August 3rd, 2026

The Honorable Kathy Hochul
Governor of New York State
New York State Capitol
Albany, NY 12224

RE: NY S9934A (Stealth Crawlers) – REQUEST FOR VETO

Dear Governor Hochul,

The undersigned civil society and digital rights organizations write to respectfully urge your veto of Senate Bill 9934A,[1] the New York Stealth Crawler Prohibition Act. While framed as a measure to protect local journalism, this legislation harms free expression and establishes a dangerous precedent by effectively deanonymizing and criminalizing automated access to the open web. By requiring all web crawlers to disclose their identity and explicit purpose, and by granting media outlets unchecked authority to obtain judicial subpoenas to unmask unidentified automated web traffic without any showing of misconduct or actual injury, this bill threatens digital privacy, compromises the foundational architecture of the internet, and will ultimately stifle the very independent journalism it seeks to protect.

[1] NY State Senate Bill 2025-S9934A. (2026) https://www.nysenate.gov/legislation/bills/2025/S9934/amendment/A

I. Constitutional and First Amendment Concerns

The Act fundamentally threatens First Amendment rights by forcing citizens to state who they are and how they will use public information just to access it. The bill targets public web data: information that news sites choose to make freely available without a paywall or log-in. Journalism providers already can and do restrict access to content via paywalls, log-ins, and various other means. By requiring any automated tool to disclose its identity and "all possible future uses" of web text the second it accesses a site, the law penalizes internet users based entirely on who they are and why they are browsing (§ 1752). This creates a dangerous system that gives publishers the power to act as a gatekeeper to public data. The right to gather and read information anonymously is a core pillar of a free press and an open society, yet this bill effectively dismantles it.[2]

The Supreme Court has long recognized that the First Amendment protects the right to receive information without first identifying oneself, striking down in *Lamont v. Postmaster General* a law that required recipients to self-identify before receiving certain mail.[3] Because online speech receives the same full protection as the printed page, New York cannot condition access to fully protected, freely published web content on a surrender of the reader's anonymity.[4]

Section 1752 also compels speech. It forces every crawler operator to affirmatively declare its identity, its software, its company, and "all uses and purposes that the content . . . could be used for" as a condition of access. Because the bill compels web crawlers to disclose specific content about themselves and what they plan to do with the content they crawl, the bill flatly violates the First Amendment.[5] Even under the most lenient First Amendment standard for mandated disclosures, the government may compel only purely factual, uncontroversial information, and only where the requirement is not unjustified or unduly burdensome.[6] A mandate to predict and disclose every conceivable future use of information is not a factual disclosure. It demands speculation about uses that do not yet exist, and as such is unduly burdensome on its face. Section 1752 is therefore unlikely to qualify for lenient First Amendment review, and could not survive it if it did.

Additionally, demanding that an internet user declare every single future use or application of the information they are accessing is a functional impossibility that serves as a legal trap. When a public-interest researcher, accountability watchdog, or investigative journalist gathers public text or information to study a trend or track a story, they often do not yet know where the facts will lead or how that information might be used in future analysis. Forcing them to predict and disclose

---

[2] Brian Carver. (2015) The Right to Read Anonymously. *Free Law Project.* https://free.law/2015/09/17/the-right-to-read-anonymously/; Carrie Robison and Daniel Ortner. (2023) Anonymous speech is as American as apple pie. *The Foundation for Individual Rights and Expression (FIRE).* https://www.fire.org/news/anonymous-speech-american-apple-pie

[3] *Lamont v. Postmaster Gen. of U.S.,* 381 U.S. 301, 305–06 (1965).

[4] *Reno v. Am. C.L. Union*, 521 U.S. 844, 868–70 (1997).

[5] *Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp. of Bos.*, 515 U.S. 557, 573–74 (1995); *Nat'l Inst. of Fam. & Life Advocs. v. Becerra*, 585 U.S. 755, 768–69 (2018).

