

AI and the Future of Independent Journalism.

May, 2026

The promise and peril of privately controlled data marketplaces for media content.

By Courtney C. Radsch | Edited by Phillip Longman

Greenfield County Post

HOME LOCAL WEATHER POLITICS INVESTIGATIONS NEWSLETTERS



Harper calls for school funding, property tax cap in run for County Council seat

by Angelica Kraus

The race for the Greenfield County Council seat has taken an unexpectedly energetic turn this season, as residents find themselves drawn into a debate that cuts to the heart of everyday life: how to fund local schools without overburdening homeowners.

At the center of the conversation is first-time candidate Daniel Harper, a former teacher and small business owner who has built his campaign around two closely linked promises—boosting school funding and implementing a cap on property taxes. His message has resonated across neighborhoods where concerns about rising living costs are matched only by worries about overcrowded classrooms and aging school facilities.

"People shouldn't have to choose between a strong education system and staying in their homes," Harper said at a recent town hall meeting. "We can—and must—do both responsibly."

Greenfield County's public schools have struggled in recent years with budget shortfalls, leading to larger class sizes, reduced extracurricular programs, and delayed infrastructure repairs. Parents and educators have voiced frustration, arguing that the current funding model consistently leaves too much uncertainty from

01100110 10011001 0110011 11001100 01010101 11111100 00000011 10110010
01101001 10010110 00101101 11010010 01011111 11101010 00010101 10100101
01111110 10000001 00111010 11000101 01010011 11101100 00001010 10111100
01100011 10011100 00110101 11001010 01011010 1100101 00011110 10101001
01100101 10011010 00110010 11001111 01010110 11111001 00000110 10110001
01101010 10010101 00101110 11010001 01001110 11101001 10101010 01100110
10011001 0110011 11001100 01010101 11111000 00000111 10111000 01100101
10011010 00101101 11010010 01011111 11101010 00010101 10100101 01111110
10000001 00111010 11000101 01010011 11101100 00001010 10111000 01100011
10011000 00110101 11001010 01011010 1100101 00011110 10101001 01100101
10011010 00110010 11001111 01010110 11111001 00000110 10110001 01101010
01011001 11100010 00011101 10101010 01100110 10011001 0110011 11001100
01010101 11111100 00000011 10110010 01101001 10010110 00101101 11010010
01011111 11101010 00010101 10100101 01111110 11010001 01001110 11101001
10101010 01100110 10011001 0110011 11001100 01010101 11111100 00000011
10100101 01101001 10010110 00101101 11010010 01001111 11101010 00010101
10100101 01111110 10000001 11111001 11011 00000110 10110001 01101010 10010101
00101110 11010001 01001110 11101001 10101010 01100110 0011 01 00110101
10011001 0110011 11001100 01010101 11111100 00000011 10110010 01101001
10010110 00101101 11010010 01001111 11101010 00010101 10100101 01111110
10000001 00111010 11000101 01010011 11101100 00001010 10111000 01100011
10011100 00110101 11001010 01011010 1100101 00011110 10101001 01100101
10011010 00110010 11001111 01010110 11111001 00000110 10110001 01101010

Search... Search

More in Local Politics

- Greenfield Voters Weigh Competing Visions for Growth and Affordability
- The Debate over Data Center Construction is Heating Up
- Council Weighs Incentives to Attract More Small Businesses to Greenfield

Read More

Entertainment



Art Walk Expands as County Creatives Gain Regional Attention

- Summer Concert Series Returns to Greenfield Park with Biggest Lineup Yet
- Local Theater Group Revives Classic Musicals with a Modern Twist
- Food & Film Festival Brings Crowds to Downtown
- A Look at the Summer's Biggest Movie Releases

Sports

- Greenfield High Advances to State Finals in Thrilling Overtime Victory
- County Youth Soccer League Sees Record Participation This Season

Courtney C. Radsch, PhD, is the director of the Center for Journalism & Liberty at Open Markets Institute and a global thought leader on technology, AI, and the media.

Phillip Longman is a senior editor at the *Washington Monthly* magazine and policy director at the Open Markets Institute.

