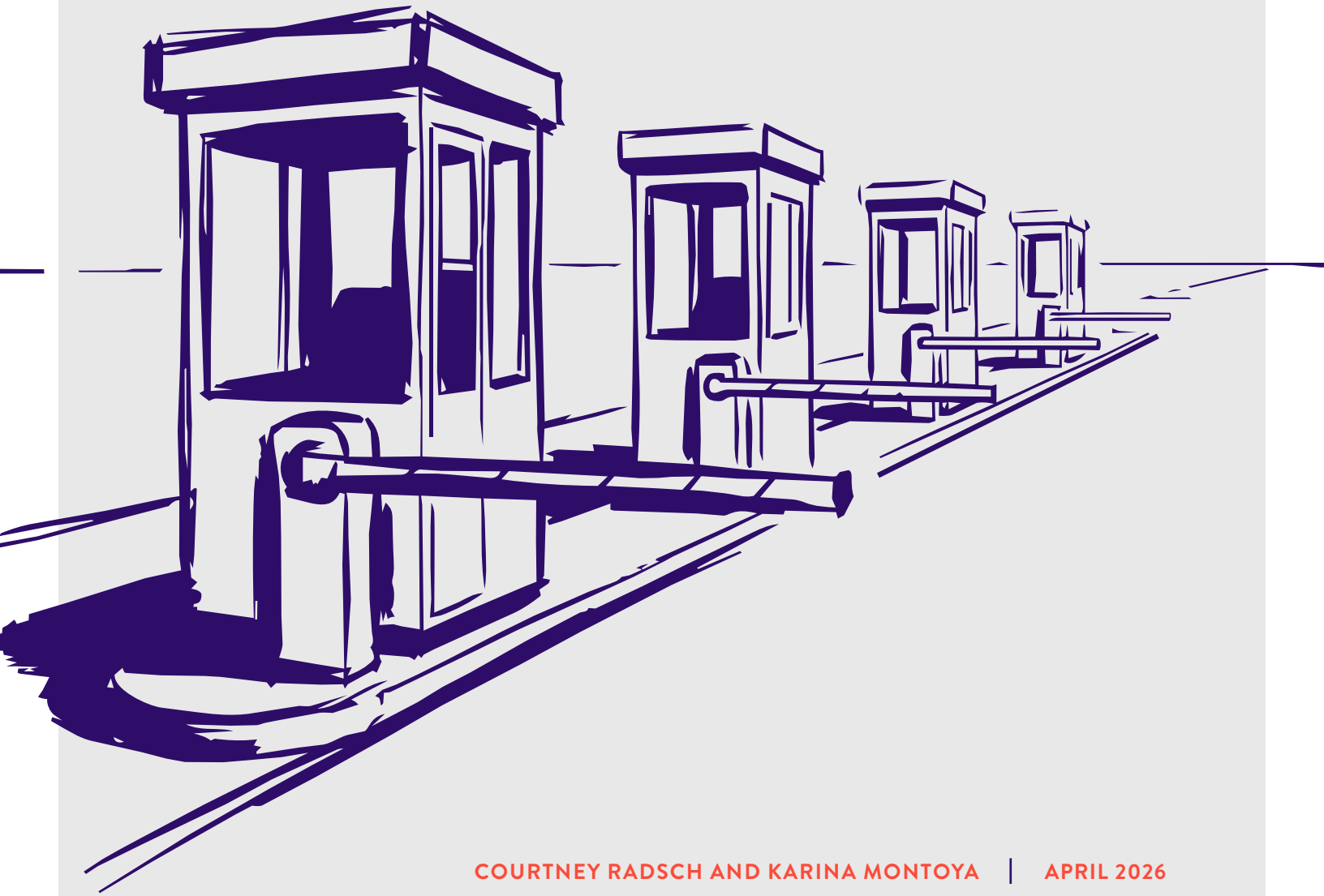
Same Gatekeepers, New Tollbooths

Mapping the AI Content Licensing Market



COURTNEY RADSCH AND KARINA MONTOYA | APRIL 2026

OPEN MARKETS INSTITUTE | CENTER FOR JOURNALISM & LIBERTY

Executive Summary

A market for AI content licensing is taking shape, but the conditions under which it is forming closely resemble those that made prior platform-journalism relationships structurally damaging to news. While a small number of large publishers, entertainment, and social platforms have secured direct bilateral deals with dominant AI companies, a startup intermediary ecosystem aimed at connecting AI companies with content and data, particularly from news outlets, is emerging. However, Big Tech corporations are encroaching into this fragile AI licensing market with a distinct value proposition.

The AI licensing market could provide an opportunity for a more level playing field, but it could also once again create dependency on intermediaries for monetization and access to audiences in the AI age. Furthermore, most of the compensation flowing through this nascent market reaches only the publishers with the brand recognition and legal infrastructure to negotiate directly or be included in marketplace pilots, leaving local, regional, foreign, ethnic, indigenous, and non-English language journalism largely untouched. Since tech firms have access to mountains of information including traffic and user behavior data that publishers do not, they are able to unilaterally define the terms of their discretionary and voluntary compensation for journalistic labor. Meanwhile, lawsuits against Big Tech corporations for copyright violations continue to accumulate, with more than 100 in the U.S. and at least six other countries, creating uncertainty in the market that benefits the biggest AI players.[1]

Three problems, each compounding the others, require policy attention. First, the prevailing framework for valuing publisher content by referral traffic from platforms such as Google and Meta is no longer feasible in the AI era when AI interfaces no longer send users to news sites. Furthermore, AI systems systematically undervalue what publishers actually contribute, repeating the foundational error of the search and social media era. Based on industry research and interviews this report finds that retrieval augmented generation (RAG) licensing and training data licensing are not separate markets; they are layers of the same value stack. Therefore, only pricing the retrieval layer while treating the underlying training model as a cost publishers have (involuntarily) donated for free misrepresents the economics of how AI systems are built and continue to develop. The second issue is that AI companies believe that agentic AI systems directed by a user should be akin to search and thus not covered by licensing or considered within the valuation system for content access or usage. Third, AI companies are selecting their own compensation counterparts (or rather subordinates), replicating a patronage dynamic that risks platform capture and coercion.[2] Given the current market structure in which confirmed illegal monopolies control search and advertising technology, a few large buyers currently dominate[3] the AI content market.

In today's world, publishers and content creators depend on Big Tech and, increasingly, AI corporations to produce, distribute, and monetize their content. By providing this infrastructure, platforms informally regulate media through their defaults around access, transparency, standards, and monetization.[4] As the first two decades of the 21st century underscored, voluntary commitments will not level the playing field without a public policy backstop.

This report maps the current AI licensing market, develops original analysis of full-stack valuation, and addresses ongoing infringement arguments that have not yet been adequately resolved in either the policy or commercial literature. The report concludes with recommendations for policymakers, publishers, philanthropy, and AI companies that could shift the market's trajectory before its structure becomes too entrenched to reform.

I. State of Play: A Second Reckoning

In the early days of search engines, publishers missed an opportunity to demand licensing frameworks when they first began indexing their content since there appeared to be a relatively fair value exchange. Similarly, when social media platforms began distributing journalism to build their own audiences, publishers could have insisted on revenue floors before traffic became indispensable. In both cases, urgency, optimism, and legal regulatory disparity led the journalistic media industry to accept structurally unfavorable terms that proved impossible to revisit once platforms achieved dominance throughout the tech stack.

