

A D D E N D U M
to
**TOO BIG TO PROSECUTE?: EXAMINING THE
AI INDUSTRY’S MASS INGESTION OF
COPYRIGHTED WORKS FOR AI TRAINING**

This Addendum is available at:

<https://www.govinfo.gov/content/pkg/CHRG-119shrg61891/pdf/CHRG-119shrg61891-add1.pdf>

Submitted by Chair Hawley:

Article III Project, letter	2
Artificial Intelligence Threatens Ownership of Online Content	5
Association of American Publishers (AAP), statement	10
CreativeFuture, letter	16
Motion Picture Association (MPA), letter	19
News Media Alliance, statement	23
Rumble, statement	28
Society of Composers & Lyricists (SCL), letter	30

Submitted by Ranking Member Durbin:

Center for AI and Digital Policy (CAIDP), statement	32
Copyright Alliance, statement	38
CreativeFuture, letter	16
News Media Alliance, statement	23
Society of Composers & Lyricists (SCL), letter	30

Submitted by Senator Klobuchar:

Artificial Intelligence Threatens Ownership of Online Content	5
---	---

Submitted by Senator Coons:

Motion Picture Association (MPA), letter	19
--	----

The Honorable Josh Hawley
212 Russell Senate Office Building
Washington, DC 20510

The Honorable Dick Durbin
711 Hart Senate Office Building
Washington, DC 20510

Dear Chairman Hawley and Ranking Member Durbin:

We, the undersigned organizations committed to technological innovation and strong intellectual property rights, commend you for holding a hearing on artificial intelligence (AI) companies' mass ingestion of copyrighted works to "train" large language models (LLMs).¹ As the Senate Judiciary Committee examines the impact of AI on creative industries, we urge the Committee, Congress, and the Trump Administration to reject calls by technology companies and their advocates for a blanket exemption from copyright law in training commercial AI models on copyrighted works.

The U.S. Copyright Office continues to develop policy on AI² and there is pending litigation across the country that will further clarify the application of copyright law on AI training.³ While AI poses new questions, copyright law has a long history of adapting to new technologies. As the Copyright Office and courts grapple with the complexities on a case-by-cases basis, it would be premature, reckless, and economically harmful for Congress or the Administration to put its thumb on the scale against rights holders.

Unauthorized and unlicensed use of copyrighted material for LLMs—or what AI companies call "training"—is already having a negative impact on many American creators. For example, Google is using publishers' content to serve AI-generated summaries that are reducing web traffic to original sources.⁴ This is already impacting the advertising revenue that publishers depend on for their survival. Worse, Google is leveraging its market power in the search market to give publishers an impossible choice. If publishers don't allow Google to use copyrighted content for an AI product that harms their bottom line, the search giant will remove publishers from search results entirely.⁵

In other words, Google is offering creators the choice between a slow or sudden death.

We are not against the development of AI or the use of copyrighted materials in the development of AI. Intellectual property (IP) is the foundation of creative and technological industries alike—copyright

¹ Too Big to Prosecute?: Examining the AI Industry's Mass Ingestion of Copyrighted Works for AI Training: Hearing Before the Subcomm. on Crime & Counterterrorism of the S. Comm. on the Judiciary, 118th Cong. (July 16, 2025), <https://www.judiciary.senate.gov/committee-activity/hearings/too-big-to-prosecute-examining-the-ai-industrys-mass-ingestion-of-copyrighted-works-for-ai-training>.

² U.S. COPYRIGHT OFFICE, *Copyright and Artificial Intelligence Part 3: Generative AI Training* (pre-publication version, 2025), <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>.

³ Baker & Hostetler LLP, *Case Tracker: Artificial Intelligence, Copyrights and Class Actions*, BAKERHOSTETLER, <https://www.bakerlaw.com/services/artificial-intelligence-ai/case-tracker-artificial-intelligence-copyrights-and-class-actions/> (last visited July 15, 2025).

⁴ Ryan Tracy, *Google Negotiates With Publishers Over AI News Content*, Wall Street Journal (June 27, 2025), <https://www.wsj.com/tech/ai/google-ai-news-publishers-7e687141>

⁵ "Currently, publishers using Google search claim they are unable to opt out from their material being ingested for Google's AI large language model training and/or being crawled for summaries, without losing their ability to appear in Google's search results." Jessica Davies, *Generative AI, Not Ad Tech, Is the New Antitrust Battleground for Google*, DIGIDAY (July 8, 2025), <https://digiday.com/media/generative-ai-not-ad-tech-is-the-new-antitrust-battleground-for-google/>, at para. 2 (last visited July 15, 2025).

protections and AI need not be at odds. For centuries, IP licensing has ensured that creators benefit economically from their works while enabling the diffusion of that creativity for future inventions and innovations.

Commercial LLM products benefit from copyrighted works while in many cases offering a substitute and competitor to those products. These LLMs are now capable of producing virtually any creative work that can compete with, and dilute the market for, human-created works like books, music, films, news writing, visual art, etc.⁶ These companies should simply have to obtain a license from copyright holders—as many already have⁷—before using those copyrighted works for LLM development.

Republican voters share these concerns. A recent poll⁸ of Trump voters found that:

- 87% believe AI developers should get permission before using creative works to train for-profit AI systems,
- 81% believe that an artist’s music or vocals shouldn’t be ingested or used by AI without the artist’s permission, and
- 84% believe that companies that want to use America’s most popular and valuable creative works should get permission and pay market rates to do so.

AI companies contend that obtaining licenses from copyright holders is impractical for the development of AI. This assertion is not only belied by these firms’ own licensing with copyright holders, but also the rapid advancement and development of technology to create licensing mechanisms and intermediaries to ensure that IP rights are upheld as AI is trained. While we understand why AI companies would rather pay nothing to creators, that desire for a free ride doesn’t make licensing “impractical.”

We also ask Congress to reject the cynical argument that the unlicensed and authorized use of copyrighted works for AI training is somehow critical to national security.⁹ There is a long tradition in this country of advocates for harmful policies using phony national security concerns as a crutch to prop up bad arguments. This case is no different. There is simply no evidence that pilfering human creativity to sell subscription-based chatbots is a prerequisite for the nation’s defense.¹⁰

We share the goals of the Trump Administration and Congress in ensuring that the U.S. leads the world in AI. IP rights are not a hindrance to that goal—in fact, respect for IP is critical to protecting U.S. innovations and ensuring our creativity industries continue to project American soft power across the globe. We urge the Committee and Congress to continue supporting the property rights enshrined in the Constitution and working to ensure that AI development and human creativity be mutually beneficial.

Sincerely,

⁶ “That training use is also not ‘transformative’ because its purpose is to enable the creation of works that compete with the copied works in the same markets – a purpose that, when pursued by a for-profit company like Meta, also makes the use undeniably ‘commercial.’” Brief for Copyright Law Professors as Amici Curiae Supporting Plaintiffs at 3, *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417-VC (N.D. Cal. Apr. 11, 2025), ECF No. 525, <https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.525.0.pdf> (last visited July 15, 2025).

⁷ Sara Guaglione, *2024 in Review: A Timeline of the Major Deals Between Publishers and AI Companies*, DIGIDAY (Dec. 27, 2024), <https://digiday.com/media/2024-in-review-a-timeline-of-the-major-deals-between-publishers-and-ai-companies/> (last visited July 15, 2025).

⁸ Human Artistry Campaign, *ROC Survey*, <https://www.humanartistrycampaign.com/roc-survey> (last visited July 15, 2025).

⁹ Rebecca L. Grant, *AI, Fair Use and the Arsenal of Democracy*, REALCLEARDEFENSE (July 7, 2025), <https://www.realecleardefense.com/articles/2025/07/07/ai-fair-use-and-the-arsenal-of-democracy-1121032.html> (last visited July 15, 2025).

¹⁰ Mitch Glazier, *To Beat China in the AI Race We Have to Stop the Theft of American IP*, DC JOURNAL (May 13, 2025), <https://dcjournal.com/to-beat-china-in-the-ai-race-we-have-to-stop-the-theft-of-american-ip/> (last visited July 15, 2025).

Mike Davis
Founder & President
Article III Project

Otto Heck
Chief Operating Officer
Internet Accountability Project

Aiden Buzzetti
President
Bull Moose Project

Jeff Mazzella
President
Center for Individual Freedom

Eric Teetsel
Chief Executive Officer
Citizens for Renewing America

Gerard Scimeca
Chairman
Consumer Action for a Strong Economy

Joel Thayer*
President
Digital Progress Institute

SOCIETY COMMENTARY

To Fight Fake News, We Must Rein in AI

Joel Thayer June 25, 2025



Joel Thayer

Joel Thayer is the president of the Digital Progress Institute.

The promise of artificial intelligence is undoubtedly exciting and possesses the ability to further human flourishing. AI is also transforming and challenging almost every traditional notion. But with great change comes potentially damaging market distortion. Specifically, AI may be creating an acute crisis to our media landscape and even how we access information as it continues to degrade our concept of intellectual property rights.

The central issue concerns whether any person owns their own words or thoughts if they put them online. Big Tech companies don't seem to think so. In fact, they seem to view these property rights as an inhibitor to AI and that they should be free to use all data (including those that are clearly copyrighted) with reckless abandon.

Indeed, the AI era may mean the end of owning your own content once it can be accessed on the web. As one commenter describes, “[c]ontent creators’ work has been used or ‘scraped’ [by large AI companies] without permission and without compensation.” This implies that they can use your face, voice, words, or written thoughts in any way they choose, and you wouldn't be entitled to a dime, while they rake in billions on your likeness and work.

The Daily Signal depends on the support of readers like you. Donate now

But this issue is more than just about sour grapes over a licensing fee between AI companies and the press. The underlying issue could have grave implications.

To understand why, let's start with our nation's priors. Securing intellectual property rights was so essential to our democracy that the Framers found it prudent to incorporate it in Article I of our Constitution. There's good reason they did. The Framers viewed intellectual property akin to physical property because providing exclusive (albeit limited) rights for such works "enable[s] th[e] marketplace for creative works."

