free sites often mirror the interfaces, metadata tags, and presentation styles of professional news organizations, the

average user struggles to distinguish peer-reviewed, human-gathered reporting from synthetic, pay-for-play public relations.

Taken together, the evidence in this report suggests a three-tier structure rather than a simple two-tier split. The first tier is the verified record itself, available to subscribers and institutionally affiliated researchers. The second is derivative access: headlines, snippets, shared screenshots, and free AI summaries that point to the verified record without delivering it—and that fail in a substantial share of cases (European Broadcasting Union & BBC, 2025). The third is the subsidized and synthetic layer produced by the Mass Distribution Loop. Movement between tiers depends less on need or curiosity than on income and institutional affiliation.

Figure 13, The "Luxury Tier" of Truth

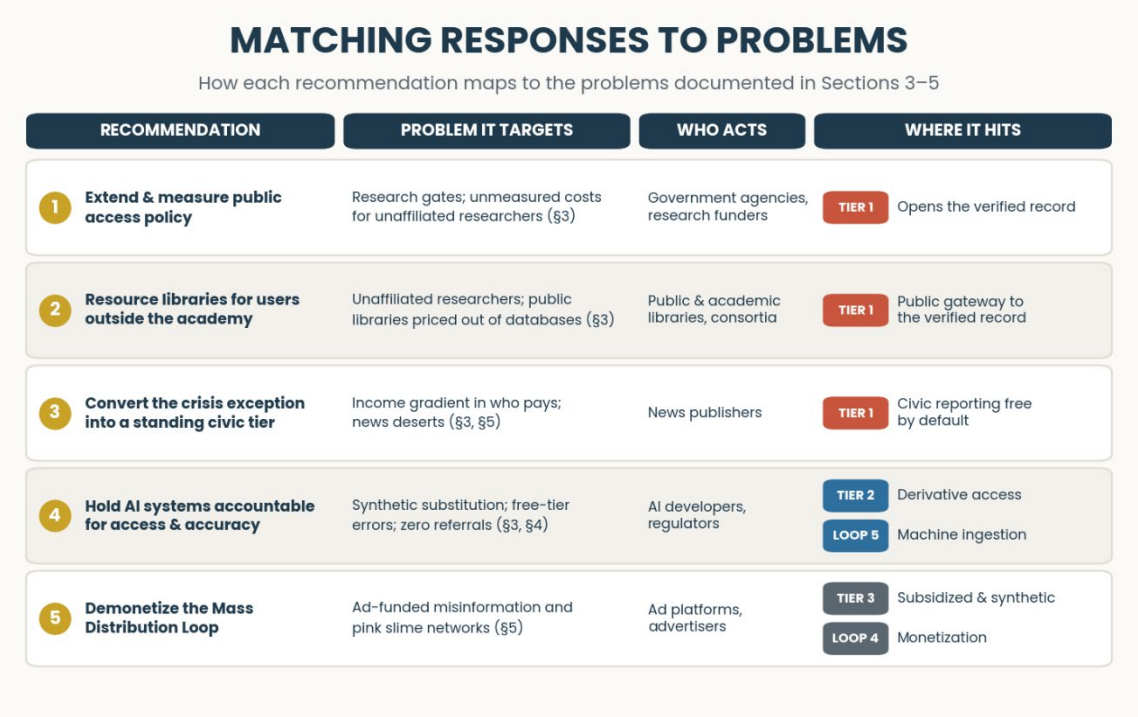


Classifying truth as a luxury good inverts the logic on which much of the public information infrastructure was built. Public libraries, public broadcasting, and public access mandates all rest on the premise that an informed public is a shared good rather than a private purchase. The paywall does not abolish that premise, but it shifts the burden of honoring it from institutions to individuals, leaving those least able to bear it most likely to be left in the lower tiers.

6. Recommendations

No single intervention will resolve the asymmetries documented in this report. The evidence does, however, point toward five complementary directions. Each one targets a different point where the current system sorts readers by ability to pay.

Figure 14, Matching Responses to Problems



Extend and Measure Public Access Policy

The 2022 OSTP memorandum and NIH's 2025 implementation are important corrections, but their scope remains limited. They apply only to federally funded research and require authors to make accepted manuscripts—not final published versions—available. They also do nothing for the substantial body of research published before the policies took effect or for work funded by private, state, or international sources. Future policy should extend immediate public access to research supported by state and philanthropic funders. It should also encourage depositing final published versions whenever publisher agreements permit.