**Washington
Monthly**
INSTITUTE



*This paper is released in collaboration with the Center for Journalism and Liberty at Open Markets Institute.
It was produced thanks to the support of Lumina Foundation.*

Table of Contents

Executive Summary & Background 1

Issue Definition 3

**Cloudflare’s New
Architecture of Consent** 5

Potentials for Abuse 7

Policy Recommendations 9

Conclusion 11

Executive Summary

Predatory business practices and failed public policy have made the spread of artificial intelligence a potentially existential threat to journalism. Artificial intelligence companies rely on journalists' work to train their models while offering little or no compensation.¹ At the same time, AI-generated news summaries divert readers from original reporting.² Small, local, and independent publications are especially vulnerable, lacking the resources to pursue legal remedies or negotiate fair terms with AI companies.

Despite numerous public policy proposals, none have been implemented at the federal level,³ let alone provided meaningful relief to journalists and content creators. New technology could, however, offer an important step toward a solution if properly regulated. Cloudflare,⁴ which manages about 20 percent of global web traffic,⁵ now makes it harder for AI companies to scrape content from third-party websites without permission or payment. This could reduce the growing market power imbalance between journalists and Big Tech. However, as internet history shows, there are significant downside risks that government policy must address.

Background

Conventional wisdom holds that the declining fortunes of independent journalism are largely the result of changes in communications technology.⁶ According to this view, inventions like the World Wide Web, smart phones, social media, and targeted digital advertising rendered the business models that had traditionally supported journalism obsolete. Yet this received idea overlooks the role of political economy in the development of the internet and thereby misses clear lessons for the future of local and independent journalism in the age of AI.

It is easy to forget that both large and small media organizations initially benefited from the rollout of the World Wide Web and associated digital technology. Not only did many newspapers, magazines, and other traditional media outlets successfully attract readers with online products,⁷ but by doing so, they often attracted new sources of advertising revenue⁸ that strengthened their ability to underwrite serious journalism. The primary reason was the emergence of so-called AdTech marketplaces.⁹ Pioneered by a company known as DoubleClick¹⁰ in the mid-1990s, these automated, online auctions allowed many local newspapers, small magazines, and even individual “bloggers,” as they came to be known, to earn significant revenue by selling banner ads to marketers—including national and global brands—that they could otherwise never have attracted.¹¹

1 <https://www.seattletimes.com/opinion/researchers-document-ai-firms-pilfering-of-news-sites/>

2 See page 14 of this Fastly report

3 <https://aimultiple.com/is-web-scraping-legal/>; S.3813 - 119th Congress (2025-2026): CLEAR Act

4 <https://www.cloudflare.com/en-gb/learning/what-is-cloudflare/>

5 Cloudflare Just Changed How AI Crawlers Scrape the Internet-at-Large

6 <https://casi.stanford.edu/news/press-peril-how-power-politics-and-profit-are-reshaping-journalism>

7 <https://web.mit.edu/comm-forum/legacy/papers/harper.html>

8 Between 2000 and 2020, digital advertisements generated \$3.1 billion in revenue for media outlets

9 See page 8 of this OPI report

10 <https://en.wikipedia.org/wiki/DoubleClick>

11 Ibid

The automation and global reach of early AdTech was particularly valuable for local newspapers and other small and new, start-up publications that by and large lacked the technical expertise to handle digital ad sales. Yet the same the AdTech technology that was initially a huge boon the financing of journalism soon transformed into a deep threat once it came under monopoly control.

Specifically, by failing to enforce traditional antitrust laws and bans on self-dealing,¹² policy makers and enforcers allowed Google to form an illegal monopoly that came to control the buying and selling of nearly all digital advertisements in independent online publications throughout the world.¹³ As has now been confirmed by a landmark court decision, Google used this policy failure to gain overwhelming market power and drive down the amount of ad revenue it shared with media organizations participating in its AdTech marketplace.¹⁴

Google also used its vast trove of personal data from products like Google Search and Google Mail to divert more and more surveillance advertising dollars away from news organizations and toward its own media properties like YouTube.¹⁵ This resulted in billions in revenue going to Google's stockholders instead of newsrooms.¹⁶

We face a similar crossroads today. A new digital technology can enable a viable business model for independent journalism by ensuring fair compensation for media organizations. However, this outcome depends on establishing market rules to prevent monopolization or control by self-dealing digital platforms.