The emergence of generative AI presents the same dynamic, but at a scope and scale that are more existential. This ingestion of effectively the entire accessible internet—decades of journalism, investigation, reported fact, and empirical analysis—into training datasets without consent, compensation, credit, or even click-throughs is shaking the industry.[5] AI systems collectively account for well under one percent of total publisher traffic, while traditional search engines and, in particular, Google continue to drive the overwhelming majority of external referrals. Generative AI products are designed to synthesize and present answers directly within an interface, obviating the need for users to click through to the underlying material even when sources are cited.

More than 100 copyright lawsuits have been filed against AI companies as of early 2026, and the U.S. legal system—where more than 90 percent of the lawsuits have been filed—has yet to decide on the legality of content extraction. What is not in dispute is the economic asymmetry: AI companies are among the most highly valued enterprises in history, built substantially on labor and data for which the contributors have received nothing or an amount that bears no relationship to what any fair market and legitimate accounting of contribution would suggest. Thus far, compensation has only been at the institutional level, meaning that individual journalists or creators are largely left out of the market.

The deal structures, price precedents, intermediary take rates, and governance norms taking shape now will be difficult to revise once they are normalized. The question of whether publishers, journalism, or creators of any sort can make a credible collective claim before market structures crystallize will not stay open indefinitely.

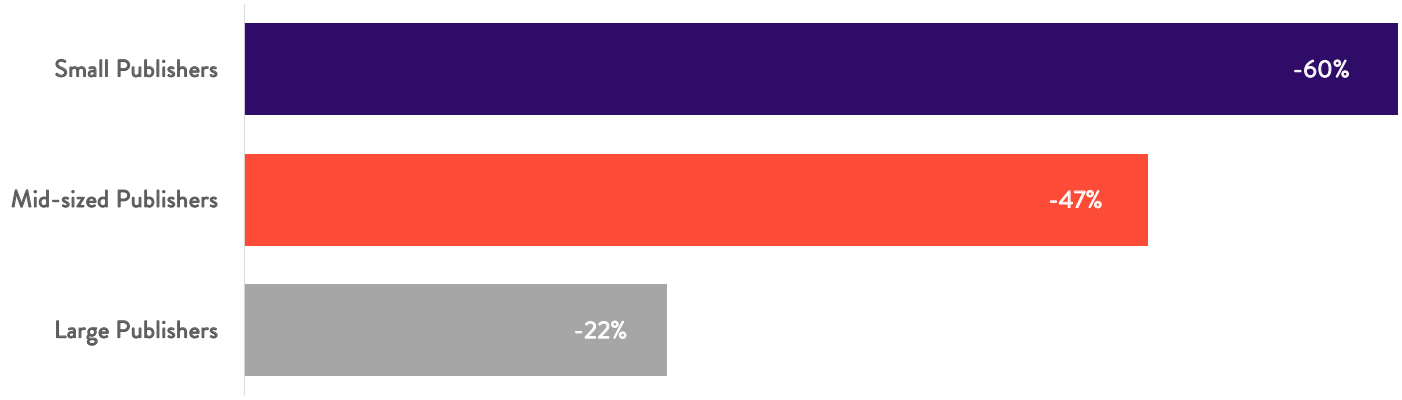


Fig 1: Search Referral Traffic Lost, 2023–2025 — Decline in Referrals from Traditional Search Engines, by Publisher Size

Source: Center for Journalism & Liberty / Open Markets Institute. Created with the assistance of Claude AI (Anthropic).

II. State of the Market: Who Is Getting Paid, and Why It Is Not Enough

THE BILATERAL DEAL TIER

The most visible portion of the AI content licensing market consists of confidential bilateral agreements between major AI companies such as OpenAI and Perplexity and select news publishers with large, national, or international presence, and sizable portfolios of specialized content.[6] OpenAI has signed deals with approximately 35 publishers; Microsoft's Publisher Content Marketplace launched with eight invited publishers; while Perplexity's publisher program covers around 20 outlets. News Corp's reported five-year agreement with OpenAI valued at roughly $250 million is one of the biggest. DotDash Meredith receives at least $16 million annually from OpenAI. Thomson Reuters reported $33 million in year-to-date AI licensing revenue. Amazon agreed to pay the *New York Times* approximately $20 million annually. The creative industry is experiencing a similar dynamic as large entertainment companies and record labels, such as Walt Disney, make deals with OpenAI and other big AI players,[7] leaving individual independent creators out in the cold.

Beyond headline numbers, the underlying deal terms are confidential. However, based on interviews and analysis conducted by Center for Journalism & Liberty at Open Markets, most include a combination of flat-fee payments and credits for newsrooms to use the company's AI products and cloud services. With most of the data for AI training already scraped, many publishers fear they will not be able to retroactively obtain compensation. As such, many of the deals are reportedly transitioning to seeking compensation based on how frequently AI scrapers access data for inference purposes.[8]

The most active commercial negotiations in the AI licensing market are focused on RAG, which enables chatbots or search engines to generate more accurate and real-time responses in response to user queries, occluding price discovery across the tech stack or insight into the tying or bundling of other services. AI systems use RAG to retrieve and incorporate relevant or up-to-date information that is not contained within the model parameters, rather than relying only on what the model learned during training. Without access to accurate and updated information to integrate with their large language models, "answer engines" and chatbots would be much less useful.

Bilateral agreements have not yet provided the referral traffic protections publishers may have expected. Data tracking publisher outcomes across 2025 shows that publishers with direct AI licensing agreements initially enjoyed a substantial click-through advantage from AI interfaces, according to analysis by AI licensing firm TollBit. Early in the year, their referral rates were roughly an order of magnitude higher than those of publishers without deals. But that "deal premium" eroded rapidly by the fourth quarter of 2025 amid a sixfold decline in click-through rates from 8.8 percent to 1.3 percent. Publishers without deals fared worse in absolute terms but experienced a smaller proportional drop — from 0.8 percent to 0.27 percent.[9] In other words, both groups lost referral traffic over the course of last year, with the available evidence suggesting that bilateral licensing agreements did not insulate publishers from the broader erosion of AI-driven referrals.

The strategic implications for publishers are significant. If licensing agreements do not meaningfully preserve referral traffic much less the accompanying advertising and subscription revenue, it undermines the commercial rationale for pursuing one-off bilateral negotiations. Furthermore, bilateral deals are not scalable and are confined almost entirely to publishers with sufficient brand recognition and legal capacity to attract AI company interest, which is but a fraction of the market nationally, much less globally. Notably, as of January 2026, 88 percent of top-ranked U.S. news outlets were blocking AI crawlers.[10]

Although bilateral licensing agreements may still provide important revenue streams, as AI interfaces become a primary gateway to search for and access information, the policy debate must shift focus towards broader questions of market structure, bargaining power, and data governance.

THE INTERMEDIARY TIER

Several AI licensing startups have emerged, primarily targeting news publishers, though some also serve book authors and creators in adjacent content industries. The field has expanded from a handful of Silicon Valley startups to more than a dozen intermediaries, including several large tech firms, since 2024.

There is no single model used by these licensing companies, reflecting genuine uncertainty about which mechanisms will prove most durable and giving rise to two categories of platforms. The first operate as intermediaries between publishers and AI developers, creating marketplaces for licensed content. This includes companies such as TollBit, Sphere.ai, ScalePost, Created by Humans, and Defined.AI, as well as Microsoft and Cloudflare. The second are publisher-side AI providers, enabling publishers to implement AI-based experiences, deploy AI-powered search and personalization on their own properties while retaining audience relationships, and identify how their content is used in AI-generated results. This is a key element of ProRata's offering, for example. Miso.ai allows publishers to create their own AI search engines that draw solely on their own content corpus, keeping the audience relationship within publisher-owned interfaces.

