The Licensing Mirage

What AI content deals are really worth to publishers - a financial analysis

An independent research paper. British English throughout. | Draft for long-form publication. Figures current to mid-2026.

Abstract

AI content licensing deals between publishers and AI companies are routinely reported as a significant new revenue stream for journalism and digital media. This paper tests that proposition against the financial evidence. Drawing on company filings, earnings-call transcripts, traffic studies, trade reporting and the published pricing of the data-broker and marketplace industry, it finds that licensing deals are real and, for a small number of large publishers, genuinely valuable, but that they are (a) heavily concentrated among a handful of premium, brand-name publishers; (b) substantially smaller than their headline figures imply once decomposed against total publisher revenue; and (c) far too small to offset the structural revenue decline now being driven by AI-mediated traffic loss. A central and underexamined finding is that the market-clearing price of publisher content, as revealed both by the third-party scraper economy and by the new AI content marketplaces, sits somewhere between a fraction of a penny and a few pence per article. That floor, not the headline lump sums, is the true economic signal. The paper concludes that licensing is best understood not as a new business model but as a transitional hedge for the few, and sets out the conditions under which that assessment would change.

Keywords: AI licensing, publisher monetisation, generative AI, referral traffic, AI Overviews, web scraping, content marketplaces, pay-per-crawl, copyright litigation.

Key findings (executive summary)

1. AI licensing is proliferating and, for a small set of large publishers, lucrative, but the evidence supports the thesis that these deals are concentrated, smaller than headlines suggest, and insufficient to offset AI-driven traffic loss.
2. The largest deal on record, News Corp and OpenAI, is reported at "up to" US$250m over five years (around US$50m a year). That figure is a press estimate that News Corp has never confirmed in any filing or on any earnings call. Even taken at face value it is roughly 2 per cent of News Corp's news-segment sales. Most other publishers were reportedly offered between US$1m and US$5m a year.
3. Not one major news publisher discloses AI licensing revenue as a separate, material line item in its accounts. Under US and UK reporting rules, material revenue streams must be disclosed. The silence is itself strong evidence that, relative to these businesses, the sums are immaterial.
4. The market price of content is far below the headlines. Third-party data brokers resell scraped publisher content for as little as a fraction of a penny per article, and the leading AI content marketplace prices crawls at roughly one penny per article. Decomposed to a per-use basis, the value of a single article to the AI supply chain is sub-penny to low-pence. The lump-sum deals are a negotiated premium over a near-zero commodity floor, available only to those with the brand and legal leverage to extract it.
5. Traffic displacement is large and accelerating. Google search referrals to publishers fell about 33 per cent globally in the year to November 2025, zero-click searches rose from 56 to 69 per cent, and independent studies put the click-through-rate reduction from AI Overviews at between roughly 35 and 89 per cent depending on method. The revenue lost dwarfs licensing income by one to two orders of magnitude.

6. The trajectory for 2025 to 2027 points to usage-based marketplaces, pay-per-crawl, collective licensing and litigation as the battlegrounds. Most analysts, and many publishers themselves, doubt any of these will replace lost search traffic at scale. In the Reuters Institute's 2026 survey, only a fifth of publishers expected substantial future AI revenue, and a further fifth expected none at all.

1. Deal inventory

AI content licensing began in earnest in December 2023, when Axel Springer signed the first major news deal with OpenAI, and accelerated through 2024 and 2025 into an ecosystem spanning bilateral licences, revenue-share programmes, marketplace intermediaries and academic data deals [1, 2]. A persistent feature, and an analytical theme of this paper, is that the overwhelming majority of financial terms are undisclosed, and the handful of figures in circulation are estimates produced by journalists rather than confirmed company disclosures.

1.1 Master deal inventory

Table 1. Selected major publisher–AI licensing deals. Values marked "est." are journalistic estimates, not company-confirmed figures.

Publisher	AI company	Announced	Reported value (status)	Structure
Axel Springer (Politico, Business Insider, Bild, Welt)	OpenAI	Dec 2023	"Tens of millions of euros" a year; around US$13m/yr over 3 yrs (est.)	Multi-year licence and product
Associated Press	OpenAI	Jul 2023	Undisclosed	Multi-year licence
Le Monde	OpenAI	Mar 2024	Undisclosed (5-yr)	Licence and attribution
Prisa Media (El País, AS)	OpenAI	Mar 2024	Undisclosed	Licence
Financial Times	OpenAI	Apr 2024	Around US$5–10m/yr (est.; OpenAI reportedly offered most outlets US$1–5m/yr)	Licence and product collaboration
Dotdash Meredith (People Inc.)	OpenAI	May 2024	Undisclosed	Licence and attribution
News Corp (WSJ, NY Post, The Times, The Sun, The Australian, Barron's, MarketWatch)	OpenAI	May 2024	"Up to" US$250m over 5 yrs (est.; never company-confirmed)	Cash plus technology credits, multi-year
Vox Media	OpenAI	Jun 2024	Undisclosed	Licence
The Atlantic	OpenAI	Jun 2024	Undisclosed	Licence and product
Time	OpenAI	Jun 2024	Undisclosed	Multi-year, 100-yr archive
Condé Nast	OpenAI	Aug 2024	Undisclosed	Multi-year licence
Hearst	OpenAI	Oct 2024	Undisclosed	Licence
Future plc	OpenAI	Dec 2024	Undisclosed	Licence and internal tooling
Schibsted	OpenAI	Feb 2025	Undisclosed	Licence
The Guardian	OpenAI	Feb 2025	Undisclosed	Licence and product