Policy on this scale also requires evidence. No baseline data show how much unaffiliated and independent researchers spend accessing the literature, or how often paywalls delay or halt their work. Without that data, policymakers cannot assess whether open access efforts are reaching the taxpayers they are intended to serve.

Resource Libraries to Serve Users Outside the Academy

Public libraries are the natural gateway for researchers without an institutional affiliation. Yet many cannot afford the specialized scholarly databases those researchers need. Efird-Green et al. (2024) show that institutions can deliberately redesign access for unaffiliated users, including by vetting whether a resource is genuinely usable without an institutional login. That approach could be scaled through statewide consortium licensing, community borrower and alumni access programs at academic libraries, and dedicated funding for public library database subscriptions. Libraries are also well positioned to collect the usage and demand data that would make unaffiliated researchers visible to publishers and policymakers.

Convert the Crisis Exception into a Standing Civic Tier

Publishers already lower their paywalls when the harm of gated information becomes visible at scale, such as during elections, natural disasters, and public health emergencies. That exception could become a standing policy. Reporting on elections, public health, and public safety would be freely available by default rather than through emergency declarations. This preserves subscription revenue for the bulk of a publisher's output while removing the barrier to the information citizens most need to participate in public life.

Pricing structures can also expand the paying audience beyond its current ceiling. A fifth of non-paying Americans have expressed interest in bundled access to multiple news providers at a reasonable price (Reuters Institute for the Study of Journalism, 2025). Bundles, income-based pricing, and library-mediated news access all offer ways to lower the toll without sacrificing reader revenue.

Hold AI Systems Accountable for Access and Accuracy

AI developers now stand between many readers and the verified record, and they should share both the costs and the accountability. That means licensing and compensating the publishers whose work their systems summarize. It also means providing accurate attribution with working links to the original sources. And it means holding free-tier products to the same accuracy standards as paid ones. As the EBU and BBC evaluators argued, access to accurate news should not depend on whether a user pays for an AI assistant (European Broadcasting Union & BBC, 2025). Regular, public reporting of error rates by product tier and language would enable users, publishers, and regulators to judge whether these systems are improving.

Demonetize the Mass Distribution Loop

The economics of misinformation must be addressed alongside the economics of journalism. Ahmad et al. (2024) found that most misinformation websites are sustained by advertising and that most company decision-makers are unaware that their ads appear on such sites. Greater transparency would weaken the Mass Distribution Loop at the point where it converts reach into revenue. Advertising platforms could disclose where programmatic ads are placed, and advertisers could have easy tools to exclude sites that repeatedly publish false information. Reducing the subsidy for falsehood narrows the gap between the free and verified tiers from the other direction.

Conclusion

The paywall is neither a villain nor a solution. It is a rational response to a broken market. As Section 2 showed, the collapse of advertising-supported publishing left news organizations and scholarly publishers with few alternatives to direct reader revenue. The authentication barrier delivered on its promise: predictable income, cleaner first-party data, and insulation from search volatility and AI-driven traffic loss. For the individual publisher, gating content is often the difference between survival and closure. Any honest assessment of paywalls must begin there.

The central finding of this report, however, is that a strategy rational for each publisher produces an irrational outcome for the information ecosystem as a whole. The authentication barrier sorts readers by ability to pay rather than by need. Only about 17% of U.S. adults pay for news, and payment tracks closely with income, education, and race (Pew Research Center, 2025). News deserts now cover 213 U.S. counties, leaving roughly 50 million Americans with limited or no access to local news (Medill Local News Initiative, 2025). Scholarly publishing shows the same pattern in a more mature form. Research funded largely by the public is sold back to the public at margins approaching 40% (RELX, 2026), while unaffiliated researchers pay per article, rely on informal workarounds, or abandon their work. Across news, local information, and scholarship, the burden of exclusion falls hardest on those with the least institutional support.