What's more, copyright is critical to break up the centralization over news, while creating a more sustainable press industry at large. Indeed, smaller, conservative publications, like The Daily Wire, the Daily Caller, and Blaze Media, use paywalls to justify the massive investments and extreme risks it takes to compete with larger organizations, like The New York Times and The Washington Post. Better yet, these outlets, armed with copyright protections, are outpacing the larger providers.

This is a net positive for society because healthy competition in media markets means more diversity of opinion. This is especially needed as larger, metropolitan papers continue to run local news outlets out of business.

However, Big Tech AI companies can upend all these positive gains entirely if the status quo continues. In no uncertain terms, if AI is permitted to recklessly crawl the internet,

even past paywalls, then we've effectively stripped these smaller companies of their competitive advantage and, worse, solidified these Big Tech companies to be the ultimate distributor of our information for good.

Given Big Tech's penchant for censoring content that runs counter to their worldview, this is a terrible outcome. Especially when one considers that AI is even replacing actual journalists. As Forbes reported, "One of the increasingly common ways that we're seeing AI being used with journalism and news is with automated news writing." This will increasingly be a huge problem for those seeking objective journalism given that, as the Brookings Institution described, "there is a clear left-leaning political bias" to AI-generated responses. If this becomes more ubiquitous, so too will media bias.

To combat this, we must maintain human reporting to counterbalance the rapacious efficiency and proliferation of AI-generated stories. A way to do that is to give traditional news agencies a fighting chance by allowing them to protect their intellectual property rights. One that prohibits Big Tech from using their articles without their permission or at the very least requires them to ask the authors to opt in or otherwise not go behind their paywalls.

Big Tech AI companies, like Google and Meta, argue that unfettered use to this data and the degradation of our basic understanding of intellectual property rights is paramount to ensure our dominate position in AI over China.

But that's complete hokum, given their blind acquiesce to China at every turn, especially on AI. Google, for example, negotiated "with the Chinese government to develop a censored search engine known as Project Dragonfly." There's strong evidence to suggest that Meta may have provided the Chinese Communist Party with "AI tools" to surveil its citizens. Apple has a multibillion-dollar deal with the CCP to degrade its encryption on Mainland China and has called China one of the most important partners in its supply chain.

They are hardly our national AI champions.

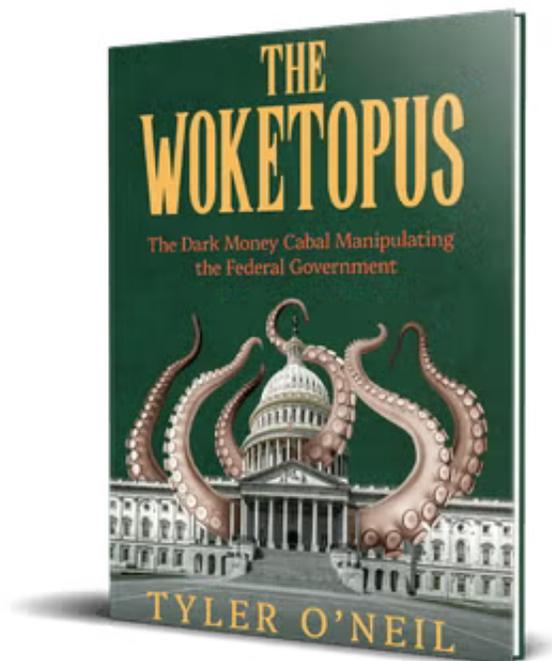
Let's be clear, Big Tech's actions here have nothing to do with innovation. Rather, Big Tech companies understand that degradation of intellectual property rights only helps them obtain more market dominance over our data and information. To ensure a diversity in our news, our government must turn its attention to ensuring that intellectual property is maintained and not eliminated by AI.

We publish a variety of perspectives. Nothing written here is to be construed as representing the views of The Daily Signal.

Related Posts:

- 1. Coalition Calls for Lawmakers, Tech Industry to Make AI Safer for Kids**
- 2. Artificial Intelligence Regulation Hotly Debated on Hill**

Read the first chapter of *The Woketopus* right now for FREE



Today, even with President Trump's victory, leftist elites have their tentacles in every aspect of our government.

The Daily Signal's own Tyler O'Neil exposes this leftist cabal in his new book, *The Woketopus: The Dark Money Cabal Manipulating the Federal Government*.

In this book, O'Neil reveals how the Left's NGO apparatus pursues its woke agenda, maneuvering like an octopus by circumventing Congress and entrenching its interests in the federal government.

You can read the first chapter of this new book for FREE in this eBook, *The Woketopus: Chapter One* using the secure link below.



Statement of
THE ASSOCIATION OF AMERICAN PUBLISHERS
before the
U.S. SENATE JUDICIARY COMMITTEE
SUBCOMMITTEE ON CRIME AND COUNTERTERRORISM
Hearing on
“Too Big to Prosecute?: Examining the AI Industry’s
Mass Ingestion of Copyrighted Works for AI Training”
July 16, 2025

The Association of American Publishers respectfully submits this written testimony for the record in connection with the Senate Judiciary Committee Subcommittee on Crime and Counterterrorism’s hearing titled *Too Big to Prosecute?: Examining the AI Industry’s Mass Ingestion of Copyrighted Works for AI Training*.

We commend the Committee for holding this important hearing to spotlight the threats to America’s continued leadership on both copyright and AI. It should be of great concern to all Americans that certain AI companies, some of the largest and wealthiest technology ventures in the world, chose to cut corners, violate federal copyright laws, misappropriate the intellectual property America’s authors, artists, and creators, and then argue that they are too big—or too important—to be held accountable for their misdeeds. The so-called ethos of “move fast and break things” at any cost is counter to American values and should not be permitted to continue.

The Association of American Publishers (AAP) is the industry representative for book, journal, and education publishers in the United States on legal and policy issues, including, especially, the fast-moving opportunities and challenges relating to artificial intelligence.

American publishers drive the most economically successful publishing market in the world, producing revenue of nearly 30 billion dollars annually in the United States alone.¹ We are a key part of the broader U.S. copyright industries, many of which produce news, education, and entertainment content based on

¹ AAP StatShot Annual Report: Publishing Revenues Totaled \$29.9 Billion for 2023, ASS’N OF AM. PUBLISHERS (AAP) (Aug. 22, 2024), <https://publishers.org/news/aap-statshot-annual-report-publishing-revenues-totaled-29-9-billionfor-2023/>.

books, and which collectively add more than \$2.09 trillion in annual value to U.S. gross domestic product.²

Books are the original copyright industry. When Congress passed the first Copyright Act in 1790, just three years after enshrining the exclusive rights of authors in the Constitution, it secured protection for maps, charts, and books.

And now, over two centuries later, books are a critical component in the most cutting-edge generative AI technologies. The literary and informational works published by AAP members are especially valuable as training materials for large language models. They are high quality, professionally written and edited, and carefully reviewed for reliability.

Big AI developers could have engaged with publishers to use their works as training data, just as they engage with chip manufacturers for computing resources and engineers for talent and ingenuity.

But they didn't.

The largest, wealthiest companies in the world instead stole the work of countless authors, artists, and creators to train their models.

Worse, in the case of books and journals, they didn't even scrape them from the public web—which is already infringing unless the copyright owner has provided explicit permission.

Instead, they scoured the darkest regions of the web and connected with notorious foreign pirate sites to illegally acquire millions of copyrighted works—widespread theft on an unprecedented scale, by commercial actors.

These sites include Library Genesis (LibGen), Z-Library, and Anna's Archive, and have long been a scourge on American investments in intellectual property. Many of these pirate sites have been identified by the U.S. government as notorious agents operating against U.S. interests³ and targeted by U.S. law enforcement.⁴ Their IP theft often subsidizes other criminal activities beyond the jurisdictional reach of U.S. law enforcement.⁵

² Robert Stoner and Jéssica Dutra, Copyright Industries in the U.S. Economy: The 2024 Report, INT'L INTELL. PROP. ALL. (IIPA) (February 2025), https://www.iipa.org/files/uploads/2025/02/IIPA-Copyright-Industries-in-the-U.S.-EconomyReport-2024_ONLINE_FINAL.pdf. Beyond these important economic contributions, an independent and thriving publishing industry supports the nation's political, intellectual, and cultural systems.

³ Office of the U.S. Trade Representative, USTR Releases 2024 Review of Notorious Markets for Counterfeiting and Piracy, USTR.gov (Jan. 8, 2025), <https://ustr.gov/about-us/policy-offices/press-office/pressreleases/2025/january/ustr-releases-2024-review-notorious-markets-counterfeitng-and-piracy>.

⁴ Kevin Schaul et al., Inside the Secret List of Websites that Make AI like ChatGPT Sound Smart, WASH. POST (Apr. 19, 2023), <https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/>; Ernesto Van der Sar, Z-Library Aftermath Reveals That the Feds Seized Dozens of Domain Names, TorrentFreak (Nov. 7, 2022), <https://torrentfreak.com/z-library-aftermath-reveals-that-the-feds-seized-dozens-of-domain-names-221107/>.