On the demand side, these intermediaries partner with large AI developers seeking to refine their AI-generated output and obtain better quality data for AI model training. Most of them focus on intermediating data for RAG (TollBit, Sphere.ai, ScalePost, and Cloudflare), with a subset of them (ScalePost, Created by Humans, and Defined.AI) offering data for both RAG and AI model training. The advantage they offer to AI developers is access to a more transparent pipeline of quality, licensed data sources — in lieu of having AI bots crawl the entire web, increasing the risk of tapping into inaccurate or poisoned data.[11] In turn, this also reduces their exposure to copyright infringement. Additional services, such as collection and annotation of field-specific data for AI training, are provided by companies such as Defined.AI.

A key advantage for publishers connecting to these marketplaces seems to be that they access better insights on how their content is used. "We've noticed that with a lot of the big one-on-one deals, they encourage publishers to just throw their data in [a cloud storage] bucket, and do not give any visibility into the analytics [of data use]. But we're trying to give more visibility and control about how AI companies are working with content," said Olivia Joslin, co-founder and COO of TollBit. In that sense, TollBit generates audits for publishers to see the type of content accessed per licensing term. Founder and CEO of ScalePost, Ahmed Malik, described the company's data analytics in similar terms in an interview with CJL: "We've built a pretty in-depth tool for helping [publishers] clearly see what's important to these AI chatbots — what parts of the content, which categories, which authors, what traffic they're getting, and what they can do to get more clicks."

AI licensing intermediaries offer at least three distinct products, often overlapping within individual companies: 1) bot detection, blocking, and paywall redirection; 2) content marketplaces with pay-per-use or pay-per-access pricing, and; 3) attribution-based revenue distribution.

Bot detection, blocking, and paywall redirection is the most operationally immediate service. TollBit's bot paywall — which intercepts AI crawler traffic and redirects it to a payment interface rather than allowing free access — saw bot traffic directed through it increase by 732 percent from Q4 2024 to Q1 2025. Cloudflare, operating at a different scale entirely by virtue of managing approximately 20 percent of global internet traffic, has extended analogous functionality to its entire publisher network by making AI crawler blocking the default. These tools address the consent problem directly, creating technical enforcement where robots.txt, the primary voluntary mechanism publishers use to restrict AI access, provides only a social norm. Describing robots.txt as a gentleman's agreement, TollBit's Joslin said, "It's a suggestion or a guideline, but it's not active enforcement. That's why we developed the TollBit bot paywall: to actually enforce it a little bit more."

Content marketplaces with pay-per-use or pay-per-access pricing are the most commercially active model and are significant for a structural reason that bilateral deals obscure: they allows access to the long tail of the AI industry — the small startups, domain-specific tools, and emerging AI businesses that cannot afford a flat-fee archive license but can pay per query. Sphere.ai, ScalePost, and Created by Humans all operate variants of marketplace platforms that connect publishers with AI developer buyers on a transactional rather than bilateral-deal basis. Methods to charge AI developers include pay-

per-use (based on appearances in results) and pay-per-crawl (based on access to the content) models, but the market has yet to settle on a standard.

Pay-per-use is akin to the idea of micropayments,[12] which have long been floated in the industry as a potential revenue source for media but are considered unworkable and suboptimal.[13] However, this model seems to be finding its place in the AI licensing world, with Microsoft adopting pay-per-use,[14] as have TollBit, Sphere.ai, and ScalePost, while Cloudflare has adopted pay-per-crawl.[15] "You have this pay-per-use model in our marketplace, and this has been super popular with AI applications — small teams, five to ten million dollars raised, probably two to five people on their team," said ScalePost's Malik, describing the demand profile that marketplaces are servicing. No bilateral licensing framework reaches these buyers, a significant and growing portion of the AI economy.

Cloudflare, which manages roughly 20 percent of global internet traffic,[9] began blocking AI crawlers by default for new domains in July 2025. The company launched a pay-per-crawl marketplace, in which Cloudflare appears likely to charge a 20 to 30 percent fee, or take rate, to the publisher.[16] "I believe in journalism, and I believe that the health of local, independent news is essential for a healthy Internet and a healthy society," Cloudflare CEO Matthew Prince said when he announced a free program to help independent media and non-profits control access from AI bots.[17] He was not the first to observe that content creators need incentive to keep creating. Although the default blocking is available to all its customers, including free services to publishers and NGOs around the world,[18] its marketplace partners currently include only major news brands.

Both small and large companies have proposed several ways of pricing content, including path-based pricing, which would charge different rates based on the section of a website used, for example premium investigative reporting versus general archives. Others have proposed purpose-based rates determined by the intended use of the data, including training (bulk ingestion), inference (answering real-time queries), or search (appearing in AI-generated results) and/or the type of user (e.g. commercial versus non-profit or research). Prince has suggested a shift towards "knowledge-advancement-based compensation," where content is valued based on its contribution to specific AI knowledge domains rather than solely on its raw traffic.

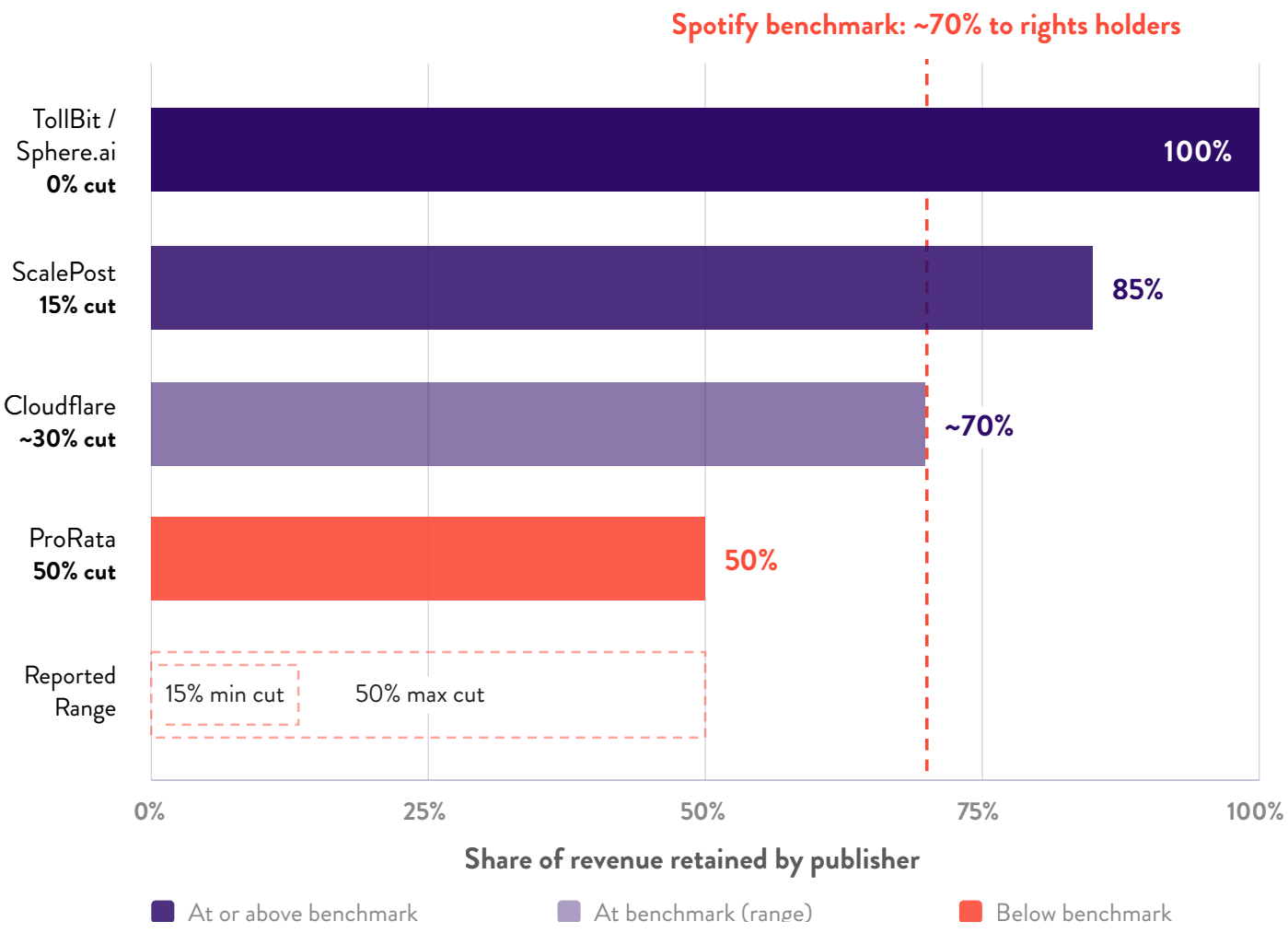
Attribution-based revenue distribution is a technically ambitious model with the most direct implications for the long-term structure of the market. AI search engine Perplexity and startups like ProRata use this approach. Following criticism and lawsuits, Perplexity incorporated attributions in its chatbot, launched a pilot Publishers Program in June 2024 that included advertising revenue-sharing for a handful of major publications and WordPress.com, and began collaborating with ScalePost.ai to provide publishers with deeper insights.[19] AI startup ProRata has built patent-pending attribution technology that analyzes AI-generated outputs, measures the proportional contribution of each source, and routes compensation accordingly. The company reports attribution certainty of 98 percent across its licensed content network and launched its own answer engine, Gist, on which it will sell advertising. "Our product is, in its core, an attribution platform. We create an answer, we dissect it with real RAG attribution, and then we divide the money between the attributed sources," ProRata's cofounder Josh Freeman told CJL. "Somebody needs to pay the journalistic salary." The company's 50/50 revenue share model splits subscription and advertising revenue equally between the platform and publisher, going far beyond prior models in the industry. As one of the early entrants, and given its explicit focus on creating its own advertising-based product, ProRata says it has already attracted more than 500 publication partners in the U.S. and Europe as of early 2026.