Publisher	AI company	Announced	Reported value (status)	Structure
Axios	OpenAI	Jan 2025	Undisclosed (3-yr; funded 4 local newsrooms)	Licence and newsroom funding
The Washington Post	OpenAI	Apr 2025	Undisclosed	Licence (display)
Reddit	Google	Feb 2024	Around US$60m/yr	API and data licence
Associated Press	Google	Jan 2025	Undisclosed	Real-time feed for Gemini
The New York Times	Amazon	May 2025	Around US$20–25m/yr (est.)	Licence (Alexa, training)
Condé Nast, Hearst	Amazon	Jul 2025	Undisclosed	Licence (Rufus shopping assistant)
Informa (Taylor & Francis)	Microsoft	2024	Over US$10m first year; recurring to 2027	Data-access licence
AFP	Mistral	Jan 2025	Undisclosed	Multi-year licence
News Corp	Apple	Oct 2025 (confirmed)	"Significant" (undisclosed)	Partnership
News Corp (WSJ et al.)	Meta	Mar 2026	"Up to" US$50m/yr over 3 yrs (est.)	Licence and training
CNN, Fox News, USA Today, People Inc., Le Monde and others	Meta	Dec 2025	Undisclosed (multi-year)	Licence for Llama / Meta AI
Reach (Mirror, Express)	Amazon	2025	Undisclosed	Usage-based (Nova, Alexa)

1.2 Intermediary and marketplace deals

A second layer sits beneath the bilateral deals: marketplaces and collectives that aggregate publisher supply and meter access. Perplexity's Publishers' Program, relaunched in August 2025 as Comet Plus, established a US$42.5m pool with an 80/20 revenue split in publishers' favour, drawn from US$5-a-month subscriptions, with partners including Time, Fortune, Gannett's USA Today, The Independent and the Los Angeles Times [3]. ProRata.ai operates a 50/50 revenue-share model and signed the FT, Axel Springer, The Atlantic and Fortune, alongside a member-wide agreement with the News/Media Alliance. TollBit runs a per-crawl marketplace across roughly 7,000 sites, of which only about a fifth earn money, ranging from hundreds to tens of thousands of dollars a month [4, 5]. Microsoft launched a Publisher Content Marketplace in late 2025 with early participants including the Associated Press and People Inc., and Snowflake's Cortex Knowledge Extensions offer enterprise retrieval licensing to 17 publishers including The Washington Post and the AP [4, 6]. In academic publishing, Wiley disclosed US$40m of total AI licensing revenue in its 2025 financial year, up from US$23m the year before [7].

1.3 Litigation as the alternative to licensing

The two routes, in News Corp chief executive Robert Thomson's phrase "woo and sue," run in parallel. The landmark case is The New York Times v. OpenAI and Microsoft, filed in December 2023 and consolidated in the Southern District of New York. The court allowed the core copyright claims to proceed in April 2025, and in January 2026 the presiding judge affirmed an order requiring OpenAI to produce around 20 million anonymised ChatGPT logs, with summary judgment expected around April 2026 [8]. Separately, Anthropic settled the Bartz authors' class action for US$1.5bn, described by lead counsel as the largest copyright recovery on record, at roughly US$3,000 per work across about 500,000 titles, with a final approval hearing set for May 2026. Litigation functions both as a substitute for licensing, where talks fail, and as leverage to improve licensing terms.

2. Financial reality check

The central finding of this section is stark: no major news publisher discloses AI licensing revenue as a separate, material line item. Where AI money appears at all, it is folded into "other revenue" or segment-level "content licensing" without a discrete figure. Under US and UK accounting standards, companies must separately disclose material revenue streams. The absence of such disclosure is itself strong evidence that, relative to these businesses, AI licensing is immaterial.

2.1 News Corp

News Corp is the most important test case. The reported "up to US$250m over five years" OpenAI figure is a press estimate that News Corp has never confirmed [1]. On the company's first earnings call after the deal, an analyst asked directly how much AI revenue had come in and how it split between Dow Jones and the news-media segment. Thomson replied that he could not be more specific given confidentiality requirements. Across every subsequent call, management has used qualitative framing such as "high-margin content licensing revenues" but has never broken out a discrete number [9]. If the press figure is correct, straight-line recognition implies around US$50m a year, which is roughly 2 per cent of News Corp's news-segment sales and well under 1 per cent of total company revenue, against group revenue of approximately US$8.5bn in FY2025. The company's digital pivot is real, with digital now 62 per cent of revenue and advertising down to 16 per cent from 32 per cent in 2018, but that transformation is driven by Dow Jones professional information and digital real estate, not by AI licensing [9].

2.2 The New York Times Company

The New York Times Company reported total revenue of US$2.59bn in 2024, up 6.6 per cent, with subscription revenue of around US$1.78bn and advertising of around US$506m [10, 11]. Its first AI deal, with Amazon in May 2025, is estimated at around US$20–25m a year, under 1 per cent of total revenue [1]. The Times chose to litigate rather than license OpenAI, and its filings disclose generative-AI litigation costs as an expense, not AI licensing as revenue [12]. The cost of fighting AI thus appears in its accounts before any AI licensing income does.

2.3 Future plc

Future plc illustrates the other side of the ledger. It has an OpenAI tooling and licensing arrangement but no material disclosed AI licensing revenue. Its FY2025 revenue fell 6 per cent to £739.2m, and first-half FY2026 profit before tax fell 67 per cent to £18.4m, with the company attributing the damage to falling Google search traffic. Future reports that only 27 per cent of sessions originate from Google search, yet a majority of group revenue still comes from brands dependent on Google traffic [13, 14]. Licensing is not offsetting this.