Generative AI has sharpened this asymmetry rather than resolving it. Large language models and retrieval-augmented systems reconstruct and summarize gated content, eroding the traffic and revenue that paywalls were meant to protect. At the same time, they offer excluded readers a free substitute that is incorrect or poorly sourced in a substantial share of cases (European Broadcasting Union & BBC, 2025). Meanwhile, subsidized and ad-funded content moves through the Mass Distribution Loop without friction. It is produced at near-zero cost, amplified by engagement-ranked platforms, monetized through programmatic advertising, and ingested by the same AI systems that authoritative outlets increasingly block. The result is the three-tier environment described in Section 5. The verified record sits behind the gate, derivative fragments circulate just outside it, and a subsidized synthetic layer fills the space beyond. In that environment, verified knowledge functions as a luxury good.

The publishing industry has already acknowledged, in practice, that this arrangement carries a public cost. When elections, natural disasters, and the COVID-19 pandemic made the harm of gated information visible at scale, publishers lowered their paywalls. The recommendations in Section 6 start from that acknowledgment. The everyday harms of a news desert or an unaffordable article accumulate just as surely as the harms of a crisis, only more quietly. Responses should not wait for an emergency.

This report has limitations. It synthesizes secondary sources rather than presenting an original empirical study. Some figures describing AI crawling and content reconstruction come from industry sources, including technology vendors with commercial interests in the problems they describe; the text identifies these, and they should be treated as estimates rather than measured findings. It draws heavily on U.S. data, and the AI landscape it describes is changing faster than the literature can follow. Several of its claims about the costs borne by unaffiliated researchers rest on inference from affiliated populations, precisely because those costs have never been measured directly. Future research should fill that gap. It should quantify what unaffiliated and independent researchers spend, how paywalls delay or end their work, and how free-tier AI tools perform for users who have no paid alternative.

Paywalls emerged to preserve the institutions that produce verified knowledge, and in many cases they have. But a society that preserves its sources of truth by pricing most of its members out of them has not solved the problem of an informed public. It has privatized it. The toll for truth is now being paid by the people least able to afford it, in money, in time, and in the quality of the information they are left with. Deciding who should bear that cost is ultimately a public question, not a market one.

Bibliography

Ahmad, W., Sen, A., Eesley, C. E., & Brynjolfsson, E. (2023). Companies inadvertently fund online misinformation despite consumer backlash. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4514749

Ahmad, W., Sen, A., Eesley, C., & Brynjolfsson, E. (2024). Companies inadvertently fund online misinformation despite consumer backlash. *Nature, 630*(8015), 123–131. https://doi.org/10.1038/s41586-024-07404-1

Alyukov, M., Makhortykh, M., Voronovici, A., & Sydorova, M. (2025). LLMs grooming or data voids? LLM-powered chatbot references to Kremlin disinformation reflect information gaps, not manipulation. *Harvard Kennedy School Misinformation Review, 6*(5). https://misinforeview.hks.harvard.edu/article/llms-grooming-or-data-voids-llm-powered-chatbot-references-to-kremlin-disinformation-reflect-information-gaps-not-manipulation

Bilton, R. (2018, February 1). Learning from the New Yorker, Wired's new paywall aims to build a more "stable financial future." *Nieman Journalism Lab*. https://www.niemanlab.org/2018/02/learning-from-the-new-yorker-wireds-new-paywall-aims-to-build-a-more-stable-financial-future/

Bishop, D. (2020, December 1). Nature's OA fee seems outrageously high – but many will pay it. *Times Higher Education*. https://www.timeshighereducation.com/blog/natures-oa-fee-seems-outrageously-high-many-will-pay-it

Bohannon, J. (2016, April 29). Who's downloading pirated papers? Everyone. *Science, 352*(6285), 508–512. https://doi.org/10.1126/science.352.6285.508

Castro, Andres. (2026, May 31). *Securing AI citation integrity and attribution accuracy to prevent misleading generative reports*. https://andresseo.expert/geo/securing-ai-citation-integrity-attribution-accuracy/

Concept47. (2019, April 19). Are online paywalls helping the proliferation of Fake news? *Medium*. https://medium.com/@concept47/are-online-paywalls-helping-the-proliferation-of-fake-news-93049b8cb833

Coronium.io. (2026, April 7). *The AI crawler war: 50 billion daily bot requests reshaping web scraping (2026)*. https://www.coronium.io/blog/ai-web-scraping-crawler-war-2026

DeepDyve. (2026, June 29). *Accessing research papers without a university: The 2026 guide*. https://blog.deepdyve.com/2026/06/29/accessing-research-papers-without-university/