Big Tech Market Power: Amazon, Microsoft, Google, and Apple

Big Tech firms like Amazon, Microsoft, Google, and Apple are also entering the licensing market. Their encroachment risks producing a structural realignment over who controls access to premium content at the infrastructure layer, not because of contractual exclusivity but by capture through platformization. Microsoft's Publisher Content Marketplace, launched in February 2026 with an initial cohort of major U.S. media brands,[20] provides a structured, pay-per-use model to compensate publishers for content used by Microsoft's AI products.[21] It started the program with Copilot, an AI assistant embedded in Edge and Windows, thus guaranteeing a buyer and

Fig 2: What Publishers Actually Keep: Intermediary Take Rates Against the Spotify Benchmark

Source: Center for Journalism & Liberty / Open Markets Institute. Created with the assistance of Claude AI (Anthropic).



stimulating demand in this marketplace. Unlike licensing startups TollBit or Sphere.ai, Microsoft's Publisher Content Marketplace is a private two-sided platform built to feed its own AI products but also potentially available to other AI systems since no exclusive arrangement bars publishers from dealing with other buyers.[22] Amazon's reported marketplace plans operate on similar terms. The opportunity for publishers could be significant given the "scale and built-in audiences with enterprise clients who rely on Microsoft 365 and Azure," according to Axios.[23]

However, openness at the transactional layer does not preclude structural capture at the platform layer. Standardization lock-in increases as publishers who build content operations around a dominant platform's technical specifications face real switching costs, as the Google search trial underscored regarding the monopolist's AMP protocol.[24] Data asymmetry remains acute as the AI platform accumulates transaction-level intelligence on pricing, usage, and content performance that neither publishers nor competing AI developers can match, effectively replicating the structural advantage Google exploited by running the publisher ad server, the ad exchange, and the demand platform simultaneously.[25]

What makes the dynamic particularly harmful to journalism is the double bind it creates. Google's Gemini and AI summaries, which automatically appear atop 60 percent of search results, absorb the informational value of publisher content without routing users to the source. That traffic erosion is precisely what is pushing publishers toward licensing revenue as a replacement — and now

apparently back into the arms of the same Big Tech giants whose AI products are destroying traffic. Licensing deals with individual publishers like Associated Press function largely as diplomatic arrangements that manage publisher relations and reduce litigation risk rather than as mechanisms proportionally compensating publishers for the value they create.[26]

Intermediaries serving both the supply and demand for data leads to structural dependency by content creators and publishers on one side and AI companies on the other. Public policy interventions designed to prohibit exclusive deals would leave this dynamic untouched. Therefore, regulatory attention is warranted on these platform operators in order to mitigate their data access advantages and ability to set de facto (and potentially coercive)[27] standards for an industry in which no independent standards yet exist.[28]

THE LONG TAIL LEFT OUT

Perhaps the most consequential structural feature of the current market is that it excludes local newspapers, regional broadcasters, ethnic and indigenous media, non-English language publishers, and specialized outlets serving communities without political or economic power. All are effectively absent from the AI licensing market. They lack the brand recognition required to attract bilateral deal interest and the organizational capacity and the right to collective bargaining, not to mention the technical and legal resources to engage meaningfully with intermediaries. Licensing startups are typically focused on signing up bigger publishers, whose large supply of data entices AI companies to use their licensing platforms.

The long tail are precisely the publishers whose effacement is likely to cause the greatest civic harm, as well as those that AI companies have the least commercial incentive to compensate voluntarily.[15] One startup, Sphere.ai, has identified local news as its intended niche because of its uniqueness and importance. The startup is looking at practices such as surge-pricing that could benefit small-scale content providers such as local news organizations. Sphere.ai cofounder Kyle Buckingham told CJL he believes there's a significant opportunity for local news publishers who are the sole content creators for their geographical areas. Noting that content that isn't repetitive can command higher rates on a per-view basis, Buckingham said, "Local news publishers have a significant opportunity to monetize unique content about their specific areas."

The distribution of AI licensing revenue is thus likely to show a steep power curve: a handful of large publishers capturing the overwhelming majority of value, while thousands of local and independent outlets receive effectively nothing. This is not a feature of how journalism's economic value is actually distributed; it is a feature of how market power is distributed. Local news has among the lowest scrape-to-referral ratios in TollBit's data, for example, meaning its content retains high click-through value but generates virtually no monetization.

THE INTELLIGENCE GAP THAT PRECEDES THE MARKET

Publishers, particularly lower-resourced ones, are operating blind in the AI era because they neither know which AI systems are accessing their content and at what frequency, nor whether those systems are honoring the restrictions publishers have set or what the commercial value of their content is to the AI companies that consume it. Intermediaries have found a commercial foothold by trying to address this information deficit.

While these new intermediaries provide value through analytics on content access, bot identification, and paywall mechanisms, it is increasingly unclear whether the nascent startup sector will be able to remain viable amid the entry of dominant market players who can bundle products from their various lines of business. This creates structural vulnerabilities.

Take rates of 25 to 50 percent of publisher earnings are high relative to the value added by intermediation, particularly since there is little transparency into why certain sources are selected for access or use over others. Since most startups are venture-backed, they are therefore exposed to acquisition by the same large technology companies from which they nominally protect publishers. This was the same dynamic that hollowed out the independent ad tech ecosystem over the previous decade and transferred value steadily upward as it concentrated the market into the Google-Meta ad tech duopoly.