[6] *Zauderer v. Off. of Disciplinary Couns. of Supreme Ct. of Ohio*, 471 U.S. 626, 651 (1985).

this information upfront puts an impossible burden on lawful data collection. Ultimately, S9934A sets a dangerous standard where citizens must ask a website owner for permission and surrender their privacy simply to read what should already be openly visible to the public.

II. Automated Access is Central to a Free and Open Internet

The ability to access and analyze publicly available information using automated tools is a core pillar of free expression and digital transparency. Automated access (commonly called crawling or scraping) powers the essential tools used to locate, preserve, and understand online information. This technical process is not a specialized loophole; it is simply a computer requesting and reading public web content.[7] For decades, researchers, universities, journalists, public interest groups, and ordinary users have relied on these automated tools to access and collect public web data. These tools have served as the lynchpins of the open web much longer than generative AI has even existed.

Anonymous crawling enables some of the most publicly beneficial uses of the open web. Researchers, journalists, and other watchdog groups use unidentified automated tools to gather the information necessary to hold powerful institutions accountable and protect the public.

Anonymous crawling has fueled important investigative journalism. For example, *The Markup*, a non-profit news site, used anonymous crawlers to investigate potentially anti-competitive practices by tech companies, such as Amazon's tendency to prioritize Amazon brands and Amazon-exclusive products over competitors with higher ratings.[8] The crawlers identified themselves as ordinary Firefox browsers to web servers, which allowed *The Markup* to understand how Amazon search results pages would appear to ordinary users.[9] Similarly, *ProPublica* used an automated tool designed to simulate an ordinary Amazon customer to reveal that the site steered shoppers to more expensive products over cheaper alternatives.[10] Anonymous crawling has even been instrumental in efforts to combat hate speech and misinformation on social media platforms like X. The Center for Countering Digital Hate, for example, used an anonymous crawler in research into hateful content and misinformation posted on X.[11]

Anonymous web scraping is also crucial for cybersecurity professionals, who use automated tools to monitor the web for information that helps them protect against malicious attackers.[12] Privacy

---

[7] Jamie Williams. (2018) 'Scraping' Is Just Automated Access, and Everyone Does It. *Electronic Frontier Foundation*. https://www.eff.org/deeplinks/2018/04/scraping-just-automated-access-and-everyone-does-it

[8] Adrianne Jeffries and Leon Yin. (2021) Amazon Puts Its Own "Brands" First Above Better Rated Products. *The Markup.* https://themarkup.org/amazons-advantage/2021/10/14/amazon-puts-its-own-brands-first-above-better-rated-products; Adrianne Jeffries and Leon Lin. (2020) How We Analyzed Google's Search Results. *The Markup*. https://themarkup.org/google-the-giant/2020/07/28/how-we-analyzed-google-search-results-web-assay-parsing-tool

[9] Adrianne Jeffries and Leon Yin. (2021) How We Analyzed Amazon's Treatment of Its "Brands" In Search Results. *The Markup*. https://themarkup.org/amazons-advantage/2021/10/14/how-we-analyzed-amazons-treatment-of-its-brands-in-search-results

[10] Julia Angwin and Surya Mattu. (2016) Amazon Says It Puts Customers First. But Its Pricing Algorithm Doesn't. *ProPublica*. https://www.propublica.org/article/amazon-says-it-puts-customers-first-but-its-pricing-algorithm-doesnt

[11] Center for Countering Digital Hate. (2023) Toxic Twitter: How Twitter Generates Millions in Ad Revenue by Bringing Back Banned Accounts. https://counterhate.com/research/toxic-twitter/

[12] Apurv Singh Gautam. (2025) Undercover Operations: Scraping the Cybercrime Underground. *SANS*. https://www.sans.org/blog/undercover-operations-scraping-the-cybercrime-underground

tools, including the Electronic Frontier Foundation's (EFF) Privacy Badger, also crawl sites anonymously to identify trackers without compromising user privacy.[13]