⁵ INTERPOL, Digital Piracy (last visited July 15, 2025), <https://www.interpol.int/en/Crimes/Illicit-goods/Shop-safely/Digital-piracy> (“the criminals behind these pirate sites can be part of organized crime groups. They can use the proceeds to fund other illegal

Litigation against major US AI developers have revealed the sheer scale of this problem. Meta, which at first sought to license copyrighted works from publishers,⁶ abandoned those efforts and instead downloaded “at least 81.7 terabytes of data across multiple shadow libraries through the site Anna’s Archive, including at least 35.7 terabytes of data from Z-Library and LibGen.”⁷ Anthropic, according to another court decision, “ha[d] many places from which’ it could have purchased books, but it preferred to steal them to avoid ‘legal/practice/business slog.”⁸ In total, the company “pirated over seven million copies of books.”⁹ OpenAI has also allegedly used LibGen to source illicit training materials.¹⁰

Foreign AI companies have used pirate repositories to expropriate the value of American copyrighted works. For example, DeepSeek researchers have publicly acknowledged that they used content from Anna’s Archive (a site that aggregates records from major shadow libraries like Z-Library, Sci-Hub, and LibGen) to train at least one of its models.¹¹ Anna’s Archive itself has publicly claimed to have given “high-speed access” to its illegal collection of more than 140 million copyrighted texts to companies in China, Russia, and elsewhere.¹²

Mass Ingestion of Copyrighted Works for AI Training is Not Permitted by Fair Use

The U.S. Copyright Office concluded that “making commercial use of vast troves of copyrighted works to produce expressive content that competes with them in existing markets, especially where this is accomplished through illegal access, goes beyond established fair use boundaries.”¹³ We agree. Publishers regularly rely on fair use, by, for example, displaying copyrighted images for documentary purposes in historical works,¹⁴ quoting from copyrighted works for critical biographies,¹⁵ and creating critical retellings of fictional works.¹⁶ Such uses advance the goals of copyright without prejudicing the interests of the original author.

activities, such as illegal online gambling, online sexual exploitation, drug trafficking, human trafficking, arms smuggling, and money laundering.”).

⁶ *Richard Kadrey et al. v. Meta Platforms, Inc.*, No. 3:23-cv-03417-VC, slip op. (N.D. Cal. June 25, 2025).

⁷ Ashley Belanger, “Torrenting from a corporate laptop doesn’t feel right”: Meta emails unsealed, *Ars Technica* (Feb. 6, 2025), <https://arstechnica.com/tech-policy/2025/02/meta-torrented-over-81-7tb-of-pirated-books-to-train-ai-authors-say/>.

⁸ *Bartz et al. v. Anthropic PBG*, No. 3:24-cv-05417-WHA, slip op. (N.D. Cal. June 23, 2025).

⁹ *Id.*

¹⁰ “The Unbelievable Scale of AI’s Pirated-Books Problem,” *The Atlantic*, Alex Reisner, Mar. 20, 2025, <https://www.theatlantic.com/technology/archive/2025/03/libgen-meta-openai/682093/>.

¹¹ Haoyu Lu et al., *DeepSeek-VL : Towards Real-World Vision-Language Understanding* (Mar. 2024), available at <https://arxiv.org/pdf/2403.05525>.

¹² Anna’s Archive, *Copyright Reform is Necessary for National Security* (Jan. 31, 2025), <https://annas-archive.org/blog/ai-copyright.html>.

¹³ U.S. Copyright Office, *Copyright & Artificial Intelligence: Part 3: Generative AI Training* (Pre-Publication Version, May 9, 2025) at 107, available at: <https://copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>.

¹⁴ See *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605 (2d Cir. 2006).

¹⁵ See *New Era Publ’ns Int’l, ApS v. Carol Publ’g Grp.*, 904 F.2d 152 (2d Cir. 1990).

¹⁶ See *SunTrust Bank v. Houghton Mifflin Co.*, 268 F.3d 1257 (11th Cir. 2001).

The same cannot be said for AI training, which expropriates the expressive value of the content to be used to synthesize new expression that serves the same purpose as the original and competes with or impairs the market for the original works. Allowing AI developers to do so, without permission or compensation, clearly runs contrary to the statutory framework created by Congress.

American AI Leadership Must Be Built on Licensing

Lawful commerce, in the form of licensing copyrighted works for AI training, should be at the heart of the US AI industry, rather than criminal conduct. Licensing promotes innovation, supports the vital creative sector, and advances the public interest.

- **Licensing is pro-competitive.** It is not a one-size-fits-all solution—pricing and terms can be based on licensee size, revenue, and purpose. Contrary to claims that licensing would favor big companies over small, it is the world that ignores copyright that is a winner takes all world, where the competitive edge goes only to the largest companies that can devote the most resources to collecting and building training datasets.
- **Licensing creates incentives for high-quality data.** Companies will have commercial motives and build expertise in all the work required to transform raw text, images, and other media into clean, reliable training datasets. Again, without a legal framework built around copyright, only the largest companies will have the time and money to do this work, and they will be disincentivized from sharing datasets to prevent free riding.
- **Licensing preserves the open web:** If copyright law fails to protect creative labor, creators and publishers will resort to practical and technological barriers to access like paywalls, obfuscation, and removal from public platforms. The result will be a fragmented and impoverished web, depriving not just future models of the data needed to train responsibly but the general public from the benefits of an open internet.
- **Licensing rewards authors and publishers, consistent with the constitutional goals of copyright:** Creators of books, journalism, photography, and other expressive works are essential to the cultural and informational commons. Licensing allows them to share in the commercial success of AI, providing income streams that support continued creation. The public continues to benefit from their creative work, and the next generation of AI benefits from a renewable source of high-quality training materials.

Piracy Undermines AI Leadership in Innovation and Creativity

The unchecked use of piracy sites is detrimental. It improperly normalizes commercial-scale piracy, contrary to clear U.S. law enforcement objectives worldwide. It perpetuates the viral harm of digital piracy. Pirate repositories compound the challenges already caused to copyright owners from infringement arising out of the unauthorized use of copyrighted works to train AI models. The availability of copyrighted works on the black market undermines the negotiating leverage of copyright owners in the new AI markets. Moreover, the illicit distribution of copyrighted works eliminates the ability of copyright owners to employ technological measures to mitigate unauthorized training.

This conduct is all the more troubling because it contravenes the framework Congress established in 1998 by adding the DMCA to the Copyright Act.¹⁷ With the DMCA amendments, Congress sought to establish a robust digital marketplace by ensuring appropriate safeguards for works made available online, including copyright owners' ability to rely on DRM protections in distributing electronic copies of their works. As explained in the accompanying Senate Report:

Due to the ease with which digital works can be copied and distributed worldwide virtually instantaneously, copyright owners will hesitate to make their works readily available on the Internet without reasonable assurance that they will be protected against massive piracy. [The DMCA] ... will facilitate making available quickly and conveniently via the Internet the movies, music, software, and literary works that are the fruit of American creative genius.¹⁸

In addition to avoiding the inconvenience and expense of licensing and compensating copyright owners for the commercial use of their content, AI developers using pirate sites seek to evade technological protections that are essential to a functioning online marketplace for copyrighted works. This is manifestly at odds with the mandate of Congress in adopting the DMCA, the objectives of the Copyright Act, and the Constitution's directive to protect the exclusive rights of authors in their works.¹⁹

Recommendations

Pirate site usage rewards criminal conduct, stifles innovation, and places the US creative sector, which is the global gold standard, at risk.

Congress must lead on this issue and can reinforce its historic role in securing the exclusive rights of authors through the following ways:

- Denounce the training of AI models on pirate sites and content and nullify any incentives or benefits that would otherwise flow to such AI developers from the U.S. government.
- Prohibit the procurement and use of AI models that have been trained on pirate sites and content by government agencies when contracting for AI services.
- Prioritize the use of law enforcement and trade tools to combat pirate repositories which allow foreign competitors to illegally expropriate American intellectual property.
- Direct the DOJ and FTC to use their unfair competition authority against companies that use and support pirate repositories for training.

¹⁷ DMCA, Pub. L. No. 105-304, 112 Stat. 2860 (1998).

¹⁸ S. Rep. No. 105-190, at 8 (1998).

¹⁹ U.S. Const. art. I, § 8, cl. 8 ("The Congress shall have Power ... To promote the Progress of Science ... by securing for limited Times to Authors ... the exclusive Right to their respective Writings."); see also *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 558 (1985) ("The economic philosophy behind the clause empowering Congress to grant patents and copyrights is the conviction that encouragement of individual effort by personal gain is the best way to advance public welfare through the talents of authors and inventors.") (internal quotes omitted).

- Draft and pass legislation creating a robust transparency requirement on AI developers to disclose the works used for training to provide rights holders with the information necessary to protect and enforce their rights.

Conclusion

America can't lead with AI that is built on a corrupt foundation. Sacrificing American leadership on copyright protection will sacrifice one of our primary competitive advantages in AI and innovation. AI can and should be built on foundations that are lawful, competitive, and economically sound. That means ensuring that AI companies procure copyrighted works for AI training through lawful, voluntary licensing with permission and compensation, and never from notorious foreign piracy sites. Otherwise, Congress must ensure that AI companies are *not* too big to prosecute and *will* be held accountable for the theft of America's creative intellectual property.

Thank you for the opportunity to share these views with the Subcommittee on Crime and Counterterrorism.

15 July 2025

The Honorable Josh Hawley
Chairman
Senate Crime and Counterterrorism
Subcommittee
United States Senate
Washington, DC 20510

The Honorable Richard Durbin
Ranking Member
Senate Crime and Counterterrorism
Subcommittee
United States Senate
Washington, DC 20510

Dear Chairman Hawley, Ranking Member Durbin, and esteemed members of the Crime and Counterterrorism Subcommittee:

Thank you for your continued focus on internet platform accountability. We ask that you include this letter in the record of your hearing on July 16: *Too Big to Prosecute?: Examining the AI Industry's Mass Ingestion of Copyrighted Works for AI Training*.

We write on behalf of the almost ten million Americans who make their living in the core copyright industries. CreativeFuture is a nonprofit coalition that advocates for strong copyright protections for the millions of Americans working in the business. We depend on copyright to bring all of us countless moments of inspiration, learning, and joy.

Our communities span this country and the world. We are artists, craftspeople, writers, producers, guilds and unions, talent agencies, studios, independents, music labels, and mostly small businesses. 92% of the businesses in the film and television industry alone employ fewer than 10 people. The massive harm caused by the global theft of our creative works is devastating.

America's creative industries are a significant economic driver, contributing more than \$2 trillion to GDP. We are a leading source of exports, outpacing other major U.S. industries such as aerospace, agriculture, and pharmaceuticals, according to the International Intellectual Property Alliance's 2024 [Copyright Industries in the U.S. Economy](#) report. When consumers in other countries enjoy American-made creativity, job creation and economic growth occur here.