12 Google's Antitrust Paradox: Can Modern Antitrust Principles Fix Google Search?

13 Google controls over 90 percent of all digital advertising, according to the DOJ

14 <https://fortune.com/2025/11/13/cloudflare-ceo-google-abusing-monopoly-search-ai/>

15 <https://www.wired.com/story/google-dv360-banned-audience-segments-national-security/>;
<https://www.investopedia.com/investing/companies-owned-by-google/>

16 Why publishers fear traffic, ad declines from Google's AI-generated search results - Digiday

Issue Definition

Large-scale AI models, used for chatbots and digital agents, are trained on vast amounts of internet data scraped by web crawlers.¹⁷ To stay current, they need timely, accurate information. The best data comes from up-to-date online publications, databases, and digital content repositories that reflect human knowledge and expertise. Without these sources, AI applications like ChatGPT can't keep up with local sports scores, much less provide useful information about current events.¹⁸

To be reliable, AI models must also be trained on mostly true data. Training them with indiscriminately scraped internet data risks giving them the “intelligence” of a conspiracy theorist reasoning from false premises. This problem worsens with the spread of AI. People and governments worldwide are overwhelming the internet with zettabytes of AI-generated “synthetic media” and “slop.”¹⁹ Amazon is inundated with AI-generated books and product reviews.²⁰ Facebook is overrun by zombie accounts.²¹ Google Search and Google News are infested with AI-generated fake news and junk content.²²

“Perplexity had taken our work, without our permission, and republished it across multiple platforms—web, video, mobile—as though it were itself a media outlet.”

— Randall Lane, Forbes chief content officer

AI companies also know that, even under the best of circumstances, training their models using data generated by other AI models brings serious risks. When the original training data is sound, this approach is still like trying to advance human knowledge using photocopies of photocopies *ad infinitum*. The resulting models become distorted and less and less faithful to reality. Eventually they tend to malfunction, degrade, and potentially even collapse, rendering AI useless, if not downright harmful.²³ Meanwhile, as synthetically produced AI-generated content drowns out other content online, the resulting “enshittification” of the internet poses an existential threat to the very foundation of the AI paradigm.

Mindful of their dependence on quality, timely data, AI corporations have relied on web crawlers to scrape content off news organizations’ websites without permission, much less compensation.²⁴ Today, they often compound the harm by using this purloined intellectual property to generate their own competing media products. For example, Perplexity.ai,²⁵ a free AI-powered search engine, allegedly engaged in such a parasitical practice to profit off

17 <https://blog.cloudflare.com/ai-crawler-traffic-by-purpose-and-industry/>

18 Here’s a look at how the newly up-to-date ChatGPT reports the latest news

19 Over 50 percent of all internet content in English is AI-generated, study shows; 20 percent of all YouTube content is AI-generated; close to 200 zettabytes of data now exist on the internet worldwide

20 <https://www.npr.org/2024/03/13/1237888126/growing-number-ai-scam-books-amazon>

21 Meta Opens Floodgates For AI-Generated Accounts On Facebook, Instagram

22 See quote from Sara Platnick of Perplexity in this article on AI content scraping and journalism

23 AI models collapse when trained on recursively generated data | Nature

24 Generative AI Has a ‘Shoplifting’ Problem. This Startup CEO Has a Plan to Fix It | WIRED

25 <https://thefix.media/2024/09/06/how-perplexity-ai-partners-with-major-publishers/>

journalism financed and developed by *Forbes*.²⁶ “Perplexity had taken our work, without our permission, and republished it across multiple platforms—web, video, mobile—as though it were itself a media outlet,” lamented Randall Lane, *Forbes*’s chief content officer and editor.²⁷ The search engine purportedly plagiarized a major scoop by the company, spinning up an article that regurgitated much of the same prose as the *Forbes* article, while also generating an accompanying podcast and YouTube video that outperformed the original on search.²⁸