The capability to scrape anonymously is vital to civil society, independent accountability, and user autonomy. Watchdogs, journalists, and researchers rely on automated tools to monitor public information, track discrimination, identify security vulnerabilities, and report the news.[14] Crucially, many of these essential tools do not identify who deployed them—and for good reason. Important investigative efforts are often only possible when done anonymously. Without anonymity, website operators have the power to selectively block any user they wish, targeting dissidents, investigative reporters, or anyone who has not paid for a commercial license. Similarly, non-profit libraries like the Internet Archive use scraping to preserve historical records of the web before they disappear, while accessibility tools rely on automated retrieval to make public spaces readable for people with disabilities.[15] Automated comparison and aggregation tools also allow everyday citizens, small businesses, and startups to curate their own data feeds and find information without relying on corporate gatekeepers.[16] S.9934A threatens this entire ecosystem by allowing website operators to veto lawful public access, giving them the power to block not just bad actors, but competitors, researchers, dissidents, or anyone who has not paid for a license to view public text.

The Act also introduces impossible compliance thresholds by applying to any news website with more than 999 active users in New York (§ 1751.(2)(D)). For the average internet user, determining whether a smaller, lesser-known website qualifies under this definition is virtually impossible. This confusion is compounded by vague and ill-defined terminology; the Act completely fails to define critical terms like what it means to "deploy" a crawler or what constitutes "economic harm" (§ 1753). This ambiguity means that both an individual typing a search query into an AI search tool or browser, and the company that provides the tool to the public, could be legally classified as deploying a crawler, exposing everyday New Yorkers and small, open-source developers to massive liability.

More fundamentally, mandatory deanonymization threatens the core of the open web without actually addressing these technological and potential economic harms of AI-related crawling. The solution cannot be to dismantle the neutral protocols necessary to keep the web free and open. Enacting broad, reactionary restrictions on automated access will entirely fail to protect working journalists. Nor will it preserve baseline advertising revenues as AI tools shift the ways that people search the web. And behavior-based technical measures are an effective way to address the technological strain caused by bots that crawl at excessive rates. Addressing the harms of AI-

[13] Alexei Miagkov and Daly Barnett. (2023) Introducing Badger Swarm: New Project Helps Privacy Badger Block Ever More Trackers. *Electronic Frontier Foundation*. https://www.eff.org/deeplinks/2023/10/privacy-badger-learns-block-ever-more-trackers; Blacklight. (N.D.) *The Markup*. https://themarkup.org/blacklight.
[14] Andrew Crocker. (2022) Scraping Public Websites (Still) Isn't a Crime, Court of Appeals Declares. *Electronic Frontier Foundation.* https://www.eff.org/deeplinks/2022/04/scraping-public-websites-still-isnt-crime-court-appeals-declares
[15] Joe Mullin. (2026) Blocking the Internet Archive Won't Stop AI, But It Will Erase the Web's Historical Record. *Electronic Frontier Foundation.* https://www.eff.org/deeplinks/2026/03/blocking-internet-archive-wont-stop-ai-it-will-erase-webs-historical-record
[16] Corynne Mcsherry And Tori Noble. (2026) Comparison Shopping Is Not a (Computer) Crime. *Electronic Frontier Foundation.* https://www.eff.org/deeplinks/2026/04/comparison-shopping-not-computer-crime

related crawling requires policies that narrowly target the causes of these issues–without undermining free expression and the open web. The Act is anything but.

III. The Severe Cost to User Privacy and Independent Voices

The dangers of S.9934A are exacerbated by the sweeping enforcement mechanism, which weaponizes New York's legal system against anonymous reading and scraping. By granting news providers the authority to trigger broad ISP subpoenas to identify anonymous, automated web traffic, this bill forces anyone interacting with the public web via automated tools to forfeit their anonymity or face legal liability.

The enforcement mechanisms of this Act critically strips away basic due process protections, while giving certain websites unprecedented power to investigate potential critics or competitors. First, the bill levies crushing statutory fines of up to $15,000 per day per violation. The Attorney General can enforce these penalties based purely on a subjective finding of "satisfactory" evidence. Second, the Act grants unchecked subpoena power, requiring court clerks to issue identity-unmasking subpoenas based solely on a bare, unproven allegation by a news provider—no actual evidence linking the accused to website access is required. Consequently, Internet Service Providers (ISPs) would be legally compelled to violate the contractual and privacy rights of their customers due to this Act. This forces them to turn over sensitive user browsing history and incur heavy data-preservation burdens without any concrete showing by the government or any practical recourse to object.