We have long applauded the Senate Judiciary Committee's scrutiny of Big Tech and its abuses. While past hearings highlighted harms caused by social media platforms and intellectual property theft made possible by loopholes and shortcomings in our current laws governing the internet, your hearing today brings needed attention to the ways in which the rapid rise of Generative Artificial Intelligence (AI), driven by some of the same large platforms, seems to be repeating the harmful patterns we saw in the early days of the internet.

As the consumer experience moves away from internet search and toward AI chatbot interactions, AI-enabled search, and generative AI, the next frontier of piracy lies with AI. We have already witnessed widespread infringement in the creation of these AI language models and chatbots. Indeed, it is already well-documented that Meta and other AI developers have knowingly trained their systems on notorious piracy websites and repositories. Worse, these systems are then directing consumers to existing pirate sites. As well, these models are generating outputs that result in further copyright infringement.

There have already been lawsuits filed against [Meta](#)¹, [OpenAI](#)², and [Anthropic](#)³ in which all have been accused of copyright infringement in the training of their their models. In an unredacted filing from the Meta lawsuit, the court revealed internal discussions at the company about the ethics of using the pirate library and the company's choice to move ahead regardless.

A study commissioned by the U.S. Chamber of Commerce estimated that global streaming video piracy costs the U.S. economy up to **\$71 billion** in lost revenue *each year*. This is a staggering loss for the creative industries, which now provide more legitimate platforms on which to access authorized versions of audiovisual works than ever before.

This creative economy, the world's best, continues to be under siege by digital piracy aided and abetted by some of our largest American corporations. These same corporations are also some of those pioneering AI today.

These AI companies are moving at breakneck speed to ingest copyrighted material without permission from the copyright owners. At the same time, they are coming to you, our lawmakers, asking for the types of immunities that spawned the unaccountable online environment that we have today and that we have endured for too long. Once again, Big Tech insists that the ordinary accountability all other sectors abide by should not apply to them.

We commend you for turning your attention to these AI companies and their unauthorized ingestion of copyrighted material to build their businesses.

The creative communities have long been focused on how the absence of accountability has led to the proliferation of theft of its copyrighted content – to the detriment of creatives and the U.S. economy. But as your hearing suggests, we are now battling an entirely new frontier of infringement – by American AI companies.

Since 2019, members of the Senate Judiciary Committee and other members of Congress have written several letters to the platform companies, inquiring about their ability and willingness to work voluntarily toward solutions to curb piracy: Alphabet, Meta, X, and Verisign among them. Nevertheless, American consumers are still able to easily access piracy apps and websites, as well as pirated content on those companies' platforms or websites, such as YouTube and Facebook.

This, even *after* those companies told Congress not only that they diligently work to remove piracy but that they would redouble their efforts. These companies cannot be trusted to keep their word.

The problem you are addressing threatens a repeat of the mistakes of the past. A legal framework that fails to sufficiently hold internet platforms to the same standards of accountability as every other American industry has resulted in the piracy problems we have spent over two decades fighting. If we

¹ <https://www.wired.com/story/new-documents-unredacted-meta-copyright-ai-lawsuit/>

² <https://www.npr.org/2025/03/26/nx-s1-5288157/new-york-times-openai-copyright-case-goes-forward>

³ <https://apnews.com/article/anthropic-ai-fair-use-copyright-pirated-libraries-1e5cece51c2e4bd0bb21d94de2abb035>

chart the same course with today's AI companies, how can we expect any different result? In fact, with the scale and sophistication of AI technology, we can be sure the problems will be exponentially worse.

When companies like these AI behemoths fail to respect the rights of creatives in the training of their AI models, they deprive us of the critically important exclusive rights secured by copyright under the U.S. Constitution.

We applaud the Committee for its attention to this pressing issue. We stand ready to work with you to ensure AI companies respect existing protections for the rights of the creative communities. Contrary to claims by these AI companies, respecting copyrighted material need not undermine their ability to build cutting-edge AI models.

Thank you for your consideration.

Respectfully,



Ruth Vitale

cc: Senate Judiciary Committee



MOTION PICTURE ASSOCIATION

July 16, 2025

The Honorable Josh Hawley
Chairman, Subcommittee on Crime and Counterterrorism
U.S. Senate Committee on the Judiciary
226 Dirksen Senate Office Building
Washington, DC 20510

The Honorable Richard Durbin
Ranking Member, Subcommittee on Crime and Counterterrorism
U.S. Senate Committee on the Judiciary
226 Dirksen Senate Office Building
Washington, DC 20510

Chairman Hawley and Ranking Member Durbin:

Thank you for holding this important hearing, “Too Big to Prosecute?: Examining the AI Industry’s Mass Ingestion of Copyrighted Works for AI Training.” The Motion Picture Association, Inc. (“MPA”) serves as the global voice and advocate of the motion picture, television, and streaming industries. MPA works in every corner of the globe to protect its members’ content across all screens, defend the creative and artistic freedoms of storytellers, and support innovative distribution models that expand viewing choices for audiences worldwide.¹

The American motion picture and television industry is a global economic and cultural powerhouse, distributing films and television shows in more than 130 countries, projecting American soft power and promoting our values of freedom and opportunity. In 2023, the enduring value and global appeal of U.S. entertainment translated to \$22.6 billion in audiovisual exports (including a \$15.3 billion trade surplus), supported 2.3 million jobs and \$229 billion in total wages, and delivered \$21 billion in payments to more than 1940,000 local business across the country.²

¹ The MPA’s member studios are: Amazon Studios LLC; Netflix Studios, LLC; Paramount Pictures Corporation; Sony Pictures Entertainment Inc.; Universal City Studios LLC; Walt Disney Studios Motion Pictures; and Warner Bros. Entertainment Inc.

² See Motion Picture Ass’n, *The American Motion Picture and Television Industry | Creating Jobs, Trading Around the World*, at 1-2 (2025), https://www.motionpictures.org/wp-content/uploads/2025/01/MPA_Economic_contribution_US_infographic.pdf.

The United States has long set the global gold standard for intellectual property (“IP”) protection, and for effective enforcement against unscrupulous actors that steal American creativity and undermine the creative economy and the jobs it produces.

Promoting AI Innovation and Protecting IP

MPA supports efforts for the United States to promote AI innovation, but we will not win the AI race with China—or any nation—by gutting our country’s global intellectual property dominance in the process. Indeed, the American companies at the forefront of AI got there themselves with the benefit of America’s robust history of respect for and defense of intellectual property. We win the race by safeguarding America’s intellectual property, building AI systems the world can trust and rely upon, and prioritizing the nation’s global leadership in *both* the creative industries and AI innovation. David Sacks, chair of the Presidential Council of Advisors for Science and Technology has himself raised alarm that China may have stolen U.S. data to produce DeepSeek.³ And White House Office of Science and Technology Policy Director Michael Kratsios has observed that “we must safeguard U.S. intellectual property” in order to promote America’s technological leadership.⁴

Respect for and protection of copyrighted works and other IP must remain a central pillar of our nation’s artificial intelligence (“AI”) policies, both in how we facilitate innovation domestically and in how we address the global race for AI leadership. As the first Trump Administration stated, “respect for intellectual property” is a core pillar in promoting “AI with American values.”⁵ MPA applauds the Trump Administration for championing American industry and American workers, and for its stated objective of promoting U.S. leadership in establishing a “gold standard” for AI innovation that works for all Americans.

Fueled by uniquely American human creativity and technological innovation and undergirded by the U.S. Constitution’s protection of both free speech and intellectual property,⁶ the U.S. film, television, and streaming industry delivers enormous economic value, drives innovation, promotes free expression, and serves as a global ambassador for the nation’s creativity and dynamism, as well as its values. Any AI policies undertaken by Congress must maintain the preeminent position of the motion picture industry and other creative sectors.

In identifying policy options, Congress should eschew attempts to undermine the traditional copyright fair use analysis with categorical pronouncements, bright-line rules, or special rules in copyright law regarding AI training, and should reject any calls to implement blanket AI-specific exceptions to our strong American copyright laws. Copyright law’s fair use doctrine—which

³See Stephanie Samsel, *There is a 'wake-up call' for US to be the leader in AI, says White House AI and crypto 'czar'*, FOX NEWS (Jan. 28, 2025, 7:54 PM), <https://www.foxnews.com/media/wake-up-call-us-leader-ai-says-white-house-ai-crypto-czar>.

⁴The White House, *Remarks by Director Kratsios at the Endless Frontiers Retreat* (Apr. 14, 2025), <https://www.whitehouse.gov/articles/2025/04/8716/>.

⁵The White House, *Artificial Intelligence for the American People*, <https://trumpwhitehouse.archives.gov/ai/ai-american-values/>.

⁶See U.S. CONST., amend. I; art. I, § 8, cl. 8 (“The Congress shall have the power [t]o... promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.”).

applies four statutory factors to the specific facts of the particular use—strikes the right balance in both incentivizing the creation of original expression and protecting the rights of copyright owners, while also permitting third parties to exercise those rights where the particular facts show that doing so is “necessary to fulfill copyright’s very purpose, ‘[t]o promote the Progress of Science and useful Arts’”⁷

The fair use factors, as interpreted by many decades of case law, reject infringing uses of copyrighted works that are undertaken for commercial gain and cause harm to the market for a copyrighted work. As such, sweeping generalizations that training is always, or is never, lawful under the fair use doctrine are neither helpful nor correct. Instead, the specific facts regarding the purpose of the training are a major factor in determining whether it is excused under the fair use defense. As the Supreme Court has instructed, “The task [of determining whether a use is fair] is not to be simplified with bright-line rules, for the statute, like the doctrine it recognizes, calls for case-by-case analysis.”⁸

As a global leader in shaping international norms regarding copyright law and policy, the United States should adopt an approach that ensures foreign countries’ copyright regimes and AI policies do not permit the unfair exploitation of American creativity in a manner that violates international standards.

The U.S. Government should encourage AI companies to acquire copyrighted works they wish to use for AI training through lawful means, including voluntary licensing. The use of pirated content for AI training rather than content acquired from lawful and legitimate sources further contributes to the pervasive harm caused by piracy. It is important that Congress encourages the use of lawfully acquired works rather than legitimize piracy as an appropriate source.