More recently, Google has caused a massive loss of readership²⁹ for most online publications by adding AI-generated “overviews” to search engine results. Google’s publication of these synthetic summaries causes more and more users to no longer click on links to original sources.³⁰ *HuffPost*,³¹ *Business Insider*,³² and other publishers reported 40 percent declines in web traffic following the introduction of Google’s AI Overviews.³³ In response to a recent query about the fate of the *Washington Post*, Google’s AI gave this example³⁴ of its own destructive effect on journalism:

“The Washington Post has experienced a severe, multi-year decline in search engine referrals, with traffic from search engines falling by nearly 50% over the past three years. This collapse, alongside a significant drop in subscribers, directly contributed to the company’s decision to lay off approximately 30% of its newsroom staff in early 2026.”

The European Commission³⁵ has opened a formal investigation into whether Google is using publisher content to power AI services without appropriate compensation.³⁶ Other AI products, such as Anthropic’s Claude, similarly depend on the work of human journalists and content creators while increasingly competing against them for media consumers’ attention.³⁷ Making matters worse for journalists and publishers, up to 20 percent of their annual web revenue now goes to cover operating costs created by the volume of AI bots scraping their sites.³⁸

Publishers and website owners have attempted to fight back through various means. Many have put their content behind paywalls, but this deeply limits their reach with both viewers and advertisers³⁹ and often does not provide a sustainable business model. From a societal perspective, it also means surrendering even more of the information environment to foreign propaganda and junk content—turning quality information into a luxury good.

Large media enterprises have also brought lawsuits charging copyright infringement.⁴⁰ But copyright doctrines vary across jurisdictions, and courts remain undecided on the legality of training data ingestion.⁴¹ Moreover, smaller publications lack the means to engage in litigation with Big Tech companies, especially given the odds against prevailing in court.

26 Perplexity Is a Bullshit Machine | WIRED

27 Randall Lane

28 Self-reported by Forbes

29 Local Newspapers Fact Sheet | Pew Research Center

30 <https://www.semrush.com/blog/zero-clicks-study/>

31 HuffPost - Wikipedia (Huffington Post until 2017)

32 <https://www.businessinsider.com/>

33 Allyn, ‘Online News Publishers Face “extinction-Level Event” from Google’s AI-Powered Search’; European Commission, ‘Commission Opens Investigation into Possible Anticompetitive Conduct by Google’.

34 On author

35 https://commission.europa.eu/index_en

36 Commission opens investigation into possible anticompetitive conduct by Google in the use of online content for AI purposes

37 Fastly Threat Insights Report

38 Without explicit blocks, one study showed AI scraping increased operating costs by 20 percent

39 Few Americans pay for news when they encounter paywalls

40 Digital Media Outlets Sue OpenAI for Copyright Infringement; US newspapers sue OpenAI for copyright infringement

41 Copyright Law in 2025: Courts begin to draw lines around AI training, piracy, and market harm

CloudFlare’s New Architecture of Consent

In the summer of 2025, a new approach to remedying these technical and structural threats to independent journalism emerged. It came from the giant digital infrastructure company Cloudflare, which dominates the global market for critical website security and performance optimization services, including bot detection.⁴² Cloudflare announced that going forward it would prevent AI bots from scraping the websites it manages unless their owners give permission.⁴³ This policy gives website owners the ability to choose whether they want to grant AI companies access to their content, making restriction—not permission—the default. Importantly, Cloudflare is also developing billing features⁴⁴ that allow website owners to set a price for the use of their content.

These features provide a potential technical solution to the failure of policymakers and the courts to determine whether and how to enforce copyright laws in the age of AI. Danielle Coffey,⁴⁵ president and CEO of the News/Media Alliance, which represents over 2,000 primarily U.S.-based publishers, said Cloudflare’s new features could be “particularly helpful for small and local publishers who lack the resources needed to combat sophisticated AI crawlers on their own.”⁴⁶

Training AI with indiscriminately scraped internet data risks giving them the ‘intelligence’ of a conspiracy theorist reasoning from false premises.