Intermediaries thus reproduce familiar platform dynamics, like take rates and potential for capture, while tools implemented directly by a publisher on their own website, by contrast, preserve greater publisher control.

WHAT THE DATA SHOWS

Platform and third-party data document the scale of AI extraction relative to compensation. OpenAI's crawlers scraped approximately 1,700 pages for every human visitor sent back to publisher sites; Anthropic's ratio was approximately 73,000:1; Perplexity's was 369:1.[29] AI zero-click behavior collectively drove 0.04% of total external referral traffic to publishers, against Google's 85%,[30] leading to an estimated loss of roughly $2 billion in annual publisher revenue losses.[31]

AI bots' circumvention of robots.txt, the underlying voluntary protocol that has shaped web- scraping over the past three decades, is accelerating. The proportion of AI bot scrapes bypassing robots.txt instructions rose from 3.3% in Q4 2024 to 12.9% by March 2025, as AI companies modified their terms of service to assert that user-initiated requests override publisher restrictions.[32] If this is a deliberate strategy to skirt legal issues by claiming that user queries create implied consent to access content that publishers have explicitly restricted, then it requires public debate if not court sanction before being codified in code and practice.

THE COLLABORATIVE ALTERNATIVES

While the emerging AI licensing market is primarily driven by profit motives, various collectives and market actors are collaborating to propose alternative approaches through collaborative protocols. These protocols offer novel methods for structuring the future of AI content licensing and compensating publishers. Really Simple License (RSL)[33] is an initiative among major websites and media publishers, led by a steering committee that includes O'Reilly Media, Yahoo, Automattic, and some of the creators of the Really Simple Syndication (RSS) protocol.[34] It articulates a standardized rights-and-licensing machine-readable protocol that tells AI crawlers how they can use specific content for training and deployment and the amount they must pay. By treating data access as an economic transaction that can be handled at the protocol layer, RSL could give smaller publishers disintermediated routes to licensing.

From a publisher perspective, under this machine-readable protocol, "anyone who hosts content will be able to indicate not just whether the content can be scraped but how it should be attributed and, crucially, how much they want to charge for its use, either individually or as part of a coordinated group," reported New York Magazine.[35] The expectation behind the voluntary adoption of RSL is not necessarily that giants such as Meta or OpenAI suddenly start paying royalties to all websites, but rather that the movement to block unauthorized AI crawling, now with support from infrastructure-level companies such as Cloudflare, might limit access to data of such inferior quality that licensing become inevitable. Being able to scale permissions through RSL could reopen the pipes of data supply for Big Tech giants but this time on publishers' terms.

OpenMined, an AI startup launched by RadicalxChange, a community of creators, is building a free and open-source protocol for what they term network-sourced AI that would be implemented directly by publishers. While currently in the idea phase, OpenMined's protocol would be designed for publishers to "get their hands dirty and experiment with having greater control over the data they share with AI developers," said Matt Prewitt, CEO of RadicalxChange, who envisions that publishers themselves would be able to connect their own collection systems to the network to handle transactions.

THE GATEKEEPER & THE BENEFICIARY: WHO DECIDES WHO GETS PAID

AI companies are currently selecting which publishers receive compensation by choosing their own preferred counterparties through bilateral negotiations or invitation-only programs. The market that results from such bilateral arrangements concentrates compensation among publishers with strong brand value and legal departments while structurally excluding most of the information ecosystem.[36] The rationale behind many partnerships is explicitly reputational, with AI companies citing association with trusted news brands as a means of building user confidence in AI-generated answers. Perplexity, for example, articulated as much when it launched its Publisher Program, saying, "Our ability to provide high-quality answers hinges on trusted, accurate sources covering the topics people care about most." When it expanded the program in 2025, Perplexity said it "would not be able to serve factual, valuable answers without news organizations continuing to report on different topics."[37]

The history of patronage in journalism whereby Big Tech platforms "partner" with participating media outlets to access their news for free, namely Facebook's Journalism Project and Google's News Showcase, is one of dominant platforms selecting beneficiaries on their own commercial terms, extracting legitimacy benefits from association with trusted journalism and then withdrawing support when business priorities shift. Although often referred to as "partnerships," the power dynamics and control exerted by the platform would more accurately be characterized as serfdom.[38]

The alternative is governance developed by journalism for journalism that embodies the self-regulatory ideal for press freedom and quality control. Press accreditation has a long institutional history as a mechanism for distinguishing professional journalism from other content based on editorial standards and public interest that functions independently of commercial relationships with technology companies. A publisher-eligibility framework for AI content compensation could be developed and administered by journalism-sector organizations based on independent eligibility criteria such as editorial standards and correction policies, original reporting as a primary function, transparent ownership and funding, and accountability to an independent complaints mechanism.

These criteria would cover much of legitimate journalism while foreclosing the patronage dynamic and minimizing the risk of capture.[39] Legislative models based on these criteria such as Australia's News Media Bargaining Code and Canada's Online News Act established the principle that eligibility is a public policy question,[40] not a commercial preference of the platform, and could be a template for the AI sector.

III. The Full-Stack Value Problem

As lawsuits work their way through the courts, current policy and market discourse treats AI content licensing as primarily a question of who gets paid when an AI system retrieves and uses publisher content in real time to answer a user query.

Pricing inventory is a key function of AI licensing marketplaces and since intermediaries have direct insight into AI developers' demand, they are able to provide various ways to set prices, based on factors such as pageviews, keywords, publication date (as newer content might be priced higher than older content), and type of AI model, among others. The most commonly used benchmark for publishers to assess the value of their content has been related to how much a pageview is worth in the "ad or subscription world," Sphere.ai's Buckingham said, noting these criteria could be too narrow.

Sphere.ai, for example, recommends prices based on an 'originality score' generated by its proprietary technology, which considers original reporting as a price-setting factor. Noting that the originality score takes into account the time and effort put into producing content, Buckingham said, "Original reporting, anything like that, we think is going to dramatically increase in price-per-view." In all cases, pricing can be adjusted over time, with some even envisioning "surge pricing," when demand for specific content spikes due to a natural disaster or other newsworthy event. TollBit's Joslin said, "We get really excited about things like surge pricing, identifying when a unique piece of content is highly in demand, and helping publishers optimize that revenue."

Intermediaries take a cut from publishers' earnings. Although most companies do not disclose their service fees, our interviews indicate they may range from 15 to 30 percent of collected revenue.

Created by Humans, which focuses on licensing books for both AI model training and RAG, enables book publishers and authors to license bundles of content or individual works. The company negotiates the agreements — which can either be approved or rejected by authors, who first have to register on the platform — and manages the transactions and payment disbursement to rightsholders.[41] Created by Humans takes a cut from authors' earnings, and authors are also free to license their works on other platforms.

Defined.AI, on the other hand, is focused on licensing a wide variety of datasets for AI developers, mostly geared towards AI model training in specific fields such as healthcare, translation and localization, voice AI training, and music production, to name a few. Defined.AI crowdsources its datasets from creators across the world, makes samples available to AI developers, and directly negotiates between AI developers and rights holders for the acquisition of licensed content for a fee.

The compensation logic that has governed most negotiations between platforms and publishers to date rests on a narrow conception of value: referral traffic. Yet just as a search platform would be less useful without journalism, so would answer engines, chatbots, and the underlying AI models themselves. This notion of the value of journalism to AI systems focuses on how the absence of referral traffic to publisher websites results in lost advertising revenue and the inability to convert audience into paid subscribers or members. According to the prevailing thinking, publishers should be compensated for that loss. Although this framing is valid as a measure of what publishers contributed to AI development, it is radically incomplete.