2.4 Comparators: Reddit and Wiley

Two non-news comparators are instructive because their deals are large enough to be material, and they disclose them. Reddit's chief operating officer said in early 2025 that AI licensing made up about 10 per cent of revenue, around US$130m a year, with Google paying around US$60m. Wiley disclosed US$40m of AI licensing revenue in FY2025 [7]. The contrast with news publishers, who do not break out such figures, is the point: where the money is material, it is disclosed.

2.5 Deal value against publisher revenue

Table 2. Reported AI deal value against total publisher revenue (order-of-magnitude).

Publisher	Reported AI deal (annualised)	Total annual revenue	AI licensing as % of revenue
News Corp	~US$50m (OpenAI, est., unconfirmed)	~US$8.5bn	<1% (~2% of news segment)
The New York Times	~US$20–25m (Amazon, est.)	US$2.59bn (2024)	<1%
Future plc	Not disclosed / immaterial	£739.2m (FY2025)	~0% disclosed
Wiley	US$40m (FY2025, disclosed)	~US$1.65bn	~2.4%
Reddit (comparator)	~US$130m (disclosed)	~US$1.3bn (2024)	~10%

3. Concentration analysis

The deals are heavily concentrated among a small number of large, brand-name publishers, and this concentration is structural rather than transitional. The Open Markets Institute's 2025 report on the AI content-licensing market maps a three-tiered structure: a small number of large bilateral deals between major AI companies and major publishers; an emerging intermediary layer of licensing marketplaces; and an uncompensated majority, the bulk of publishers and creators, sitting outside any licensing framework [15].

3.1 Why concentration is structural

AI companies prefer a handful of premium sources for three reasons. First, brand-name content carries trust and quality signals that improve model outputs and citations. Second, transaction costs favour a few large deals over hundreds of thousands of small ones. Third, a small number of marquee licences provides legal cover and public-relations value disproportionate to the content volume. One Reuters Institute analysis observed that OpenAI appears to be signing roughly one licensing deal per country, usually with the largest publisher, with Google a handful more, while Amazon and Microsoft pursue marketplace approaches [16]. Industry scorecards counted only around 18 OpenAI publisher-licensing deals globally as of 2025, a small slice of a universe of tens of thousands of news publishers.

3.2 The addressable universe

The realistic addressable universe for meaningful bilateral licensing is therefore narrow: large, major-market, brand-name publishers with deep archives and current reporting that AI companies value for recency. Small, local, niche, investigative and non-English publishers are largely excluded. The Reuters Institute's 2026 survey of 280 leaders in 51 countries found that around a fifth of publishers, mainly upmarket news companies, expected substantial future AI revenue; about half expected only a minor contribution; and a further fifth expected no income at all [2]. The New York Times publisher A. G. Sulzberger has made the same point, that the vast majority of news publishers expect no significant income from AI platforms [17].

3.3 Do collectives change the picture?

Collective licensing efforts, including the RSL Collective, the UK's SPUR alliance and the News/Media Alliance's ProRata agreement, aim to aggregate supply so that smaller publishers can participate. As of late 2025 they had not yet shifted the concentration picture materially, and no major AI company had committed to honour the principal collective standards at scale. The structural asymmetry, a fragmented publisher side facing a concentrated and well-resourced AI side, persists.

4. Traffic displacement analysis

The evidence that AI features are reducing publisher referral traffic is now extensive, consistent across methodologies, and large in magnitude.

4.1 Aggregate search referral decline

Chartbeat data across 2,576 sites worldwide, published in the Reuters Institute's 2026 trends report, showed Google search referrals to publishers dropping about 33 per cent between November 2024 and November 2025, and falling around 38 per cent in the United States, with Google Discover referrals down about 21 per cent [16]. Similarweb found that zero-click searches rose from 56 to 69 per cent between May 2024 and May 2025.

4.2 Click-through-rate collapse

Multiple independent studies converge on a sharp reduction in click-through when an AI Overview is present, although the magnitude varies with method and query type. The Pew Research Center, analysing nearly 69,000 real searches in March 2025, found users clicked a traditional result 8 per cent of the time when an AI Overview appeared, against 15 per cent without, and that around a quarter of AI Overview sessions ended with no click at all. Ahrefs reported a 58 per cent reduction in click-through for top-ranking pages associated with AI Overviews, up from a 34.5 per cent decline measured a year earlier. Authoritas and Seer Interactive reported reductions of roughly 61 to 79 per cent for the top organic result, and DMG Media reported declines of up to 89 per cent on some AI Overview searches. The range is wide, but every credible study points in the same direction.

4.3 Publisher-level declines

Similarweb estimates of year-on-year organic-search traffic to mid-2025 included falls of around 55 per cent at Business Insider, around 50 per cent at HuffPost and Forbes, around 44 per cent at Mail Online and around 40 per cent at The Washington Post. Aggregate news-sector organic traffic fell from around 2.3bn monthly visits in mid-2024 to under 1.7bn by May 2025, a loss exceeding 600 million monthly visits in under a year. Digital Content Next's member survey found median Google referrals down 10 per cent, with news brands down 7 per cent and non-news down 14 per cent [18].

4.4 Loss against gain

The comparison is unforgiving. A large publisher with, say, US$200m to US$500m in annual digital advertising and affiliate revenue heavily dependent on search-driven sessions would lose tens to low-hundreds of millions of dollars a year from a 25 to 50 per cent decline in search referrals, a range well within the documented evidence. Against that, the best AI licensing deals deliver low tens of millions a year, and most deliver far less or nothing. Future plc's accounts make the arithmetic concrete: a 67 per cent half-year profit collapse driven by traffic loss, against no material licensing offset [14]. At the macro level, Sulzberger told an industry congress that private AI investment in the United States reached nearly US$350bn in 2025, yet industry data suggested that less than half of 1 per cent of that investment went to compensate the publishers whose content powers the technology [17]. Looking forward, the Reuters Institute found publishers expecting search-engine traffic to almost halve, by around 43 per cent, over the following three years, with about a fifth anticipating a loss of more than three-quarters of their search traffic [2]. The structural conclusion is that licensing revenue is one to two orders of magnitude too small to offset traffic-driven losses for most publishers.