In September 2025,⁴⁷ Cloudflare added technical refinements that allow journalists and publishers to make distinctions in how Cloudflare treats different kinds of web crawlers seeking access to their content. Under its so-called Content Signals Policy,⁴⁸ for example, online content providers can grant web crawlers engaged in indexing for search engines open access to their websites.⁴⁹ This can serve the interests of publications because search engines bring new viewers through referrals. By contrast, publishers can block all web crawlers engaged in AI training,⁵⁰ which is justified so long as AI companies offer little or no compensation for their use of this data. Consequently, the effectiveness of this approach depends on companies

42 Cloudflare controls 82 percent of market for DDoS and bot protection

43 Cloudflare Just Changed How AI Crawlers Scrape the Internet-at-Large

44 Introducing pay per crawl: Enabling content owners to charge AI crawlers for access

45 <https://www.newsmediaalliance.org/about-us/danielle-coffey/>

46 On author

47 Cloudflare updates robots.txt for the AI era – but publishers still want more bite against bots

48 Ibid

49 Ibid

50 Ibid

engaged in search indexing or web crawling to accurately convey the purpose of their bots—differentiating between bots used for indexing versus crawling—and assumes that bot and human traffic can be distinguished from each other, which is becoming increasingly challenging amid the transition to agentic AI.⁵¹

Cloudflare has also introduced cryptographic bot verification and transparency dashboards.⁵² These collectively allow journalists and publishers to see who (or what) is crawling their websites, and how often. The same technology makes it possible to see how much referral traffic different kinds of web crawlers generate. According to TollBit data, AI search tools as a whole deliver 96 percent less referral traffic⁵³ to publishers than traditional Google searches.

Importantly, other technical features offered by Cloudflare give online media organizations the ability to set prices for different uses of their content. For example, Cloudflare has created a beta pay-per-crawl marketplace that lets a select group of media organizations charge per-page access fees to AI companies, with Cloudflare serving as the intermediary.⁵⁴ This private marketplace for journalistic and other editorial content is analogous to the private AdTech marketplace for advertising space, but is not yet an automated, programmatic stack.

The establishment of AI licensing marketplaces could be pivotal in enabling news organizations to receive fair compensation for their contributions. In essence, such exchanges could establish a robust framework that safeguards the intellectual property of journalists and other content creators, while simultaneously providing a regulated pathway for AI companies to access this content upon payment of a fee. However, despite the promising potential of this Cloudflare technology, there are significant potential drawbacks.

51 Studies show agents are increasingly mimicking human behavior

52 Forget IPs: using cryptography to verify bot and agent traffic

53 New Data Shows Just How Badly OpenAI And Perplexity Are Screwing Over Publishers

54 Introducing pay per crawl: Enabling content owners to charge AI crawlers for access

Potentials for Abuse

A key question is whether Cloudflare’s role as a middleman—the toll booth connecting buyers and sellers of journalistic content—will allow it to charge monopolistic commissions or fees. Through its pay-per-crawl market, Cloudflare will act in effect as a “merchant of record,” handling billing and distributing the earnings to the publisher.⁵⁵ But how much will it charge for this service? A Cloudflare spokesperson would not disclose any information about how much the company will take as an intermediary.⁵⁶ According to two people with direct knowledge, and remarks made by CEO Mathew Prince⁵⁷ in a podcast, Cloudflare appears set to take 30 percent of publishers’ newfound profits (on top of their usual website management fees).⁵⁸

This is akin to fees charged by the Google app store, which a jury found in 2023 to be an illegal monopoly.⁵⁹ It is also akin to the percentage extracted from app developers by Apple’s app store, which multiple courts and regulators have found to be uncompetitive.⁶⁰

Cloudflare’s market power over journalists and publications could be at least partially constrained by the rise of competing private exchanges.⁶¹ Months after Cloudflare’s announcement, Microsoft launched a pilot to compensate publishers when their intellectual property is used by AI companies and products, including Microsoft’s own Copilot.⁶² The company has not disclosed details about its two-sided Publisher Content Marketplace,⁶³ which will initially include only select U.S. publishers. Requests for additional information were denied except on background.⁶⁴

Yet even if Microsoft does develop this marketplace, it is not clear that this will create enough open competition to ensure fairness and efficiency. By way of analogy, Google competes with Facebook—and, increasingly, Amazon—in providing AdTech services, yet its oligarchic pricing power is still sufficient to drain billions in ad revenue from journalism.⁶⁵ Similarly, Amazon competes with both online and brick-and-mortar retailers, yet its market dominance still allows it to extract an effective commission rate of 45 percent from the third-party retailers on its platform.⁶⁶ And Spotify shares the marketplace for streaming music with Apple Music, yet these platforms have still used their monopolistic market power in ways that leave musicians with smaller and smaller shares of the value they create.⁶⁷ The same could happen to journalists if the emerging online marketplace for trade in repurposed journalistic content becomes dominated by a handful of unregulated corporate platforms.