Accepting referral traffic, use, or even access as the basis for pricing reproduces the foundational mistake of the social media era, which was to allow platforms to define what journalism is worth. Publishers' contributions to AI development extend far beyond traffic, encompassing model training, fine-tuning, linguistic training, fact verification, current event information, and civic credibility.[42] Section VI addresses in greater detail what a fair valuation would include. However, in order to establish value, it is essential to have an accurate understanding of the economic structure.

RAG IS NOT A SEPARATE MARKET FROM TRAINING

The largest opportunity for recurring revenues for publishers is likely to be in retrieval augmented generation, not to mention where publishers might have the most immediate leverage. But RAG licensing cannot be fairly priced in isolation from the training problem and treating the two as separate markets produces compensation frameworks that structurally undervalue publisher contribution.

Every RAG response is generated by a foundation model whose capacity to synthesize information, construct coherent arguments, and produce readable prose was built with publisher content that was scraped without consent. Without the training foundation, the retrieval layer is useless. Pricing only the retrieval layer while treating the underlying model as a sunk cost that publishers have already absorbed for free is a category error that serves AI companies, not journalism.

The correct unit of analysis is the full AI value stack, which includes foundation model training, fine-tuning, RAG retrieval at inference, and the advertising revenues AI companies are beginning to generate from outputs built substantially on publisher content. Publishers have contributed value at every layer. The market currently envisions compensation at only the retrieval layer for a small subset of publishers, and on terms set by the AI companies.

THE MODEL IS ACTIVATED ANEW AT EVERY INFERENCE CALL

The copyright discourse around AI training has largely treated content-scraping as a discrete historical event. The scraping happened, the model was trained, and the legal question is therefore whether that one-time act constituted infringement, and whether the act falls under the statute of limitations. But with RAG, the model trained on publisher content is activated anew with every inference call – every query answered, every article summarized, every document drafted. Whether or not training data collection is ultimately found to violate copyright, the addition of RAG means infringement continues with every subsequent commercial deployment of the underlying model.

While the continuing infringement theory has been established in other copyright contexts,[43] it has not been adequately tested in the AI training cases now working their way through the courts. What makes this theory consequential across jurisdictions from the U.S. to Japan to the UK is that it rejects the framing preferred by AI companies, which is that data-scraping was a one-off historical act, the liability for which is fixed, fading, or has already been resolved through bilateral deals. The continuing infringement theory would extend both the relevant damages period and the revenue base for compensation well beyond historical training costs and does not confine it only to RAG uses. Publishers who accept current deal terms as final may be foreclosing more substantial claims that courts have not yet addressed, boxing themselves into a very narrow conception of value.

IV. The Legal Uncertainty and Its Market Consequences

The copyright status of AI training data is among the most consequential unresolved questions in the current legal landscape. More than 100 lawsuits against AI companies for violating copyright are making their way through courts around the world, mostly in the U.S. but with a handful in Europe, the UK, and India.[44] Despite guidance from the U.S. Copyright Office that AI training is likely not a valid fair-use exception,[45] the Trump administration has sought to void this assessment[46] and grant AI developers blanket protection[47] against copyright enforcement by claiming that "it's not doable."[48] In the absence of legal clarity and enforcement, AI developers continue to take data from publishers and creators on the open web and from behind paywalls.

Although some lawsuits have provided important but still narrow protections, others have not — maintaining a legal ambiguity that benefits Big Tech firms.[49] Even fewer lawsuits address the rights of creators and publishers, although Anthropic reached a reported $1.5 billion settlement arising from copyright litigation over its use of pirated books that will cover authors as well as publishers.[50] No appellate court has established clear doctrine, and resolution through litigation will take years.

This legal uncertainty, too, is asymmetric. Big AI companies benefit from delay: they are currently deploying AI models without paying for the training data, and they are building a userbase by offering their users indemnity for copyright infringement. Smaller AI companies can neither afford to offer such legal protections, nor the fees associated with mounting a legal defense, impeding their ability to raise venture capital or compete with Big AI. Publishers face pressure in the opposite direction, toward accepting current deal terms while they last, well before there is legal clarity and even when those terms undervalue their contribution and set a negative precedent. Below-market settlements and licensing arrangements struck under this pressure may prove structurally unfavorable once legal doctrine clarifies.

The continuing infringement theory provides some counterweight, however small, to the asymmetric power balance. If infringement is ongoing rather than merely historical, the damages period extends through current commercial revenues, substantially enlarging the claim and changing the legal calculus for continuing unlicensed scraping or access. Policymakers or the courts could require the creation of a revenue escrow mechanism that would require AI companies to deposit a percentage of inference revenues into third-party escrow during pending litigation. This would ensure that compensation would be available to rights holders if infringement were established or returned to AI companies if fair use were upheld.

Even in the best-case scenario, litigation cannot solve the transaction-cost problem of establishing individual publisher contribution values for specific AI systems at scale. This is the strongest argument for a statutory licensing regime with set rates. Policymakers could look to the U.S. compulsory mechanical licenses for music or the EU Copyright Directive's press publishers' right for inspiration, as several countries including Brazil and South Africa are already doing.[51] This approach sets compensation through a public administrative process and distributes it through a collective organization while removing the need to relitigate valuation for every publisher and every model, thus providing legal certainty for all parties.

V. What Fair Value Would Look Like

A compensation framework that genuinely reflects publisher contribution to AI development requires moving beyond the loss of traffic benchmark toward recognition that publishers contributed to a technology stack, not merely to a traffic source. That contribution has multiple dimensions. Publishers of all types provided the linguistic and reasoning training material on which models' coherence depends. High-quality journalism is the primary public mechanism through which facts about the world are regularly established and verified, while also providing the temporal currency that RAG systems depend on to respond to queries about current events. News publishers supply the civic credibility that AI systems inherit when they cite news sources.

Publishers' content provides linguistic and reasoning capacity (i.e. high-quality prose and argument structure, allowing models trained on it to communicate more coherently); factual grounding (the primary mechanism by which facts about the world are established and verified publicly); temporal currency (real-time journalism that RAG systems depend on to respond to queries about current events); and civic legitimacy (the credibility and trust that AI systems inherit when they cite news sources).

Establishing value is insufficient without the infrastructure to identify rights-based content and to aggregate rights holders. Developing a copyright identification system similar to YouTube's Content ID[52] at the LLM inference layer would identify when AI-generated responses draw substantially on specific publisher content and route compensation accordingly. ProRata offers a version, though with post-hoc attribution rather than internal model observation.

Mandating rather than merely incentivizing this type of infrastructure could provide a durable foundation for compensation while also mitigating challenges around information integrity and AI slop. A collective rights organization for AI content licensing would aggregate publisher bargaining power, enable collective price-setting, distribute revenue to the long tail, and provide ongoing governance as technology evolves, similar to how Broadcast Music Inc., or BMI, licenses and manages copyrighted works in music or as collecting societies do in publishing. Aligning long-term incentives between both sides could also be achieved through an equity or revenue-share component in licensing agreements that give publishers a stake in the AI company, though this is unlikely to benefit smaller publishers as much as larger ones if valuation is determined by quantity not quality.

Any compensation framework also needs to address the distinction between publishers and individual creators. Training data deals and RAG licensing intermediaries typically compensate corporate rights holders, not the individuals whose labor produced the underlying work. And it's not clear that creators want to get involved in this market: a survey by the Authors Guild found that only one-third of respondents expressed willingness to license their work through AI content platforms, reflecting significant adoption barriers even among creators aligned with the licensing model in principle. A framework modeled on equitable remuneration rights in European copyright law, which preserve individual creator rights to compensation even after copyright has been assigned to a publisher, would more accurately reflect the full distribution of human contribution.