5. The scraper economy: what publisher content actually sells for

This section addresses a question the licensing debate usually skips. If publisher content is valuable enough to license, what is a single article actually worth on the open market? The answer, revealed independently by the third-party data-broker industry and by the new AI content marketplaces, is sobering: somewhere between a fraction of a penny and a few pence. This floor, not the headline lump sums, is the true economic signal, and it reframes the entire licensing conversation.

5.1 The scraper economy floor

A mature industry of data brokers already scrapes and resells web content at industrial volume, and its public pricing sets a hard floor under the value of an article. Five public numbers tell most of the story, compiled in the published analysis of Frederick Jahn, co-founder of the Berlin anti-scraping firm Centinel Analytica [19], and corroborated here against the vendors' own pricing pages.

- Bright Data's pay-as-you-go scraping APIs are advertised at US$1.50 per 1,000 records, falling to around US$0.75 to US$1.12 per 1,000 at committed volume [20, 21]. That is about 0.15 pence per page.
- Bright Data's pre-built datasets, including news datasets, start at around US$250 per 100,000 records, roughly 0.25 pence per record [21, 22].
- Zyte's entry pricing reaches as low as US$0.06 per 1,000 successful HTTP responses at its committed tier, about 0.006 pence per response [23].
- Bright Data operates a residential-proxy pool that the company markets in the hundreds of millions of rotating IP addresses, with independent reviews citing 150 million or more, enough to make scraping traffic extremely difficult to distinguish from human visitors [21, 24].
- Apify's marketplace offers thousands of pre-built scrapers, called Actors, with various sources citing counts from around 3,000 to over 20,000, many of them targeted at named domains [25].

Run these against a publisher's own output and the result is brutal. At dataset prices of roughly 0.25 pence per record, replicating an entire archive of 10,000 articles costs about US$25, and at pay-as-you-go scraping prices nearer US$15 [19]. Even if that estimate is wrong by an order of magnitude, the cost of replicating a publisher's life's work still fits inside a procurement department's discretionary budget. The significance is not that scraping is cheap in the abstract, but that any licensing negotiation with an AI laboratory necessarily takes place above this floor, and the floor is close to zero.

5.2 The brokers, not the publishers, are capturing the value

The firms running this arbitrage are substantial businesses. Bright Data, the largest, has reported crossing US$300m in annual recurring revenue in 2025, growing more than 50 per cent year on year, and serving a customer base its reviewers describe as including 14 of the top 20 global AI laboratories [24]. The media analyst Matthew Scott Goldstein, who convenes a twice-yearly forum of publishing executives and technology companies, has documented this layer in a report on the "scraper economy" covered across the trade press. He identifies at least 21 vendors, including Firecrawl, Exa, Tavily, Perplexity's Sonar and Bright

Data, that routinely scrape publisher content and resell data services to customers including OpenAI, Amazon and even other publishers such as The Telegraph; TollBit's own index puts the count closer to 40 [26, 27, 28]. Goldstein puts the size of this scraper economy at around US$1bn, citing Mordor Intelligence data, and his central finding is that publishers capture essentially none of it: none of them hold licensing agreements with these vendors. He calls it, plainly, a black market for AI content licensing [27, 29].

The structural insight is that AI companies are already paying for content. The money is simply going to the scrapers rather than to the people who created and own it [30]. One publishing executive quoted by Digiday compared the new data brokers to demand-side platforms for content, with the difference that at least advertising-technology middlemen moved money, whereas these intermediaries take everything and return nothing; Chris Dicker, chief executive of Candr Media, called it "a hostile takeover funded by our own IP" [28]. Brent Maynard of Akamai described the shift as a separation of roles, in which the entity extracting the data is often not the one monetising it, and quoted a large publisher's summary: "we're not being scraped by one company anymore, we're being harvested by an ecosystem" [26]. The juxtaposition with the licensing deals in Section 1 is the heart of the matter. A publisher can license its archive to an AI company for an undisclosed sum, while at the same moment a broker offers a scraped dataset of the same titles for a fraction of a penny per article.

5.3 The extraction ratio

The imbalance is visible in the ratio of scraping to referral. Analyses of Cloudflare and TollBit data put OpenAI's scrape-to-referral ratio at around 179 to 1, Perplexity's at around 369 to 1 and Anthropic's at a reported 8,692 to 1, against roughly 14 to 1 for Google's traditional search crawler; one publisher, Digital Trends, documented 4.1 million bot scrapes in a single week against 4,200 human referrals, a ratio of 966 to 1 [31]. By the fourth quarter of 2025, TollBit recorded one AI bot visit for every 31 human visits across its network, and click-through from AI applications had fallen to 0.27 per cent, and to 1.33 per cent even for publishers with deals [16]. The old value exchange, in which a crawler took a page and returned a reader, has broken down.