IV. Existing Protections and Better Alternatives

The Legislature rushed to pass this extreme measure without exploring less harmful, highly effective means already at publishers' disposal. Journalism providers could prevent crawler access by utilizing existing technical firewalls, paywalls, or requiring user logins—in fact, many already do so. The legislature should additionally defer to established industry standard organizations like the Internet Engineering Task Force (IETF), which is actively developing robust protocols for bot authentication that protect website operators without restricting free speech.

Finally, robust legal remedies already exist to target malicious actors and protect publishers, including copyright protections and other authorities. Adding a new law specifically for news providers unnecessarily imposes undue burdens on lawful internet users.

V. Acknowledging Digital Media Realities Without Breaking the Web

Our organizations understand and acknowledge the very real economic and operational anxieties confronting traditional news publishers in the AI era. Anonymous crawlers have recently become a convenient boogeyman, but the technical reality is far more nuanced. We recognize that, due to recent innovations in AI, software now collects far more public web data than it used to, which can push web servers closer to their maximum capacity. If some crawlers or scrapers crawl too aggressively, they can strain a site's technical capacity and degrade performance for other users.

While this is a real problem, it is being effectively addressed with technical standards, commercial tools, and other measures that do not impede anonymous access to information. Additionally, there is a real risk that AI search tools may reduce traffic–and thus advertising revenue–to websites by summarizing search results, reducing the need or tendency to visit another website cited as a source of the information. However, anonymous crawlers are not the cause of this problem, and allowing publishers to unmask them will do little-to-nothing to mitigate it.

VI. Conclusion

Safeguarding the future of information sharing and our digital ecosystem requires preserving the legality of fundamental web utilities like scraping and crawling. Maintaining these open standards is essential to maintaining the openness of the web in the face of companies that continually seek new tools to control and monetize it, at the public's expense. Preserving the open web allows everyday citizens, small publishers, and local businesses to safely navigate the internet on their own terms—without the need to ask for permission from a gatekeeper that may try to control what they do with the information. And it preserves beneficial uses for research, journalism, innovation, and more.

Rather than enacting overly broad, reactive regulations that restrict public access to information and strip away digital anonymity and privacy, we urge you to protect the digital rights and privacy of all New Yorkers by vetoing Senate Bill 9934A.

Respectfully submitted,

Charlie Martial Ngounou, President, **AfroLeadership**
Dave Hansen, Executive Director, **Authors Alliance**
Becca Branum, Deputy Director, Free Expression, **Center for Democracy & Technology**
Joshua Tucker, Co-Director, **Center for Social Media, AI, and Politics @ NYU**
Rich Skrenta, Executive Director, **Common Crawl Foundation**
Himanshu Panday, Digital Anthropologist, **Dignity in Difference**
Rindala Alajaji, Associate Director of State Affairs, **Electronic Frontier Foundation**
Aviya Skowron, Head of Policy, **EleutherAI**
Mohamed ElGohary, Board Member, **Global Voices**
Murray Cox, Executive Director and Founder, **Inside Airbnb/Housing Justice Data Lab**
Brewster Kahle, Digital Librarian, **Internet Archive**
Ethan Zuckerman, Co-founder, **Media Cloud**
Fernando Bermejo, Executive Director, **Media Ecosystems Analysis Group**
Patrick Warren, Professor and Co-Director, **Media Forensics Hub**
Daniel Schwarz, Senior Privacy and Technology Strategist, **New York Civil Liberties Union**
Brandon Butler, Executive Director, **Re:Create**
Swapneel Mehta, Executive Director, **SimPPL**
Christy Grace Provines, Founder and Executive Director, **The 'MPOWER Project**