VI. The Governance Gap: Why Markets Cannot Solve This Alone

An array of private market mechanisms is emerging amid a public policy vacuum in this rapidly developing and competitive AI market, leading to greater risk that companies with the most to gain from keeping content acquisition costs low will set the terms. In the absence of mandatory licensing regimes, collective bargaining rights, a regulatory framework for AI training data, or policies that equitably balance rights-holder interests over AI developer deference, the encroachment of Big Tech firms into AI licensing threatens to overwhelm the startups. Furthermore, the United States, where most dominant AI firms are based, does not favor copyright enforcement and is geared towards unfettering "innovation," while increasingly wielding its power to deter regulation the government disagrees with in other countries.[53]

Across the broader AI content market, the dynamics are oligopsonistic: a small number of powerful buyer companies negotiate against thousands of fragmented publisher sellers. Mandatory disclosure of aggregate deal terms, collective bargaining rights, and eligibility criteria that prevent platform selection of preferred partners are ways that could redress this imbalance. Cloudflare's potential 30 percent fee on its pay-per-crawl marketplace mirrors fee structures that have faced antitrust scrutiny globally; its structural position as an infrastructure gatekeeper warrants the same scrutiny, which has not yet materialized.[54]

Google, meanwhile, was found to operate illegal monopolies in both general search[55] and display advertising technology.[56] These rulings matter for AI licensing because Google's market position puts publishers in a state of structural dependency since refusing Google's crawlers risks search invisibility that few publishers can absorb. That same crawling infrastructure feeds AI products that undermine publisher revenue, yet in deciding remedies on Google's search monopoly last year, the D.C. District Court did not separate search indexing consent from AI training access.

Australia's News Media Bargaining Code and Canada's Online News Act make negotiation between platforms and publisher groups mandatory when bargaining power is severely asymmetric. Both frameworks have been criticized for their treatment of smaller publishers and their susceptibility to platform selection effects, but they are critical on two fronts. One, they acknowledge the anti-competitive dynamic inherent to the relationship between platforms and journalism, and two, they enshrine eligibility and compensation as public policy questions. The EU Copyright Directive's press publishers' right[57] provides a complementary mechanism operating through a different legal architecture, but the EU AI Act creates a text-and-data mining exception that is already being exploited by commercial developers.[58] A coherent U.S. framework does not exist, and its absence is increasingly anomalous given the size of the market and the accumulation of the legal claims around the question of AI content.

Conclusion

The conditions under which AI content licensing is developing now differ from the search and social media moments in several meaningful respects: the extraction is more visible, the legal claims are more mature, and the journalism sector has had time to observe how prior platform relationships developed. Whether that accumulated experience translates into collective action and effective policy before market structures settle around arrangements that serve AI companies at the expense of publishers, creators, and particularly journalism depends on choices being made in courtrooms, legislatures, newsrooms, and investor meetings over the next several years.

The answer will be determined not by any single lawsuit, deal, or regulatory ruling, but by whether the journalism industry collectively can make a unified case that its contribution to the AI economy deserves recognition not as a favor or as a product of platform goodwill, but as a matter of economic viability, legal obligation, and democratic necessity.

Glossary of Terms

AI Model / Large Language Model (LLM)	A computational system trained on data to recognize patterns, make predictions, or generate outputs such as text, images, audio, and video.
AI application	A use-specific implementation of artificial intelligence running atop AI models, such as image and video generators, answer engines, chatbots, or fraud detection tools.
AI Developers	Developers of AI models (Meta, OpenAI, Google, Amazon, Microsoft, Anthropic) and AI applications (AI model developers themselves, Perplexity, Harvey, Codeium, etc.).
Answer engines	Implementation of AI that uses search and natural language processing to deliver direct, concise answers by synthetizing multiple sources to respond to user queries, rather than providing a list of links to the sources likely to respond to the user queries.
Advertising technologies (ad tech)	The software and systems that support the planning, buying, delivery, targeting, measurement, and optimization of digital advertising.
Fine-tuning	The process of taking an already trained AI model and further training it on a smaller, specific dataset to improve its performance for a particular task, domain, or style.
Inference	The stage where a trained AI model generates outputs (such as text, answers, or predictions) in response to user inputs. Inference is when a model may draw on what it has learned from training or fine-tuning, but it does not directly access or retrieve licensed content unless additional systems (like RAG) are used.
Monopoly rents	The excess profits earned by a firm due to its exclusive or dominant control over a market, allowing it to set prices above competitive levels.
Retrieval augmented generation	A technique used during inference that combines a language model with a search or retrieval system, allowing the model to pull in relevant external information at the time of a query before generating a response. Systems that retrieve and connect an LLM with relevant external data (e.g., from search indexes) or proprietary data. RAG is a cost-effective way to update static LLMs with more timely, relevant, or domain-specific information. It improves accuracy and predictability and reduces the likelihood and prevalence of errors.

References

1 GWU AI Litigation Database, https://blogs.gwu.edu/law-eti/ai-litigation-database/

2 Radsch, "Platformization and Media Capture: A Framework for Regulatory Analysis of Media-Related Platform Regulations"; Papaevangelou, "Funding Intermediaries: Google and Facebook's Strategy to Capture Journalism."

3 This term refers to a market condition, oligopsony, dominated by a small number of large buyers who hold significant power over numerous sellers.

4 Radsch, C. (forthcoming). AI regulation as a form of media regulation. In Media Regulation. Eds Tambini and Kaye. Edgar Publishers.

5 Bruell, A. (2026, February 16). Marketplaces are the next frontier in publisher deals with AI companies. The Wall Street Journal.

6 Virtually all publishers with bilateral deals are based in the Global North with a few exceptions.

7 https://openai.com/index/disney-sora-agreement/

8 https://www.theinformation.com/articles/news-publishers-shift-ai-licensing-focus-usage-based-deals?stream=top

9 Evidence from other analytics firms points in the same direction. Traffic data reported by Chartbeat, Similarweb, and Parse.ly shows that referrals from generative AI interfaces remain extremely small relative to traditional search

10 TollBit. (2025). State of the bots: Q1 2025 — The rise of RAG bots. TollBit Resource Center. https://tollbit.com/state-of-the-bots. Figure of 88% of top-ranked U.S. news outlets blocking AI crawlers as of January 2026 from Radsch & Montoya (2026) analysis of publisher robots.txt deployments.

11 https://www.ibm.com/think/topics/data-poisoning

12 https://pressgazette.co.uk/paywalls/micropayments-for-news/

13 https://www.cjr.org/opinion/micropayments-subscription-pay-by-article.php; Radsch, URGENT: Understanding and Responding to Global Emerging News Threats.

14 https://digiday.com/media/publishers-scorecard-for-big-techs-ai-licensing-deals/

15 https://blog.cloudflare.com/introducing-pay-per-crawl/

16 Although a Cloudflare spokesperson would not disclose any information about how much the company will take as an intermediary, according to two people with direct knowledge and podcasts with Prince. https://www.youtube.com/watch?v=8W17UqoKpjw&t=169

17 https://www.cloudflare.com/press/press-releases/2025/cloudflare-helps-protect-independent-journalism-and-non-profits-from-ai/#:~:text="I%20believe%20in%20journalism%2C%20and,to%20the%20AI%2Ddriven%20web.