5.4 What publishers have actually priced content at in marketplaces

The marketplaces that let publishers charge for access provide the second independent reading of the market price, and it confirms the floor. TollBit prices on a cost-per-thousand basis that publishers already understand from advertising. Its co-founder told an interviewer that an article might sit at a US$10 CPM and move dynamically to US$15 on demand, that paywalled and newer content commands more than open or older content, and that finance commands a premium over entertainment [32]. A US$10 CPM is one penny per crawl: that is, the leading marketplace's typical price for an AI company to fetch a single article is roughly one penny. Publishers on the platform earn, in the co-founder's own words, anywhere from a few hundred dollars a month to tens of thousands, depending on their leverage [32]. The platform lets publishers set different prices for crawls that generate summaries and crawls that display full text, and exclude some content entirely [5].

Cloudflare's pay-per-crawl, launched in 2025 and now blocking AI crawlers by default for new domains, points the same way. Commentators who have modelled it describe a toll booth that pays in loose change, with per-page charges of a few cents at most and, at the low end, fractions of a penny [33]. Raptive's chief strategy officer was blunter still, predicting that flat pay-per-crawl pricing would mainly incentivise AI companies to crawl less rather than pay more, and arguing that pay-per-inference, with proper attribution, is the only model with real promise [34].

Set the two readings side by side. The cost to a broker of scraping an article runs from around 0.006 to 0.25 pence. The most a publisher typically charges to license one through the leading marketplace is around one penny. The entire commercial contest is being fought within a few multiples of a near-zero number, and the AI buyer will only pay the higher figure if blocking forces the choice.

5.5 The strategic logic of blocking

This pricing is also why the most credible publisher-side strategy is not to license at the floor but to raise the cost of taking content for free. As more publishers block crawlers, the argument runs, scrapers gain only limited access to premium content, AI companies must either pay for quality or make do with what is left, and the economics tilt back toward paying publishers directly rather than paying scrapers [30]. The blocking movement is now substantial: by late 2025, around 5.6 million sites had added OpenAI's GPTBot to their robots.txt disallow list, up roughly 70 per cent since July, with comparable rises for Anthropic's and Apple's crawlers, and TollBit recorded a 336 per cent year-on-year increase in sites blocking AI crawlers [35, 36]. Cloudflare's most aggressive setting, blocking all AI scrapers at the HTTP level, was switched on by more than 800,000 sites [37].

The difficulty is enforcement. Robots.txt is voluntary, and TollBit found that a large share of AI bot scrapes, by some measures around 30 per cent, ignored explicit disallow instructions [27, 36]. Frederick Jahn's company exists precisely because, as he puts it, user-agent blocking is easily bypassed, legal action is slow and expensive, and most publishers rely on a single security tool while AI companies have dozens of scraping methods designed to defeat it; in his pitch to publishers, his firm offers to send a publisher a link to all its own articles, premium included, scraped in minutes, to prove the point [19, 38]. Goldstein's prescription is characteristically direct: block every one of these companies, and then block some more [27]. The economic reading of that advice is simple. Until taking content for free becomes expensive, the market price of an article will stay where the scraper economy has set it, which is close to nothing, and no licensing model can rise far above its own floor.

6. Structural critique

6.1 The sceptics

The most pointed sceptical case comes from David Skok, chief executive and editor-in-chief of The Logic, writing in Nieman Lab's predictions for journalism in 2026 under the title "Publishers will see no meaningful AI licensing revenue" [39]. His argument is that the market is structurally jammed: as long as Google's search crawler and its AI-training crawler function as a single system, the market for licensing journalism into AI models is effectively stalled, because there is no price discovery, no competitive tension and no incentive for major AI companies to write large recurring cheques. His prescription is to invest in direct audiences, owned platforms, first-party data and events, and to stop expecting the licensing money to arrive.

Jason Kint, chief executive of Digital Content Next, frames the problem as a competition issue, arguing that Google leverages its adjudged search monopoly to train on and substitute for news content, and that because it ties AI-training permissions to search inclusion, publishers cannot meaningfully opt out [40]. The Professional Publishers Association has stressed that an opt-out does not resolve the underlying value-exchange question while AI Overviews continue to replace clicks. The Open Markets Institute warns that the emerging licensing market risks repeating the mistakes of the search and social era, entrenching concentration and accelerating what it calls content cannibalisation [15].

6.2 The opposing view

Not all voices are sceptical, and the paper would be unbalanced without the other side. Roger Lynch, chief executive of Condé Nast, told staff that the OpenAI partnership began to make up for some of the revenue lost over a decade of platform disruption. Neil Vogel, chief executive of People Inc., has argued that both lump-sum and pay-per-use models can be viable so long as content is respected and paid for, and credits crawler-blocking with restoring publisher leverage. The Reuters Institute found that platform payments are now seen by publishers as the single most significant new growth opportunity, with interest almost doubling in two years [2]. OpenAI's representatives argue that AI-referred traffic can be higher in quality, with longer dwell times and a higher propensity to subscribe, a claim partly corroborated by some niche publishers. Usage-based advocates contend that a well-designed micropayment market could ultimately reward original, differentiated and local content better than the old search-traffic regime did.

The honest synthesis is that licensing is a genuine, welcome and growing revenue stream for a fortunate minority, but that the weight of evidence and expert opinion is that it is neither large enough nor broad enough to be the sustainable foundation publishers need.

7. Forward outlook

The deal landscape is shifting from one-off bilateral licences toward three overlapping models, none of which yet looks capable of replacing lost search revenue at scale.

7.1 Usage-based marketplaces

Microsoft's Publisher Content Marketplace, Perplexity's Comet Plus, TollBit, ProRata and Snowflake are moving the market toward pay-per-use. The appeal is scalability and fairness; the risk, established in Section 5, is that as AI click-through falls toward zero and subscriber bases stay small, the absolute sums remain modest. OpenAI has stated that it will not share advertising revenue with publishers, and gives flat fees rather than revenue shares, a sign that the dominant player intends to cap rather than expand publisher participation [41, 42].