55 Cloudflare Enables Websites To Block AI Bots With One-Click Solution; Cloudflare wants to help you to get paid for the use of your data in AI training | CIO

56 On author

57 Matthew Prince - Wikipedia

58 On author

59 Google to pay \$700 million in case over whether its app store is an illegal monopoly

60 The App Store Dilemma: Revenue “Cuts,” App Restrictions, and Payment Systems, Is There Unfair Competition? | Published in Houston Law Review

61 Backed by David Sacks, Garry Tan and Walter Isaacson, Created by Humans helps people license their creative work to AI models | TechCrunch ; Defined.ai: Transforming the AI Data Marketplace for an Ethical, Intelligent Future

62 Microsoft launches marketplace to compensate publishers for AI usage

63 Building Toward a Sustainable Content Economy for the Agentic Web | Microsoft Advertising

64 On author

65 One CJR study estimates that Google brings in \$12 billion from news-related summaries

66 Stacy Mitchel, ‘Amazon’s Monopoly Tollbooth in 2023, Independent Business’.

67 Musicians push back on dwindling payments from streaming services | PBS News

Cloudflare's expansion from an infrastructure provider to an AI data market maker or middleman raises serious competition concerns. By combining its dominance in web infrastructure with control over trade between content providers and AI companies, Cloudflare can effectively position itself as a de facto governance framework for data flows in the AI economy. Any default it sets shapes the bargaining environment for content access, freezing out publishers whose websites don't meet its technical specifications for data trading. Consequently, the company's shift from neutral conduit to active policy enforcer can significantly expand private power over public communications infrastructure.

Ceding these issues to Cloudflare does not ensure fair and open markets. Instead, it puts regulation in the hands of an unaccountable Silicon Valley corporation.

These power dynamics will be amplified by the growth of agentic AI. The internet as we once knew it was dominated by human users following hyperlinks to different websites.⁶⁸ As it evolves into a network in which humans primarily rely on AI to gather and manipulate digital data for them,⁶⁹ private control over data standards becomes even more problematic for journalists and other human content creators. Training and interacting with AI at scale requires that data be highly standardized so that it can be easily and economically transferred from human content creators into ingestible forms. This means that whoever controls those data standards can favor some human content creators while denying others the ability to receive compensation for their work from AI companies. In general, smaller, less technically sophisticated publications with low or nonexistent profits are most in danger of being frozen out of the marketplace for journalistic data.

Some may applaud Cloudflare for addressing a set of AI governance issues that Congress, the courts, and other public institutions have largely chosen to ignore. But ceding these issues to Cloudflare does not ensure fair and open markets. Instead, it puts regulation in the hands of an unaccountable Silicon Valley corporation. If Cloudflare, or any future intermediary, gains the same kind of control over the sale of journalistic content that Google exercises over AdTech markets, it will create similar threats to the economic foundations of journalism, including monopolistic extraction of middleman fees, self-dealing, and discrimination against weaker players.

In short, Cloudflare's move to embed consent, authentication, and monetization⁷⁰ into the internet's infrastructure is not merely a technical intervention; it is a form of private rule-making that affects the balance of market power between content creators and Big Tech, as well as the distribution of value in the emerging AI economy. Given the urgency of rebalancing market power in favor of rights holders, the challenge for policymakers is to ensure that the governance of AI data access does not become concentrated in the hands of a few digital infrastructure giants or governed solely by private protocols.