Combating Foreign Digital Piracy

Another threat to America’s creative economy and consumers is online piracy of copyrighted content. Sophisticated foreign criminal digital piracy operations operate at enormous scale, anonymously, and outside the boundaries of U.S. law.⁹ Ongoing piracy of filmed entertainment costs the U.S. economy \$29.2 billion and over 230,000 jobs annually.¹⁰ Piracy services can also directly threaten consumers’ personal and financial security.

⁷ *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 575 (1994) (quoting U.S. CONST., art. 1, § 8, cl. 8).

⁸ *Id.* at 577.

⁹ MPA works in close partnership with the Alliance for Creativity and Entertainment (“ACE”), the world’s leading coalition dedicated to protecting the dynamic legal market and reducing digital piracy. Driven by a comprehensive approach to addressing piracy through criminal referrals, civil litigation, and cease-and-desist operations, ACE has achieved many successful global enforcement actions against illegal streaming services and other sources of unauthorized content and their operators. Drawing upon the collective expertise and resources of more than 50 media and entertainment companies around the world and reinforced by the content protection operations of the MPA, ACE protects the creativity and innovation that drive the global growth of core copyright and entertainment industries.

¹⁰ DAVID BLACKBURN ET AL., IMPACTS OF DIGITAL VIDEO PIRACY ON THE U.S. ECONOMY, at ii (2019), https://www.uschamber.com/assets/documents/Digital_Video_Piracy_June_2019.pdf.

In 2023, Digital Citizens Alliance (“DCA”) investigated the impact of visiting and signing up for illegal piracy streaming services. DCA’s report revealed that consumers who visit piracy sites are more susceptible to identity theft, malware, viruses, malicious ads and pop-ups, and ransomware. Specifically, consumers who visited piracy sites are more than four times more likely to report being a victim of identity theft (44% of those visiting piracy sites compared to 10% for those who did not) and five times more likely to report having an issue with malware over the last year (46% of those visiting piracy sites compared to 9% for those who did not).¹¹

The impact of this criminal IP piracy could be dramatically reduced by the enactment of a federal judicial site-blocking law, an effective remedy that already exists in almost sixty countries, and for which MPA testified in support before this committee in May 2025. As we stated then, it is time for Congress to consider providing express authority for a no-fault injunctive-relief regime that will give rights holders what more than a decade of experience around the globe has shown is an effective tool to address piracy.¹²

MPA is excited about the opportunities enabled by recent innovations and artificial intelligence, while vigilant of the harm that will result should AI companies fail to respect the intellectual property rights of our members and other creators. We thank and commend the Judiciary Committee for highlighting these important issues and look forward to working with the committee to ensure that the U.S. motion picture, television, and streaming industry remains the world’s leader in that field.

Sincerely,



Hap Rigby
Senior Vice President, Federal Government Affairs

¹¹ Digit. Citizens All, Giving Piracy Operators Credit: How Signing Up for Piracy Subscription Services Ratchets Up the User Risk of Credit Card Theft and Other Harms, at 1 (2023), <https://www.digitalcitizensalliance.org/clientuploads/directory/Reports/Giving-Piracy-Operators-Credit.pdf>.

¹² *Foreign Threats to American Innovation and Economic Leadership: Hearing Before the S. Comm. on the Judiciary Subcomm. on Intell. Prop.*, 119th Cong. (2025) (testimony of [Karyn A. Temple, Senior Exec. Vice President and Glob. Gen. Couns., Motion Picture Association Association](https://www.judiciary.senate.gov/imo/media/doc/2025-05-14_testimony_temple.pdf)), https://www.judiciary.senate.gov/imo/media/doc/2025-05-14_testimony_temple.pdf.



Senate Judiciary Subcommittee on Crime and Counterterrorism

“Too Big to Prosecute?: Examining the AI Industry’s Mass Ingestion of Copyrighted Works for AI Training”

Statement of the News/Media Alliance

July 16, 2025

The News/Media Alliance (“N/MA”) respectfully submits this statement on this very timely issue which affects the viability of content published by thousands of trusted American publishers.

The News/Media Alliance is a nonprofit organization representing the newspaper, magazine, and digital media industries, and empowering members to succeed in today’s fast-moving media environment. N/MA represents over 2,200 diverse publishers in the United States and internationally, ranging from the largest news and magazine publishers to hyperlocal newspapers and digital-only outlets. Covering all subject matter and political viewpoints, the Alliance’s membership accounts for nearly 90 percent of the daily newspaper circulation in the United States, over 500 individual magazine brands, and dozens of digital-only properties.

The News/Media Alliance commends the Subcommittee for holding this hearing examining the AI industry’s “mass ingestion of copyrighted works for AI training.” This important issue affects rightsholders of all sizes across the country. Many AI developers engage in the predatory and often illegal practice of scraping publishers’ copyright-protected content from their online properties without authorization or compensation. This scraping affects website creators of all types, including media publishers, and threatens to up-end the rules of the road that have enabled the open web and Internet age to flourish.

The increasing use and distribution of generative AI systems and applications raises substantial, unique concerns for media publishers. Getting an AI model to generate creative content requires training it with an enormous volume of materials, and publisher content accounts for a significant and disproportionate volume of the known sources for LLM training, assumedly due

to the high quality of its human expression that forms the foundation for generative AI.¹ This publisher content provides enormous value to the AI developers and publishers deserve to be compensated for use of their materials. Generative AI requires a great amount of another input, electricity, and no one is suggesting that they should be able to simply tap electric transmission lines and take the power for free. So, too, should they be required to pay for their other major input, data containing creative content, and not simply steal it.

To ensure that the generative AI outputs contain current and accurate information, many models are customized by post-training on media content, and “augmented” by retrieving protected content in real time (a process known as retrieval-augmented generation, or RAG), plagiarizing publisher works to power answer engines that actively substitute and harm original content creators. The vast majority of AI-powered chat and search users stop their inquiry at the AI output, without ever clicking through to the publishers who created that content in the first place. For instance, the content licensing platform TollBit found that “AI search engines send 96% less referral traffic to news sites and blogs than traditional Google search.”² When that happens, publishers have no opportunity to generate advertising or subscription revenue to recoup the investment they made in creating the content—which, if left unchecked, could have devastating consequences for the news industry.

While the interests of publishers and generative AI developers could align in a fair exchange of licensing revenues for access to these materials – and N/MA applauds the growing market for AI licensing with media publishers – such partnerships have not yet broadly materialized. Instead of entering comprehensive licensing arrangements with a wide variety of publishers, AI developers make copies of a substantial portion of the Internet, including publisher material behind paywalls, in clear violation of many publishers’ terms of service and despite any potential technical protection measures employed by the publisher. These practices severely harm publishers and other rightsholders, while AI developers rake in billions.³

Publishers expend substantial resources and employ a variety of legal and technical means to protect their valuable content, ranging from copyright registration, copyright notices, and terms

¹ See News/Media Alliance, White Paper: How the pervasive copying of expressive works to train and fuel generative artificial intelligence systems is copyright infringement and not a fair use (“AI White Paper”), <https://www.newsmediaalliance.org/wp-content/uploads/2025/02/AI-White-Paper-with-Technical-Analysis.pdf>, at 19-20.

² Rashi Shrivastava and Richard Nieva, *New Data Shows Just How Badly OpenAI and Perplexity Are Screwing Over Publishers*, FORBES (Mar. 3, 2025), <https://www.forbes.com/sites/rashishrivastava/2025/03/03/openai-perplexity-ai-search-traffic-report/>.

³ OpenAI and Anthropic are projected to generate a combined \$85 billion in revenue by 2027, a substantial increase from the near total lack of revenue in 2022. See Stephanie Palazzolo and Amir Efrati, *It’s Time to Take Anthropic and OpenAI’s Wild Revenue Projections Seriously*, THE INFORMATION (Jul. 3, 2025), <https://www.theinformation.com/articles/time-take-anthropic-openais-wild-revenue-projections-seriously>.

of service to technical controls, such as paywalls, CAPTCHA, metering systems, and protocols like robots.txt. Third-party users and businesses are required to comply with and respect these measures as a condition of access. Globally, AI scraping activity has increased dramatically in the past two years, with publisher monetization tool TollBit reporting that AI scraping increased two-fold from third to fourth quarter of 2024 and content delivery network provider Cloudflare estimating that AI bots have scraped almost 40 percent of all websites and 80 percent of the top 10 most popular Internet properties.⁴ While AI scraping activity has proliferated exponentially, publishers have not witnessed a corresponding increase in referral traffic. In fact, while every two scrapes by Google used to lead to one visitor, today that ratio is 18:1.⁵ Meanwhile, OpenAI’s crawl to visitor ratio has increased from 250:1 to 1,500:1 while Anthropic has gone up from 6,000:1 to 60,000:1.⁶

Until AI companies change their practices or publishers can effectively enforce protections against them, there is no feasible way for publishers to opt out of massive scraping while remaining accessible to human readers. While technical measures such as paywalls and robots.txt exist – and all publishers have machine-readable terms of service – they are all too often not respected. There is substantial evidence that AI scrapers bypass publisher paywalls to access protected content when publishers have elected to adopt them.⁷ Meanwhile, the robots.txt exclusion protocol is a blunt instrument that offers no granular control over licensing conditions and the content made available to different types of scrapers. Some developers, including Meta and Google, also categorize bots as “fetchers,” arguing that because they operate at the request of a user, they are not required to respect robots.txt.⁸ Combined with AI scrapers’ failure to accurately and reliably identify themselves – utilizing diversion tactics such

⁴ Alex Bocharov et al., *Declare Your AI Independence: Block AI Bots, Scrapers and Crawlers with a Single Click*, THE CLOUDFLARE BLOG (Jul. 3, 2024), <https://blog.cloudflare.com/declaring-your-ai-independence-block-ai-bots-scrapers-and-crawlers-with-a-single-click/>; TOLLBIT, TOLLBIT ARTIFICIAL INTELLIGENCE USER AGENT INDEX - Q1 2025 at 10 (Feb. 24, 2025), <https://tollbit.com/bots/24q4/>.