7.2 Pay-per-crawl and technical leverage

Cloudflare's pay-per-crawl and AI Crawl Control give publishers, for the first time, a technical mechanism to charge or block, and have restored some negotiating leverage. But pricing is at the floor and adoption is early, and smaller publishers have little leverage even with the tools.

7.3 Collective bargaining and litigation

The RSL Collective, the SPUR alliance and the News/Media Alliance aim to aggregate supply, while litigation aims to establish that training and substitution require payment. The summary-judgment ruling expected in The New York Times v. OpenAI around April 2026, and any regulatory forcing of a Google crawler split by the UK Competition and Markets Authority or the European Union, are the two events most likely to change the economics. Until then the market remains, in Skok's word, stalled.

7.4 Scenarios for 2025 to 2027

- Most likely, a muddle-through, in which licensing remains a meaningful supplement for the top tier and immaterial for the long tail, usage-based models grow but stay small, and traffic decline continues. Publishers pivot hard to direct audiences, subscriptions, video, events and first-party data.
- Upside, in which a landmark legal ruling or a regulatory crawler separation creates genuine price discovery, forcing larger recurring payments and broadening participation through collectives.

- Downside, in which AI answer engines and agentic browsers drive click-through toward zero, deals are not renewed at higher values, and the uncompensated majority widens, accelerating closures among local and independent outlets.

8. What licensing deals tell us about the future of publisher monetisation

The headline numbers, the US$250m, the US$42.5m, the billions in damages, have created an impression that AI licensing is a major new pillar of publisher revenue. The evidence does not support that impression. Examined against total revenues, the deals are small. Examined across the industry, they are concentrated among a privileged few. Examined against the traffic they are implicitly meant to compensate for, they are one to two orders of magnitude too small. And examined against the open-market price of content itself, set by a scraper economy that resells articles for a fraction of a penny and by marketplaces that price them at around one penny, the lump-sum deals are revealed for what they are: a negotiated premium over a near-zero commodity floor, available only to those with the brand and the legal leverage to extract it.

The single most telling fact in this analysis is a negative one. Not one major news publisher reports AI licensing as a material, separate line in its accounts, and News Corp, holder of the largest reported deal, has repeatedly declined even to confirm its value. What licensing deals actually tell us is that the value in the publisher and AI relationship is flowing the other way. AI systems, and the broker economy that feeds them, are extracting enormous value from publisher content while returning collapsing referral traffic and modest, concentrated cheques.

Licensing is therefore best understood not as a new business model but as a partial, transitional hedge: useful legal cover and incremental cash for those large enough to negotiate it, and irrelevant for everyone else. The publishers most likely to survive will treat licensing as a welcome supplement, not a strategy, and will rebuild around the things AI cannot easily take, namely direct audience relationships, subscriptions, first-party data, distinctive original reporting, community and live events. The future of publisher monetisation will be decided less in licensing negotiations than in courtrooms, in regulators' offices, in the daily contest to own the audience relationship directly, and in raising the cost of taking content for free until the market price of an article rises above the floor the scraper economy has set for it. On current evidence, licensing is a mirage of sufficiency over a structural deficit, real enough to see, and far too small to live on.

9. Recommendations

For large, brand-name publishers, the immediate priority is to pursue non-exclusive, multi-model licensing to maximise competitive tension, but to insist on usage-based or hybrid structures with audited reporting rather than flat fees that cap the upside. Any single deal should be treated as worth no more than around 2 per cent of revenue and planned for accordingly. A renewal that comes in below its original annualised value should be read as a signal that the market is deflating, not growing.

For mid-sized and smaller publishers, bilateral deals should not be expected, and the marketplace and pay-per-crawl options should be seen for what the evidence here suggests they are: at present clearing prices, a measurement and leverage layer, not a meaningful revenue line. The sensible first step is to instrument the site to see who is taking what - which AI systems are crawling, at what volume, and against how little referral - using whichever bot-analytics or bot-management provider fits the publisher's stack. That visibility is now available from a range of vendors and does not require committing to any single marketplace. Any subsequent decision to route enforcement or monetisation through an intermediary should be weighed as an infrastructure dependency rather than a free revenue add-on, since interposing a third party at the edge cedes a degree of control over a publisher's own traffic and data at exactly the layer this paper has shown to be strategically valuable, and the resulting switching costs accrue to the intermediary over time. With the per-unit sums currently negligible, that trade is seldom worth making reflexively. The one route with a genuine structural rationale for the long tail is collective licensing, where aggregated bargaining is the only plausible source of compensation.

For all publishers, the strategic task over the next six to eighteen months is to reallocate investment toward direct audiences, first-party data, video and audio, events, and hard-to-summarise original journalism, the categories least substitutable by AI Overviews, while reducing dependence on search-driven, commoditised service content.

Three thresholds would change this advice. A Google crawler split forced by a competition regulator would create real price discovery and justify heavier investment in licensing infrastructure. A summary-judgment or trial outcome against fair use in The New York Times v. OpenAI would shift leverage decisively toward publishers and likely raise deal values industry-wide. And AI-referral click-through stabilising above roughly 1 to 2 per cent, with rising subscriber conversion, would make usage-based models genuinely additive. Absent at least one of these, AI licensing should be treated as a hedge, not a plan.

10. Notes on method and limitations

Figures in this paper are overwhelmingly estimates. The News Corp US$250m, the FT US$5–10m a year against an OpenAI baseline of US$1–5m for most outlets, the Axel Springer "tens of millions of euros" and the New York Times and Amazon US$20–25m a year are journalistic estimates, not company-confirmed, and each has been flagged as such. They should not be treated as audited.