Duke University Libraries. (2018, September 5). *Paywalls and information costs*. Library 101 Instructors. https://sites.duke.edu/library101_instructors/2018/09/05/paywalls-and-information-costs/

Efird-Green, L., Marion, E., Willis, D., Gierisch, J. M., & Corsino, L. (2024). Designing and implementing a community-engaged research e-library: A case study for adapting academic library information infrastructure to respond to stakeholder needs. *College & Research Libraries, 85*(7), 994–1005. https://doi.org/10.5860/crl.85.7.994

Pelcro. (2025, August 5). How AI chatbots are reconstructing paywalled content (and what publishers can do about it). *Pelcro*. https://www.pelcro.com/blog/how-ai-chatbots-are-reconstructing-paywalled-content

European Broadcasting Union & BBC. (2025). *News integrity in AI assistants: An international PSM study*. https://www.ebu.ch/Report/MIS-BBC/NI_AI_2025.pdf

Faculty Advisory Council. (2012, April 17). *Faculty Advisory Council memorandum on journal pricing: Major periodical subscriptions cannot be sustained*. Harvard Library. https://gantercourses.net/wp-content/uploads/2013/11/Faculty-Advisory-Council-Memorandum-on-Journal-Pricing-%C2%A7-THE-HARVARD-LIBRARY.pdf

Fletcher, R. (2024, February 22). *How many news websites block AI crawlers?* Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/how-many-news-websites-block-ai-crawlers

Galter Health Sciences Library. (n.d.). *NIH public access*. Northwestern University. https://libguides.galter.northwestern.edu/NIH-public-access

Howard, R. (2022, August 10). Three reasons why journalism paywalls still don't work. *Quartz*. https://qz.com/three-reasons-why-journalism-paywalls-still-dont-work-1849392226

Larivière, V., Shu, F., & Sugimoto, C. R. (2020, March 5). The coronavirus (COVID-19) outbreak highlights serious deficiencies in scholarly communication. *LSE Impact Blog*. https://blogs.lse.ac.uk/impactofsocialsciences/2020/03/05/the-coronavirus-covid-19-outbreak-highlights-serious-deficiencies-in-scholarly-communication/

MacDonald, S. (2024, May 15). What is a paywall? Definition, types, and examples. *Squarespace*. https://www.squarespace.com/blog/what-is-a-paywall

Medill Local News Initiative. (2025). *The state of local news 2025*. Northwestern University. https://localnewsinitiative.northwestern.edu/projects/state-of-local-news/2025/report/

NewsGuard. (2024, June 11). *Sad milestone: Fake local news sites now outnumber real local newspaper sites in U.S.* https://www.newsguardtech.com/press/sad-milestone-fake-local-news-sites-now-outnumber-real-local-newspaper-sites-in-u-s/

NewsGuard & Comscore. (2021). *Special report: Top brands are sending $2.6 billion to misinformation websites each year*. https://www.newsguardtech.com/special-reports/brands-send-billions-to-misinformation-websites-newsguard-comscore-report/

Nieman Lab. (2026, February). Most Americans don't pay for news and don't think they need to. *Nieman Journalism Lab*. https://www.niemanlab.org/2026/02/most-americans-dont-pay-for-news-and-dont-think-they-need-to/

Novo, L. A. B., & Onishi, V. C. (2017). Could Sci-Hub become a quicksand for authors? *Information Development, 33*(3), 324–325. https://doi.org/10.1177/0266666917703638

Olsen, R. K., Kammer, A., & Solvoll, M. K. (2019). Paywalls' Impact on Local News Websites' Traffic and Their Civic and Business Implications. Journalism Studies, 21, 197-216. https://doi.org/10.1080/1461670x.2019.1633946

Pareene, A. (2020, October 21). Liberals are losing the journalism wars. *The New Republic*. https://newrepublic.com/article/159825/liberals-losing-journalism-wars-local-news-paywalls

Pattabhiramaiah, A., Overby, E. M., & Xu, L. (2020). Spillovers from Online Engagement: How A Newspaper Subscriber's Activation of Digital Paywall Access Affects Her Retention and Associated Revenue. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3694089

Pattabhiramaiah, A., Sriram, S., & Manchanda, P. (2018). Paywalls: Monetizing Online Content. Journal of Marketing, 83, 19-36. https://doi.org/10.1177/0022242918815163