⁵ Rob Thubron, *Cloudflare Tests “Pay-for-Crawl” System to Charge AI Firms for Scraping Website Content*, TECHSPOT (Jul. 1, 2025), <https://www.techspot.com/news/108521-cloudflare-tests-pay-crawl-system-charges-ai-firms.html>; Ethan Hays (@ethanhays), X (Jun. 27, 2025, 6:32 PM), <https://x.com/ethanhays/status/1938651733976310151>.

⁶ Rob Thubron, *Cloudflare Tests “Pay-for-Crawl” System to Charge AI Firms for Scraping Website Content*, TECHSPOT (Jul. 1, 2025), <https://www.techspot.com/news/108521-cloudflare-tests-pay-crawl-system-charges-ai-firms.html>; Ethan Hays (@ethanhays), X (Jun. 27, 2025, 6:32 PM), <https://x.com/ethanhays/status/1938651733976310151>.

⁷ See, e.g., Sara Guaglione, *As AI Lawsuits Mount, Publishers Still Struggle to Block the Bots*, DIGIDAY (Apr. 30, 2025), <https://digiday.com/media/as-ai-lawsuits-mount-publishers-still-struggle-to-block-the-bots/>. Paywalls also do not fit the business model of all publishers and can substantially reduce the public’s access to free or low-cost content, including local journalism.

⁸ Meta Web Crawlers, Meta, <https://developers.facebook.com/docs/sharing/webmasters/web-crawlers/> (last accessed Feb. 18, 2025) (“The Meta-ExternalFetcher crawler may bypass robots.txt because it performs fetches that were requested by the user.”); List of Google User-Triggered Fetchers, Google Search Central, <https://developers.google.com/search/docs/crawling-indexing/google-user-triggered-fetchers> (last accessed Feb. 18, 2025) (“Because the fetch was requested by a user, these fetchers generally ignore robots.txt rules.”).

as impersonating reputable AI bots⁹ and the human readers publishers are trying to reach, using headless browsers to evade identification,¹⁰ using proxies to get around geolocation and IP blocks, evading CAPTCHA security measures, and switching between IP addresses – together with third-party companies providing services to evade publishers’ technical protection measures,¹¹ publishers are left with few options to defend themselves and bad actors are able to continue scraping publisher content without authorization.

The evasion of these access control measures and the subsequent unauthorized use of publisher content is illegal under contract and intellectual property law, including the anti-circumvention provisions contained in Section 1201 of the Copyright Act. While the Department of Justice could and should address the systemic unauthorized scraping of protected content by increasing the enforcement of the Computer Fraud and Abuse Act, meaningful and efficient private enforcement of publishers’ rights is both complex and expensive, especially considering the vast array of AI developers engaging in these activities, rendering it next to impossible even for the largest publishers. As a consequence, this scraping leads to enormous costs to publishers and other website operators, who invest substantial resources to produce high-quality content for their readers.¹² At worst, these costs can lead to job losses and the closure of local news sources, facilitating the spread of news deserts across American local communities and depriving many citizens, especially in rural and smaller communities, of information about their community and local events. Further, the current situation puts the onus on publishers to prevent unauthorized copying of their protected content, while leaving companies engaging in these infringing activities largely unaffected. This is antithetical to the guiding principles of U.S. copyright law and the exclusive property rights afforded to rightsholders.

These harms are exacerbated by some actors storing and reselling content they have collected from rightsholder properties to unrelated third parties. This includes the sale of content

⁹ Iliia Bromberg and Christine Ferrusi Ross, *Managing AI Bots as Part of Your Overall Bot Management Strategy*, AKAMAI BLOG (Nov. 20, 2024), <https://www.akamai.com/blog/security/managing-ai-bots-part-overall-bot-management-strategy>.

¹⁰ See, Behind the Scenes: How Headless Browsers are the Bots' Secret Weapon, TrafficGuard (Aug. 23, 2024), <https://www.trafficguard.ai/blog/behind-the-scenes-how-headless-browsers-are-the-bots-secret-weapon>.

¹¹ See, e.g., *AI Web Unblocker for Scraping Any Website*, Zenrows, <https://www.zenrows.com/products/universal-scrapers/web-unblocker> (“ZenRows – the best way to bypass any anti-bot and get all the data you're looking for. ... Cloudflare, Akamai, Datadome... they represent a major obstacle for web scrapers. As these systems evolve daily, manual bypassing becomes a time-consuming chase. ZenRows provides you with a fingerprinting bypass, premium residential proxies, and every tool you need to scrape any website without getting blocked.”); *7 Anti-Scraping Techniques You Need to Know*, Data Journal Blog (Jan. 27, 2025), <https://medium.com/@datajournal/anti-scraping-techniques-2c8a92f700a6> (last accessed Feb. 18, 2025); Bright Data, <https://brightdata.com/>.

¹² According to a recent study, AI scrapers are also “hammering the servers of libraries, archives, museums, and galleries, and are in some cases knocking their collections offline,” posing a substantial threat to these civic institutions. <https://www.glamelab.org/products/are-ai-bots-knocking-cultural-heritage-offline/?ref=404media.co>

legitimately scraped for largely authorized use purposes – such as for search indexing, in adherence to robots.txt limitations – for unknown and unanticipated uses, including AI training or retrieval augmented generation (RAG). While publishers by and large allow the crawling of their websites for search indexing purposes to ensure that they show up in search engine results and to bring in more traffic, they have not consented to the downstream wholesaling of this indexed content, and especially not for purposes that cannibalize their traffic and for which licensing opportunities exist. Publishers have no control over these downstream uses, and it can have significant effects on publishers, undermining their longstanding and sustainable economic models, including licensing, advertising, and subscription revenues.

AI developers are engaged in practices that systemically misappropriate publisher and other rightsholder content, whether through scraping or downstream distribution, to bolster their bottom line, while posing an existential threat to the creative industries. This fundamental imbalance is unsustainable and potentially devastating to the American creative industries that lead the world and sustain American cultural influence around the globe. In this case, moving fast may result in the breaking of America's creative industries. We appreciate the Subcommittee's attention to these issues, and moving forward strongly urge the Committee to continue to consider these issues in a balanced, thoughtful and critical way in order to ensure policy outcomes that will support and strengthen the continued viability of high-quality publishing in the United States.

We appreciate the opportunity to present these views to the Subcommittee, and stand ready to answer any questions Members may have on these important issues affecting millions of American publishers and creative industry professionals.



ADDRESS 444 Gulf of Mexico Dr
Longboat Key, FL 34228

WEBSITE corp.rumble.com

The Honorable Josh Hawley
Chairman, Subcommittee on Crime and Counterterrorism
U.S. Senate Committee on the Judiciary
381 Russell Senate Office Building
Washington, D.C. 20510

The Honorable Richard J. Durbin
Ranking Member, Subcommittee on Crime and Counterterrorism
U.S. Senate Committee on the Judiciary
711 Hart Senate Building
Washington, D.C. 20510

July 16, 2026

Dear Senators:

As founder and Chief Executive Officer of Rumble, a fast-growing video sharing platform and cloud services provider, I write to thank you for your decision to hold a hearing on the dangers of artificial intelligence models being trained using copyrighted material and intellectual property that has been illegally scraped from the owners or licensees of such property. Rumble is a platform where content creators share their opinions, show their creativity, and conduct commerce. A ravenous artificial intelligence industry, given free rein, would bulldoze through every well-established legal protection we have for the products of creators' minds.

It would destroy all of it.

I am certain the Members of this Judiciary Subcommittee are familiar with the history of these protections as they are specifically upheld by the U.S. Constitution. Article I, Section 8, Clause 8 gives Congress the authority to secure for "Authors and Inventors the exclusive Right to their respective Writings and Discoveries." Furthermore, intellectual property, copyrights, and patents, in one way or another, have been protected by Congress in federal law since 1790.

American creative works from literary classics, to black and white motion pictures, to the greatest radio hits helped shape American culture and share that culture, along with American values, around the world. It promoted an idyllic way of life, the stories of diversity, and even educated hard truths and experiences of personal struggle. Just because the emerging technologies of today

are not necessarily sold for profit in a commercial mall or online store, doesn't mean that the digital creators of today do not deserve ownership of their creations as those who came before them.

For one thing, it is only fair to acknowledge that creators today have the same right to own what they have imagined and created themselves. We should also acknowledge that to eliminate that protection would be to destroy the incentive for creative people to produce anything. Finally, we need to use the current laws in place to protect these creations as they are today.

Creative content is not just for personal education, enjoyment, and laughter. For many creators, this is their livelihood. I am proud of the creator revenue program we have built on Rumble and believe it to be one of the best models in the industry. Rumble offers a percentage breakdown that is already more favorable than other "Big Tech" platforms, and I would be amiss not to note that app stores already take a percentage of creator's revenue through unfair app stores fees.

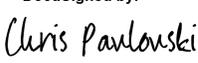
It should come as no surprise then that some of these same companies are looking to profit off Rumble's creators' content. While copyright laws have been in place for decades, we are learning of case after case of emerging AI models scraping (i.e., stealing!) content in the dead of night to train their models on language and the nuances of human communication.

Ironically, these companies have amassed billions of dollars in revenue from their own intellectual property. IP protections for "me" ("Big Tech") – but not for "thee" (digital creators).

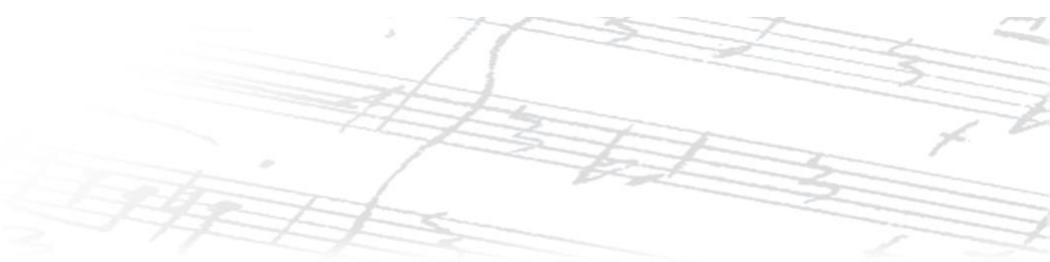
The United States has proven itself to be the strongest, most creative, and most innovative country in the world. From the development of GPS technology and F-35s, to the software and technology of Microsoft Windows and Apple's iOS – IP protections have helped fuel the economic and security dominance that has sustained U.S. hegemony for the last 250 years.