Fortunately, there is a better way. It is possible for independent journalists and publications to receive fair market compensation from AI companies for the use of their content. For that to happen, the public must establish market rules that are fair and open, including regulations governing interoperability, non-discrimination, and transparency, along with restraints on monopolization by self-dealing platforms.

68 https://en.wikipedia.org/wiki/History_of_the_web_browser

69 A majority of Americans now report using AI tools

70 Cloudflare Wades into the Battle Over AI Consent and Compensation | TechPolicy.Press

Policy Recommendations

To steer this ecosystem toward fairness, accountability, and socially beneficial competition, policymakers should prioritize the following actions:

1. Use antitrust and merger control to combat concentrated buyer power in the AI sector.

Publishers are unlikely to receive fair compensation for the use of their content by AI companies if there are only a handful of dominant AI companies competing as buyers of human-generated content. This dynamic will become more acute with the spread of AI agents, which, because of first-mover advantages, network effects, high barriers to entry, and other factors, are particularly prone to gaining monopolistic market shares. On the current course, a small number of dominant, vertically integrated generative AI providers will hold significant buyer power over weak and fractured content providers—even as these AI giants also control adjacent sectors like cloud storage, satellite, and cable infrastructure.⁷¹

2. Enforce bans on discrimination that historically have applied to all basic infrastructure.

Governments and competition authorities should also ensure publishers and smaller AI developers are not subject to discriminatory or opaque decisions by dominant players. This is especially crucial for programs like Cloudflare’s pay-per-crawl or Microsoft’s pay-per-use marketplaces, as well as cryptographic bot verification, which effectively function as quasi-regulatory systems. Regulations must not favor some players over others, and should include anti-retaliation clauses.

3. Establish transparency and due-process requirements for the use of data and bots by AI companies.

Cloudflare and other infrastructure providers should be required to set and disclose rules for web crawling, and publish clear, accessible information about how their systems operate—including criteria for bot verification, enforcement mechanisms, participation requirements, and revenue-sharing rules.

4. Encourage open, interoperable standards through public, multi-stakeholder processes.

To avoid fragmentation, regulators and public bodies should encourage multi-stakeholder efforts to modernize “robots.txt”⁷² and similar open-source standards for controlling access to websites by web crawlers and bots. Public-interest guidance and regulatory incentives can help ensure that emerging standards remain interoperable and accessible to small publishers, rather than becoming captured by a handful of powerful infrastructure providers or intermediaries. This includes supporting the development of machine-readable signals that distinguish between training, inference, retrieval, and other uses, and that are enforceable across platforms.

71 Market features in AI infrastructure: Competition in artificial intelligence infrastructure | OECD

72 Cloudflare updates robots.txt for the AI era – but publishers still want more bite against bots

5. Support rights-based licensing frameworks to reduce dependence on private gatekeepers.

Governments, foundations, and competition authorities should support the adoption and evolution of open licensing frameworks that enable competitive markets for AI training data rather than consolidating power in the hands of Content Delivery Networks, cloud providers, or Big Tech platforms.

6. Prevent market concentration in the AI content licensing marketplace.

Competition authorities should monitor whether exchanges like Cloudflare’s pay-per-crawl marketplace risk becoming gatekept chokepoints. If Cloudflare, Amazon, Microsoft, or any other major provider emerges as a “default” pathway for AI-ready content ingestion, policymakers should intervene to set clear rules before market power becomes entrenched. Abuses to guard against include anti-competitive data standards that prevent interoperability, exclusive dealing requirements, and vertical integration into adjacent sectors, such as cloud storage, publishing, and advertising. Special scrutiny is needed to prevent revenue-sharing practices that resemble the historic abuse of content creators in app stores and AdTech marketplaces.

7. Require baseline compliance and auditability for web crawlers.

Policymakers should require AI companies that train on or retrieve public web content to respect all machine-readable rules for data use by third parties, so-called “consent signals,” and maintain auditable logs of crawling, training, and data-use sources and patterns. Crawlers and bots should be required to identify themselves and their purpose and comply with instructions and terms of use. Compliance should be mandatory for firms operating in major digital markets, with penalties for evasion or misrepresentation.