Pew Research Center. (2025, June 24). *Few Americans pay for news when they encounter paywalls*. https://www.pewresearch.org/short-reads/2025/06/24/few-americans-pay-for-news-when-they-encounter-paywalls/

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ, 6*, e4375. https://doi.org/10.7717/peerj.4375

RELX. (2026, February 12). *Results for the year to 31 December 2025*. https://www.relx.com/~/media/Files/R/RELX-Group/documents/press-releases/2026/results-2025-pressrelease.pdf

Reuters Institute for the Study of Journalism. (2025). *Digital news report 2025*. University of Oxford. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/dnr-executive-summary

Reuters Institute for the Study of Journalism. (2026). *Digital news report 2026*. University of Oxford. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2026/dnr-executive-summary

Ryaboy, M. (2023, November 14). Comprehensive reader's guide: Techniques for bypassing digital paywalls. *Medium*. https://medium.com/@mryaboy/comprehensive-readers-guide-techniques-for-bypassing-digital-paywalls-e431e15b3e64

Sadeghi, M., & Blachez, I. (2025, March 6). A well-funded Moscow-based global 'news' network has infected Western artificial intelligence tools worldwide with Russian propaganda. *NewsGuard Reality Check*. https://www.newsguardrealitycheck.com/p/a-well-funded-moscow-based-global

Segado-Boj, F., Martín-Quevedo, J., & Prieto-Gutiérrez, J.-J. (2022). Jumping over the paywall: Strategies and motivations for scholarly piracy and other alternatives. *Information Development*. Advance online publication. https://doi.org/10.1177/02666669221144429

Simon, F. M., & Graves, L. (2019). *Pay models for online news in the US and Europe: 2019 update*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/our-research/pay-models-online-news-us-and-europe-2019-update

Springer Nature. (n.d.-a). *Publishing options*. Nature. https://www.nature.com/nature/for-authors/publishing-options

Springer Nature. (n.d.-b). *Article processing charges*. Communications Engineering. https://www.nature.com/commseng/article-processing-charges

Stengel, R. (2024, March 18). Democracy dies behind paywalls. *The Atlantic*. https://www.theatlantic.com/ideas/archive/2024/03/democracy-dies-behind-paywalls/677785/

Streyczek, J. (2025). Political Effects of Newspaper Paywalls *. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5124655

The New Yorker. (2020, March 19). The fate of the news in the age of the coronavirus. *The New Yorker*. https://www.newyorker.com/news/q-and-a/the-fate-of-the-news-in-the-age-of-the-coronavirus

van Ess, H. (2025, February 10). How AI bots quietly dismantle paywalls via web search. *Digital Digging*. https://www.digitaldigging.org/p/how-ai-bots-quietly-dismantle-paywalls

Vasipenok, V. (2025, January 28). AI browsers are sneaking past paywalls: How do they do it? *Quasa News*. https://quasa.io/news/ai-browsers-are-sneaking-past-paywalls-how-do-they-do-it

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Walter, P., & Mullins, D. (2019). From symbiont to parasite: The evolution of for-profit science publishing. *Molecular Biology of the Cell, 30*(20), 2537–2542. https://doi.org/10.1091/mbc.E19-03-0147

Walters, W. H. (2019). Beg, borrow, and steal: Formal and informal access to the scholarly literature at U.S. master's universities. *The Journal of Academic Librarianship, 45*(6), 1–14. https://doi.org/10.1016/j.acalib.2019.102059

Wellcome. (2020). *Publishers make coronavirus (COVID-19) content freely available and reusable* [Press release]. https://wellcome.org/press-release/publishers-make-coronavirus-covid-19-content-freely-available-and-reusable

White House Office of Science and Technology Policy. (2022, August 25). *Ensuring free, immediate, and equitable access to federally funded research* [Memorandum]. https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

Williams, K., & Stroud, S. R. (2021, April 5). *The ethics of news paywalls*. Center for Media Engagement, University of Texas at Austin. https://mediaengagement.org/research/the-ethics-of-news-paywalls/

Zara, C. (2013, January 5). Paywalls, profits and pageview journalism: Ethics in the digital age. *International Business Times*. https://www.ibtimes.com/paywalls-profits-pageview-journalism-ethics-digital-age-994464