There's no question that the United States can and will be the leader on this emerging technology – but there must be some tactics that are not allowed in its development and training – and legalized thievery should be one of them.

Sincerely,

DocuSigned by:

9FE2662F79A64B7...

Chris Pavlovski
Founder, Chairman, and Chief Executive Officer
Rumble



PRESIDENT
Ashley Irwin

8306 Wilshire Blvd. • Suite 512 • Beverly Hills, California 90211 • (310) 281-2812 • www.TheSCL.com

VICE PRESIDENTS
David Das
Charles Bernstein

July 21, 2025

TREASURER
Christopher Farrell

Senate Judiciary Committee, Crime Subcommittee
U.S. Senate
Washington, D.C. 20510

RECORDING SECRETARY
Danny Gray

Dear Senators Hawley, Durbin and Welch,

CHIEF OPERATING OFFICER
Marie Kingsley

On behalf of the Society of Composers & Lyricists (SCL) and our membership of music creators working in film, television, streaming, video games, and emerging media, I write to express our deep gratitude for the powerful and principled statements that you all expressed during the recent Senate Judiciary Committee Crime Subcommittee hearing, *“Too Big to Prosecute? Examining the AI Industry’s Mass Ingestion of Copyright Works for AI Training.”*

THE SCORE
Rajasri Mallikarjuna, Editor

Your willingness to speak out in defense of American creators is not only commendable—it is patriotic. At its core, this issue isn’t just about copyright or commerce. It’s about the very values that define the United States of America: individual liberty, the sanctity of private property, and the right to reap the rewards of one’s own labor. These rights are enshrined in our Constitution and must be upheld in the digital age just as they were in the founding days of our republic.

ADVISORY BOARD
Charles Bernstein
Bill Conti
Charles Fox
James Newton Howard
Alan Menken
Thomas Newman
Marc Shaiman
Howard Shore
Alan Silvestri
Diane Warren
Christopher Young
Hans Zimmer

The SCL fully agrees that this is “a moral issue as much as anything else”—a matter of whether America will continue to be a nation that respects and protects the rights and creations of individuals. The warning against allowing “mega-corporations” to profit off the uncompensated work of others with impunity resonated deeply with our members.

IN MEMORIAM
Alan Bergman
Marilyn Bergman
Elmer Bernstein
John Cacavas
Jerry Goldsmith
Ron Grant
Maurice Jarre
Quincy Jones
Peter Matz
David Raksin
Lalo Schiffrin
Patrick Williams

At a time when rapidly advancing technology threatens to outpace the laws that govern it, the Subcommittee’s leadership reminds us that America must never trade its founding principles for convenience or profit. The strength of our nation lies in its ability to protect the rights of the individual—even, and especially, in the face of enormous corporate power.

BOARD OF DIRECTORS
Michael Abels
Esin Aydingoz
Ramon Balcazar
Lori Barth
Michael Lehmann Boddicker
Russell Brower
Cliff Eidelman
Jack D. Elliot
Lynn F. Kowal
Mark McKenzie
Hélène Muddiman
Cindy O’Connor
Hannah Parrott
Greg Pliska
Carlos Rafael Rivera
Lolita Ritmanis
Elizabeth Rose
Garry Schyman
Tony Scott-Green
Jay Weigel
Chris Wirsig

The SCL is firmly committed to ensuring that creators are not left behind in the AI era. We advocate for the foundational “Three Cs” that every generative AI system must observe:

Consent from creators before their works are ingested by generative systems,
Credit for all audiovisual creators whose work is used, and
Compensation at fair market rates for both ingestion and derivative outputs.

PAST PRESIDENTS
John Addison
Richard Bellis
Bruce Broughton
Jay Chattaway
Ray Colcord
James Di Pasquale
Dan Foliart
Arthur Hamilton

As you know, the core copyright industries—including publishing, film, theater, television, music, software, gaming, newspapers, and magazines—not only play a vital role in the U.S. economy but also carry a profound responsibility in shaping our culture and society. Yet the freelance creators who fuel these industries

remain dangerously under-protected, as generative AI systems are increasingly trained to mimic their work without permission or compensation.

We were especially encouraged by the acknowledgments during the hearing that Generative AI systems are being trained on copyrighted works without authorization, and many of these works were sourced from illegal online repositories.

This is not innovation—it is appropriation. And it undermines the very engine of American ingenuity.

The SCL stands ready to work alongside all of you to ensure that our laws reflect our values. Let us continue building a future where creativity is honored, the Constitution is upheld, and the American Dream remains possible for those who dare to create.

With sincere appreciation,

A handwritten signature in black ink, appearing to read "Ashley Irwin". The signature is fluid and cursive, with a long, sweeping underline that extends to the right.

Ashley Irwin
President
The Society of Composers & Lyricists

cc. Members of the Senate Judiciary Committee Crime Subcommittee

July 16, 2025

Chairman Josh Hawley
Ranking Member Richard J. Durbin
United States Senate Subcommittee on Crime and Counterterrorism
HVC 210 Capitol, Washington, D.C.

Re: CAIDP Statement for the record: *Too Big to Prosecute?: Examining the AI Industry’s Mass Ingestion of Copyrighted Works for AI Training, July 16, 2025*

Dear Chairman Hawley, Ranking Member Durbin, and Members of the Committee,

The Center for AI and Digital Policy (CAIDP) commends this Committee’s sustained focus and efforts to address AI risks and harms. We welcome this hearing on *Too Big to Prosecute?: Examining the AI Industry’s Mass Ingestion of Copyrighted Works for AI Training*.¹ The AI industry is rapidly acquiring the commercial value of the work of artists, authors, and others. The current trajectory of the AI industry threatens to upend copyright protections by mimicking and then commercializing human identity and human creativity. We offer two brief recommendations to the Committee:

1. Enact protections for creators without compromising human authorship requirements for copyrightability, as recommended by the US Copyright Office. The **No Fakes Act**² sets a good baseline for protections.
2. Urge the Federal Trade Commission to complete its investigation into OpenAI, initiated by CAIDP, and establish safeguards for the creative economy

About CAIDP

The Center for AI and Digital Policy (“CAIDP”) is an independent research and education organization, based in Washington, D.C.³ CAIDP’s mission is to ensure that artificial intelligence (AI) serves the public interest based on fundamental rights, democratic institutions, and the rule of law. CAIDP routinely provides nonpartisan advice to Congressional committees on matters

¹ United States Senate, Committee on Crime and Counterterrorism, *Too Big to Prosecute?: Examining the AI Industry’s Mass Ingestion of Copyrighted Works for AI Training* (Jul. 16, 2025), <https://www.judiciary.senate.gov/committee-activity/hearings/too-big-to-prosecute-examining-the-ai-industrys-mass-ingestion-of-copyrighted-works-for-ai-training>

² S.4875 - 118th Congress (2023-2024): NO FAKES Act of 2024, S.4875, 118th Cong. (2024), <https://www.congress.gov/bill/118th-congress/senate-bill/4875>.

³ CAIDP, <https://www.caidp.org/>.

involving AI policy.⁴ CAIDP publishes annually the *Artificial Intelligence and Democratic Values (CAIDP Index)*, a comprehensive review of AI policies and practices worldwide,⁵ and *The AI Policy Sourcebook* (CAIDP 2025), a compendium of AI governance frameworks and policy resources. We have previously the US Copyright Office on the impact of AI.⁶

The AI Industry Does not Play Fair, While Claiming Fair Use

Generative AI programs—such as Open AI's DALL-E and ChatGPT, Stability AI's Stable Diffusion, and Midjourney—are trained on writings, photos, paintings, and other works, created by authors, photographers, painters, and other creative artists.⁷ More than 40 lawsuits have been filed against AI companies since 2022, seeking to hold these companies accountable for appropriating millions of copyrighted works to develop training models without compensation.⁸ Recently in two separate author lawsuits against Anthropic and Meta, the courts decided that in favor of the companies holding that training generative AI models with copyrighted material was fair use, though the reasoning in the cases differed.⁹

1. Generative AI Models Devalue Human Creativity

Creators depend on the uniqueness of their work to support their careers. Allowing AI companies to ingest, train on, and produce content mimicking specific artists' styles devalues creators' rights and impacts human creativity and passion. Under the Copyright Act, the fair use exemption was intended for limited use of copyrighted material for purposes with transformative impacts, something that improves “the progress of science and the arts, without diminishing the incentive to create.”¹⁰ As one commenter notes “When AI slop devalues creative effort, will people’s motivations for putting work out into the world start to fall away?”¹¹

⁴ CAIDP, *Statements*, <https://www.caidp.org/statements/>.