Non-disclosure cuts both ways. The absence of separate AI-revenue line items strongly implies immateriality, but it is an inference rather than a direct disclosure; a definitive test would require line-item review of segment notes in annual filings, which to date do not break out AI licensing.

Traffic and click-through studies vary in method and incentive, with reported declines ranging from around 35 to 89 per cent, and some originate with vendors that have a commercial interest. This paper has prioritised the most rigorous sources, including Pew, Ahrefs, Chartbeat and Digital Content Next, and has presented ranges rather than single figures.

The scraper-pricing figures are list prices drawn from vendors' public pricing pages and from independent pricing reviews; effective prices vary with volume, site difficulty, rendering mode and negotiated commitments, and some vendors price per successful response rather than per page, which can raise the real cost. The per-article conversions are illustrative and rounded, and assume an exchange rate near US$1.25 to the pound. The figures attributed to Frederick Jahn and Matthew Scott Goldstein are drawn from their public commentary and from trade-press reporting of their work; direct access to the underlying LinkedIn posts was not possible, the pricing claims have been independently corroborated against vendor sources, and the market-sizing figure of around US$1bn is attributed to Goldstein citing Mordor Intelligence and has not been independently verified.

Finally, the market is moving quickly. Deal structures, marketplace models and litigation are evolving month to month, and figures and rulings cited here are current to roughly mid-2026 and will date. Counter-evidence is real: some publishers report higher-quality AI-referred traffic and genuine incremental revenue, and a favourable legal or regulatory shock could materially improve the economics.

References

[1] Press Gazette, News generative AI deals revealed: who is suing, who is signing, 2025. URL: https://pressgazette.co.uk/platforms/news-publisher-ai-deals-lawsuits-openai-google/, accessed: 2026-06-04.

[2] Reuters Institute for the Study of Journalism, Journalism, media, and technology trends and predictions 2026, 2026. URL: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2026-01/Trends_and_Predictions_2026.pdf, accessed: 2026-06-04.

[3] Subscription Insider, Perplexity unveils $42.5m revenue-sharing model as Comet browser takes aim at Google, 2025. URL: https://www.subscriptioninsider.com/article-type/news/perplexity-unveils-42-5m-revenue-sharing-model-as-comet-browser-takes-aim-at-google, accessed: 2026-06-04.

[4] Digiday, The Washington Post's Arc XP adds TollBit to help publishers make money from AI bot traffic, 2026. URL: https://digiday.com/media/the-washington-posts-arc-xp-adds-tollbit-to-help-publishers-make-money-from-ai-bot-traffic/, accessed: 2026-06-04.

[5] Crazy Stupid Tech, Why the AI revolution needs tollbooths, 2025. URL: https://crazystupidtech.com/2025/06/22/why-the-ai-revolution-needs-tollbooths/, accessed: 2026-06-04.

[6] The Media Copilot, Some publishers are turning AI into a revenue stream through Snowflake, 2025. URL: https://mediacopilot.ai/snowflake-ai-licensing-deals-publishers/, accessed: 2026-06-04.

[7] John Wiley & Sons, Inc., Form 8-K: fourth-quarter and full-year fiscal 2025 results, 2025. URL: https://www.sec.gov/Archives/edgar/data/0000107140/000010714025000071/wly-2025430xex991.htm, accessed: 2026-06-04.

[8] National Law Review, OpenAI loses privacy gambit: 20 million ChatGPT logs likely headed to copyright plaintiffs, 2026. URL: https://natlawreview.com/article/openai-loses-privacy-gambit-20-million-chatgpt-logs-likely-headed-copyright, accessed: 2026-06-04.

[9] PPC Land, News Corp reports AI content partnerships as digital subscriptions reach 62% of revenue, 2025. URL: https://ppc.land/news-corp-reports-ai-content-partnerships-as-digital-subscriptions-reach-62-of-revenue/, accessed: 2026-06-04.

[10] Statista, New York Times Company revenue 2024, 2025. URL: https://www.statista.com/statistics/192848/revenue-of-the-new-york-times-company-since-2006/, accessed: 2026-06-04.

[11] The New York Times Company, Form 8-K: quarterly results, 2024. URL: https://www.sec.gov/Archives/edgar/data/0000071691/000007169124000158/pressrelease09302024.htm, accessed: 2026-06-04.

[12] Statista, New York Times Company revenue by source 2024, 2025. URL: https://www.statista.com/statistics/192911/revenue-of-the-new-york-times-company-by-source/, accessed: 2026-06-04.

[13] Press Gazette, Why Future plc share price is so low: Google blamed by CEO, 2025. URL: https://pressgazette.co.uk/publishers/magazines/why-future-share-price-is-so-low/, accessed: 2026-06-04.

[14] Press Gazette, Future reveals it is still heavily reliant on Google as profit falls 67%, 2026. URL: https://pressgazette.co.uk/media_business/future-reveals-it-is-still-heavily-reliant-on-google-as-profit-falls-67/, accessed: 2026-06-04.

[15] Open Markets Institute, Mapping the AI content-licensing market, 2025. URL: https://www.openmarketsinstitute.org/publications/report-mapping-the-ai-content-licensing-market, accessed: 2026-06-04.

[16] A Media Operator, AI bot traffic's effect on publishers: the indirect costs adding up, 2026. URL: https://www.amediaoperator.com/analysis/ai-bots-crawlers-web-hosting-costs-indirect/, accessed: 2026-06-04.

[17] Reuters Institute for the Study of Journalism, New York Times publisher A. G. Sulzberger on why (and how) news publishers should fight AI platforms, 2025. URL: https://reutersinstitute.politics.ox.ac.uk/news/new-york-times-publisher-g-sulzberger-why-and-how-news-publishers-should-fight-ai-platforms, accessed: 2026-06-04.