18 Cloudflare's Project Galileo provides free cybersecurity to organizations working in human rights, civil society, journalism, or democracy; Radsch was involved in raising awareness of these services when she such work at the Committee to Protect Journalists https://blog.cloudflare.com/ai-crawl-control-for-project-galileo/

19 https://www.perplexity.ai/hub/blog/introducing-the-perplexity-publishers-program

20 Among them are the Associated Press, Vox Media, USA Today, Condé Nast, and Hearst's People Inc.

21 https://digiday.com/media/people-inc-strikes-microsoft-ai-licensing-deal-as-googles-ai-overviews-hit-programmatic-ad-revenue/

22 Microsoft Advertising Blog, "Building Toward a Sustainable Content Economy for the Agentic Web," February 3, 2026. https://about.ads.microsoft.com/en/blog/post/february-2026/building-toward-a-sustainable-content-economy-for-the-agentic-web

23 https://www.axios.com/2025/09/23/microsoft-ai-marketplace-publishers

24 Montoya and Radsch, "Beyond Court Remedies in the Google Search Case: A Competition Reform for the Search Ecosystem."

25 The ad tech analogy refers to documented findings in U.S. v. Google (ad tech), E.D. Va. 2024, and the CMA Digital Advertising Market Study, 2020.

26 https://www.ap.org/media-center/ap-in-the-news/2025/google-signs-deal-with-ap-to-deliver-up-to-date-news-through-its-gemini-ai-chatbot/

27 Montoya and Radsch, "Beyond Court Remedies in the Google Search Case: A Competition Reform for the Search Ecosystem."

28 The Internet Engineering Task Force AI-controls group, of which Radsch is a member, is working to develop a consensus-based standard which could be used by policymakers as a benchmark or guideline. Despite the seeming advisability of a consensus-based self-regulatory process in which individuals participate not as formal representatives of their companies, there are power dynamics involved in who can and does participate in the IETF and these types of processes. While a discussion is out of scope for this paper, civil society and small publishers do not have the same resources in terms of time, money and capacity (either in terms of employees or consultants who can follow and participate in person or on a day-to-day basis).

29 Over 70 lawsuits figure from Tech Policy Press analysis cited in the AI Licensing Corpus Pattern Analysis (Radsch & Montoya, February 2026).

30 United States v. Google LLC, No. 1:20-cv-03010 (D.D.C. Aug. 5, 2024). Judge Mehta ruled Google holds an illegal monopoly in general search and general search text advertising. Google has appealed to the U.S. Court of Appeals for the D.C. Circuit; as of early 2026, the appeal and remedies proceedings are ongoing.

31 United States v. Google LLC, No. 1:23-cv-00108 (E.D. Va. Nov. 1, 2024). The court found Google liable for monopolization of the publisher ad server market and the open-web display advertising exchange market. Remedies have not yet been determined as of early 2026.

32 Cloudflare handles approximately 20% of all global internet traffic. Cloudflare. (2025, July). Cloudflare just changed how AI crawlers scrape the internet at large. https://cloudflare.net/news/news-details/2025/Cloudflare-Just-Changed-How-AI-Crawlers-Scrape-the-Internet-at-Large-Permission-Based-Access-Makes-Way-for-A-New-Business-Model/default.aspx

33 https://rslstandard.org/guide/what-is-rsl

34 https://en.wikipedia.org/wiki/RSS

35 https://nymag.com/intelligencer/article/ai-scraping-free-for-all-by-openai-google-meta-ending.html

36 Australia News Media and Digital Platforms Mandatory Bargaining Code, Broadcasting Services (Digital Platforms) Act 2020 (Cth), as amended. For an assessment of outcomes, see Competition Commission of South Africa. (2024). Media and digital platforms market inquiry: Day 7 public hearings transcript. https://www.compcom.co.za/media-and-digital-platforms-market-inquiry/

37 https://www.forbes.com/sites/johnwerner/2025/06/16/perplexity-gives-money-to-publishers/

38 Online News Act, S.C. 2023, c. 23 (Canada). The Act requires digital news intermediaries meeting defined thresholds to negotiate compensation agreements with eligible news businesses.

39 Radsch, Making Big Tech Pay for the News They Use.

40 Center for Journalism & Liberty, "Tech & Media Fair Compensation Global Tracker.":"Center for Journalism & Liberty, "Tech & Media Fair Compensation Global Tracker."",

41 https://www.createdbyhumans.ai/post/created-by-humans-is-now-live

42 Radsch, What Is the Value of Journalism to AI? A Framework for Establishing Journalism's Value in Artificial Intelligence Systems; Radsch, "AI Needs Us More Than We Need It."

43 Bulger, "US Supreme Court Holds No Limit to Number of Years for Which Copyright Infringement Damages Are Recoverable Under the 'Discovery Rule.'"

44 https://blogs.gwu.edu/law-eti/ai-litigation-database/

45 https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf

46 https://www.journalismliberty.org/publications/journalists-and-artists-lose-out-to-ai-corporations-as-trump-fires-copyright-director

47 https://www.theguardian.com/technology/2025/jul/25/trump-ai-action-plan

48 https://www.wired.com/story/president-trump-ai-action-plan-speech/

49 On the one hand, Reuters"`landmark win`against now-defunct competitor Ross Intelligence clarified Ross Intelligence violated the "fair use" doctrine by using Reuter's legal research database to build its own AI products, and Anthropic's $1.5 billion settlement could result in a class of book authors and publishers receiving $3,000 per work for copyright infringement. On the other hand, in the case of Andersen et al. v. Stability AI, where a group of artists accused Stability AI's image generators violated copyright law, the court dismissed the claims because the artists hadn't registered their works with the U.S. Copyright Office.

50 Bartz v. Anthropic PBC, No. 3:24-cv-02655 (N.D. Cal.). This is the largest publicly reported copyright recovery in U.S. history, compensated book authors and publishers whose works were pirated from shadow libraries used for AI training. The settlement covers approximately 500,000 pirated works at ~$3,000 each. Preliminary approval was granted September 2025; final approval is pending as of this writing. The settlement addresses only past training on pirated materials, not AI outputs or news publisher content. See NPR (September 5, 2025), https://www.npr.org/2025/09/05/g-s1-87367/anthropic-authors-settlement-pirated-chatbot-training-material

51 Carlini et al., IPD Working Paper: How to Update EU and US Copyright Regimes in the Age of AI; Radsch, "AI Governance as Media Regulation."

52 ContentID is a hashing technology used by YouTube to identify rights-protected works at the point of upload, though there are risks with such a filtering approach as well. See Radsch, Weaponizing Privacy and Copyright Law For Censorship.

53 Radsch and Von Thun, Defending Europe's Digital Sovereignty: Confronting Platform Power and Political Interference.

54 Google charged 30 percent app store commission and in Epic Games v. Google LLC, No. 3:20-cv-05671 (N.D. Cal.), a jury found in December 2023 that Google maintained an illegal monopoly over the Android app distribution market. Radsch, "Cloudflare Wades into the Battle Over AI Consent and Compensation."

55 United States v. Google LLC, No. 1:20-cv-03010 (D.D.C. Aug. 5, 2024). Judge Mehta ruled Google holds an illegal monopoly in general search and general search text advertising.

56 United States v. Google LLC, No. 1:23-cv-00108 (E.D. Va. Nov. 1, 2024). The court found Google liable for monopolization of the publisher ad server market and the open-web display advertising exchange market.

57 Directive (EU) 2019/790, arts. 18–23 (fair remuneration of authors and performers).

58 Radsch, "The AI Act and the Existential Risk Facing Journalism."

OPEN MARKETS INSTITUTE

CENTER FOR JOURNALISM & LIBERTY