8. Expand public-interest infrastructure for small and local publishers.

Extending infrastructure-level protections against hacking and viral attacks to vulnerable publishers can meaningfully rebalance market power. Policymakers, philanthropic institutions, and trade associations should expand such models—ensuring that local news outlets and independent creators have access to security tools, crawler controls, and licensing infrastructure that otherwise remain confined to large publishers. This means ensuring a competitive AI licensing market, access to competitive cloud services, and digital public infrastructure.⁷³

9. Make AI agents subject to fiduciary responsibility.

To further rebalance market power to the benefit of human content providers, and to protect human consumers from abuse, owners of AI agents should be required to disclose conflicts of interest. They should also be prohibited from engaging in practices such as self-preferencing or skewing advice in exchange for payments (kickbacks) from third-party payers.

10. Repeal or seriously reform Section 230.

Ultimately, ensuring a fair and open digital marketplace for journalists and other content providers requires closing a destructive loophole embedded in Section 230 of the Communications Decency Act.⁷⁴ This provision shields digital platform providers from accountability for illegal information they publish while allowing them to engage in discrimination against disfavored users or preference their own content.⁷⁵

73 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4766427

74 <https://www.congress.gov/crs-product/R46751>

75 Ibid

Conclusion

For years, the generative AI boom has proceeded in a legal vacuum. Copyright law remains unresolved on whether ingesting copyrighted content to train AI models constitutes fair use, fair dealing, or legitimate text-and-data mining exceptions.⁷⁶ Courts have yet to rule decisively, and litigation is slow. Policy proposals from data trusts and licensing collectives are still on whiteboards.

Into this vacuum stepped Cloudflare—not with legal doctrine, but with code. The company didn’t need to wait for Congress to pass legislation, or for judges to rule on the legality of their permission-less scraping. It simply altered its bot detection protocols, rewrote default settings for crawler access, and gave (some) publishers the choice to monetize their content.⁷⁷ It did in weeks what policymakers have been debating for years.

That’s impressive but also revealing. When a single company can implement a rights-respecting, consent-based access regime at scale, it is worth asking why public institutions have failed to do the same—and what type of power that gives a private actor. Is relying on private actors, however well-intentioned, to establish norms a sustainable path forward, given what we learned from the social media age? We have seen the repercussions of relying on voluntary industry initiatives over the past two decades, as a handful of companies have come to control and politicize the platforms we use to conduct business and civic affairs.⁷⁸

Taken together, these recommendations aim to align Cloudflare’s infrastructural innovations with a broader public-interest framework—one that preserves consent, ensures fair compensation, prevents market concentration, and anchors the governance of AI data access in transparent, accountable, and interoperable standards rather than private code. The future of the web need not be a choice between extractive AI and centralized infrastructure power; with the right policy interventions, it can be a competitive ecosystem where publishers, developers, and users all have agency in shaping the rules that govern the AI-driven digital economy.

These policy implications go to the heart of who governs the web in the AI age. They remind us that technically enabled consent isn’t enough—that the stewards of consent need accountability. As such, the Cloudflare initiative should spark a deeper debate about political and economic power in a rapidly privatizing internet.⁷⁹

Infrastructure providers like Cloudflare are no longer passive conduits (if they ever were⁸⁰); they are hosts, regulators, and gatekeepers who shape the digital economy through code rather than law. Whether this represents a necessary evolution or a dangerous privatization of public functions remains an open question, one that will be answered by how we regulate these intermediaries. The principles established now about consent, compensation, and control over digital content could govern how human creativity and machine intelligence interact for years to come. Getting these rules right isn’t just about saving news or the arts; it’s about preserving the economic incentives that drive human knowledge creation in the age of AI.

76 Copyright and Artificial Intelligence, Part 3: Generative AI Training Pre-Publication Version

77 Cloudflare Wades into the Battle Over AI Consent and Compensation | TechPolicy.Press

78 Center for Journalism and Liberty submits comment to FTC on Tech Platform Censorship and Threats to Press Freedom

79 The Privatized Internet Has Failed Us

80 <https://cmds.ceu.edu/article/2021-01-11/trump-v-tech-what-censorship-and-who-gets-do-it> and/or <https://blog.cloudflare.com/why-we-terminated-daily-stormer/> and/or https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4464300