[18] Playwire, The complete publisher's guide to AI crawlers: block, allow, or optimize for maximum revenue, 2025. URL: https://www.playwire.com/block-ai, accessed: 2026-06-04.

[19] F. Jahn, Public pricing of the web-scraping market (LinkedIn commentary), Centinel Analytica, 2026. URL: https://de.linkedin.com/in/frederickjahn, accessed: 2026-06-04.

[20] SERP, Bright Data: reviews, pricing, features, alternatives, 2026. URL: https://serp.co/products/brightdata.com/reviews/, accessed: 2026-06-04.

[21] Costbench, Bright Data pricing 2026, 2026. URL: https://costbench.com/software/web-scraping/bright-data/, accessed: 2026-06-04.

[22] Oreate AI, Unlocking web data: a look at Bright Data's pricing and offerings, 2026. URL: https://www.oreateai.com/blog/unlocking-web-data-a-look-at-bright-datas-pricing-and-offerings/70cf39b8dc8b99102360bca1bf04d728, accessed: 2026-06-04.

[23] Prospeo, Zyte pricing, reviews, pros and cons (2026), 2026. URL: https://prospeo.io/s/zyte-pricing-reviews-pros-and-cons, accessed: 2026-06-04.

[24] Dupple, Bright Data review 2026: features, pricing and alternatives, 2026. URL: https://dupple.com/tools/bright-data, accessed: 2026-06-04.

[25] Proxies.sx, Best web scraping APIs compared: 2026 pricing and performance guide, 2026. URL: https://www.proxies.sx/blog/best-web-scraping-api-comparison-2026, accessed: 2026-06-04.

[26] Digiday, In graphic detail: new data shows publishers face growing AI bot, third-party scraper activity, 2026. URL: https://digiday.com/media/in-graphic-detail-new-data-shows-publishers-face-growing-ai-bot-third-party-scraper-activity/, accessed: 2026-06-04.

[27] Digiday, Media briefing: another AI threat emerges for publishers, the third-party scraper, 2026. URL: https://digiday.com/media/media-briefing-another-ai-threat-emerges-for-publishers-the-third-party-scraper/, accessed: 2026-06-04.

[28] Playwire, AI scrapers are the ad tech tax, but worse, 2026. URL: https://www.playwire.com/blog/ai-scrapers-are-the-ad-tech-tax-but-worse, accessed: 2026-06-04.

[29] Fast Company, AI scraping is creating an existential dilemma for media companies, 2026. URL: https://www.fastcompany.com/91539092/ai-scraping-become-own-media-business, accessed: 2026-06-04.

[30] Raptive, How blocking AI bots paves the way for fair compensation, 2026. URL: https://raptive.com/blog/ai-bot-blocking-strategy/, accessed: 2026-06-04.

[31] Security Boulevard, The AI content crisis: how LLMs are draining media revenue, and the technologies fighting back, 2026. URL: https://securityboulevard.com/2026/04/the-ai-content-crisis-how-llms-are-draining-media-revenue-and-the-technologies-fighting-back/, accessed: 2026-06-04.

[32] Business Standard, TollBit is about monetising AI bot traffic, says co-founder and CEO, 2026. URL: https://www.business-standard.com/companies/people/tollbit-is-about-monetising-ai-bot-traffic-says-cofounder-and-ceo-126030201054_1.html, accessed: 2026-06-04.

[33] Leaky Paywall, Cloudflare's pay-per-crawl: sustainable income or just spare change?, 2026. URL: https://leakypaywall.com/cloudflare-pay-per-crawl-income-or-spare-change/, accessed: 2026-06-04.

[34] Digiday, Media briefing: AI payouts may be entering a new era, 2025. URL: https://digiday.com/media/media-briefing-ai-payouts-may-be-entering-a-new-era/, accessed: 2026-06-04.

[35] The Register, Publishers say no to AI scrapers, block bots at server level, 2025. URL: https://www.theregister.com/2025/12/08/publishers_say_no_ai_scrapers/, accessed: 2026-06-04.

[36] TollBit, State of the bots, Q1 and Q2 2025, 2025. URL: https://tollbit.com/bots/25q1/ and https://tollbit.com/bots/25q2/, accessed: 2026-06-04.

[37] A Media Operator, Publishers tackle AI scraping - it's not going well, 2025. URL: https://www.amediaoperator.com/analysis/publishers-tackle-ai-scraping-its-not-going-well/, accessed: 2026-06-04.

[38] Centinel Analytica, AI crawler detection and content protection, 2026. URL: https://www.centinelanalytica.com/, accessed: 2026-06-04.

[39] D. Skok, Publishers will see no meaningful AI licensing revenue, Nieman Journalism Lab (Predictions for Journalism 2026), 2025. URL: https://www.niemanlab.org/2025/12/publishers-will-see-no-meaningful-ai-licensing-revenue/, accessed: 2026-06-04.

[40] J. Kint, Public commentary on AI Overviews and search-tied training permissions, X, 2025. URL: https://x.com/jason_kint/status/1959867906763657452, accessed: 2026-06-04.

[41] Press Gazette, OpenAI not planning to share advertising revenue with publishers, 2025. URL: https://pressgazette.co.uk/platforms/openai-not-planning-to-share-advertising-revenue-with-publishers/, accessed: 2026-06-04.

[42] eMarketer, OpenAI, Perplexity secure more publisher licensing deals, 2025. URL: https://www.emarketer.com/content/openai--perplexity-secure-more-publisher-licensing-deals, accessed: 2026-06-04.