⁵ CAIDP, *Artificial Intelligence and Democratic Values* (2025). <https://www.caidp.org/reports/aidv-2025/>

⁶ CAIDP, *Comments to U.S. Copyright Office, Library of Congress on Artificial Intelligence and Copyright* (Oct. 30, 2023), https://downloads.regulations.gov/COLC-2023-0006-10313/attachment_1.pdf

⁷ Congressional Research Service, *Generative Artificial Intelligence and Copyright Law*, LSB 109222 (Jun.16, 2025), <https://www.congress.gov/crs-product/LSB10922>

⁸ Alex Reisner, *Judges Don't Know What AI's Book Piracy Means*, The Atlantic (Jul. 14, 2025), <https://www.theatlantic.com/technology/archive/2025/07/anthropic-meta-ai-rulings/683526/>

⁹ Will Douglas Heaven, *What comes next for AI copyright lawsuits?*, MIT Technology Review (Jul. 1, 2025), <https://www.technologyreview.com/2025/07/01/1119486/ai-copyright-meta-anthropic/>

¹⁰ Congressional Research Service, *Generative Artificial Intelligence and Copyright Law*, LSB 109222 (Jun.16, 2025), <https://www.congress.gov/crs-product/LSB10922>

¹¹ *Id.*

2. Generative AI models misappropriate image and voice, and violate biometric privacy

While copyright law protects fixed expressions, it does not shield the unique style, voice, or likeness of an artist.¹² Current practices of training on people’s voice and image has implications beyond copyright and threatens the privacy of the billions of people who share information online.¹³ It is now well-established that the training data is procured through invasive data-scraping practices that incorporate both private data and data publicly available.¹⁴ Yet the public has little visibility into how their data is used.¹⁵ AI companies keep training sets and processes hidden from users, researchers, and regulators.¹⁶ Not too long ago OpenAI had a snafu with the launch of GPT-4o voice which was eerily like Scarlett Johansson’s.¹⁷

3. Current business practices are opaque, deceptive, and unfair to consumers and competition

In comments to the U.S. Copyright Office, CAIDP stated, “Allowing unrestricted imitation by AI could dilute the distinctiveness of such intellectual assets, potentially leading to market confusion and unfair competition.”¹⁸ **Just recently news went viral of Spotify’s AI band amassing more than 1 million streams.**¹⁹ “The rise of AI-generated bands and music entering the market points to the fact that tech companies have been training AI models using creative works – largely without authorization or payment to creators and rights-holders – in order to directly compete with human artistry.”²⁰

¹² United States Copyright Office, Copyright and Artificial Intelligence, Part 3: Generative AI Training, (May, 2025), <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>

¹³ CAIDP, *Comments to U.S. Copyright Office, Library of Congress on Artificial Intelligence and Copyright* (Oct. 30, 2023), https://downloads.regulations.gov/COLC-2023-0006-10313/attachment_1.pdf

¹⁴ CAIDP, *Complaint to FTC – in re OpenAI/ChatGPT*, 2nd Supplementary Complaint (Nov. 14, 2023), <https://www.caidp.org/cases/openai/>

¹⁵ CAIDP, *Comments to U.S. Copyright Office, Library of Congress on Artificial Intelligence and Copyright* (Oct. 30, 2023), https://downloads.regulations.gov/COLC-2023-0006-10313/attachment_1.pdf

¹⁶ CAIDP, *Complaint to the FTC – in re OpenAI and ChatGPT* (Mar. 30, 2023), <https://www.caidp.org/cases/openai/>; Kate Crawford, Twitter (Mar. 22, 2023), <https://x.com/katecrawford/status/1638524013432516610>

¹⁷ Forbes, *The Prompt: Scarlett Johansson Vs OpenAI* (May 21, 2024), <https://www.forbes.com/sites/rashishrivastava/2024/05/21/the-prompt-scarlett-johansson-vs-openai/>

¹⁸ CAIDP, *Comments to U.S. Copyright Office, Library of Congress on Artificial Intelligence and Copyright* (Oct. 30, 2023), https://downloads.regulations.gov/COLC-2023-0006-10313/attachment_1.pdf; U.S. Copyright Office, *Report on Copyright and Artificial Intelligence – Part 1: Digital Replicas* (Jul. 2024), pg. 53, <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf>

¹⁹ The Guardian, *An AI-generated band got 1m plays on Spotify. Now music insiders say listeners should be warned* (Jul. 14, 2025), <https://www.theguardian.com/technology/2025/jul/14/an-ai-generated-band-got-1m-plays-on-spotify-now-music-insiders-say-listeners-should-be-warned>

²⁰ Id

According to the FTC, the use of a digital replica that mimics an individual’s voice and likeness might qualify as an unfair method of competition or an unfair or deceptive practice, particularly if it “deceives consumers, exploits a creator’s reputation or diminishes the value of her existing or future works, reveals private information, or otherwise causes substantial injury to consumers.”²¹ The opacity in business practices is even more concerning given that AI products are designed to mimic humans. Consumers do not know whether they are interacting with an AI or a human, paying for art created by an AI or a human. The lack of transparency pervades from the training stage through the consumption stage. By arguing their models count as “fair use,” AI developers are capitalizing on the human creativity of others while hurting creators’ livelihoods and reshaping the creative economy to its benefit. Furthermore, AI companies have actively and aggressively been engaging the White House to carve out fair use exceptions in the U.S. AI Action Plan scheduled to be announced next week.²²

Recommendation 1: Enact protections for creators without compromising human authorship requirements for copyrightability

The U.S. Copyright Office advised Congress to move forward with the creation of a new federal right. In its report on Digital Replicas, the Copyright Office concluded “We recommend that Congress establish a federal right that protects all individuals during their lifetimes from the knowing distribution of unauthorized digital replicas. The right should be licensable, subject to guardrails, but not assignable, with effective remedies including monetary damages and injunctive relief.”²³ CAIDP supports the recommendation of the U.S. Copyright Office. For this reason, we urge the Committee to move forward legislation like the **No Fakes Act**,²⁴ which would:

- Grants individuals the right to authorize the use of the image, voice, or visual likeness of the individual in a digital replica
- Creates a notice-and-takedown process for unauthorized deepfakes
- Hold companies liable for unauthorized commercial use of voice, image, or likeness

²¹ Comment of the United States Federal Trade Commission to the United States Copyright Office, *Artificial Intelligence and Copyright*, Docket No. 2023-6,

https://www.ftc.gov/system/files/ftc_gov/pdf/p241200_ftc_comment_to_copyright_office.pdf

²² Axios, *AI firms push to use copyrighted content freely* (Mar. 20, 2025),

<https://www.axios.com/2025/03/20/white-house-ai-policy-copyright>

²³ U.S. Copyright Office, *Report on Copyright and Artificial Intelligence – Part 1: Digital Replicas* (Jul. 2024), pg. 57, <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf>

²⁴ Amy Klobuchar, *Klobuchar, Coons, Blackburn and Colleagues Reintroduce Bipartisan NO FAKES Act* (Apr. 22, 2025), <https://www.klobuchar.senate.gov/public/index.cfm/2025/4/klobuchar-coons-blackburn-and-colleagues-reintroduce-bipartisan-no-fakes-act>

- Respect First Amendment protections while creating national-level standards without preempting state laws

Recommendation 2: Urge the FTC to complete its investigation into OpenAI and establish guardrails against deceptive and unfair business practices

The FTC is uniquely placed to exercise oversight on business practices of AI companies and investigate what data is used to train AI models assessing risks to privacy, consumers, and competition. Especially where biometric data like voice, iris, is involved the FTC has an obligation to enforce prohibitions on unfair and deceptive practices. More than two years ago, CAIDP filed a complaint with the Federal Trade Commission regarding ChatGPT, arguing that OpenAI's business practices violated Section 5 of the FTC Act and the Commission's guidance on AI products.²⁵ CAIDP highlighted OpenAI's concession that "GPT-4 has the potential to be used to identify private individuals when augmented with outside data."²⁶ More alarmingly, AI companies' training datasets are rife with children's data including voice, image, and location.²⁷

In July 2023, both the *New York Times* and the *Wall Street Journal* reported that the FTC issued a civil investigative demand to OpenAI to investigate concerns of consumer deception, data privacy, and public safety, as CAIDP had urged.²⁸ However, after almost two years now, there is still no legal outcome, no judgment, and no settlement. There are still no guardrails for AI products sold to consumers in the United States.²⁹ Since then, we have seen alarming whistleblower reports³⁰ warning that, **"AI companies have strong financial incentives to avoid effective oversight, and we do not believe bespoke structures of corporate governance are sufficient**

²⁵ CAIDP, *In the Matter of OpenAI* Original Complaint (2023), <https://www.caidp.org/cases/openai/>

²⁶ CAIDP, *In the Matter of OpenAI* Original Complaint at pg. 22 (2023), <https://www.caidp.org/cases/openai/>

²⁷ CAIDP, *Statement to Senate Judiciary Committee Hearing on Children's Safety in the Digital Era* (Feb. 12, 2025), https://www.linkedin.com/posts/center-for-ai-and-digital-policy_aigovernance-chldrensafety-activity-7297995156975202305-my2f

²⁸ *New York Times*, *F.T.C. Opens Investigation Into ChatGPT Maker Over Technology's Potential Harms*, (Jul. 13, 2023), <https://www.nytimes.com/2023/07/13/technology/chatgpt-investigation-ftc-openai.html>; *The Wall Street Journal*, *ChatGPT Comes Under Investigation by the Federal Trade Commission* (Jul.13, 2023), <https://www.wsj.com/articles/chatgpt-under-investigation-by-ftc-21e4b3ef>; Federal Trade Commission, *Civil Investigation Demand Schedule FTC File No. 232-3044*. <https://www.caidp.org/app/download/8467488463/FTC-CID-OpenAI-CAIDP.pdf>;

²⁹ CAIDP, *ChatGPT and the Federal Trade Commission: Still No Guardrails*, at 1 (Jul. 2024), <https://www.caidp.org/app/download/8520338863/CAIDP-Still-No-Guardrails-July2024.pdf>

³⁰ Kevin Roose, *OpenAI Insiders Warn of a 'Reckless' Race for Dominance*, *The New York Times* (Jun. 5, 2024), <https://www.nytimes.com/2024/06/04/technology/openai-culture-whistleblowers.html>; *The Washington Post*, *OpenAI illegally barred staff from airing safety risks, whistleblowers say*, (Jul. 13, 2024), <https://www.washingtonpost.com/technology/2024/07/13/openai-safety-risks-whistleblower-sec/>; Julia Shapero, *Senators press OpenAI over safety concerns after whistleblower complaint*, *The Hill* (Jul. 23, 2024), <https://thehill.com/policy/technology/4788030-openai-senators-press-ai-safety/>

to change this.³¹ The Senate Judiciary Committee examined these issues in its hearing last year "Oversight of AI: Insiders' Perspective"³² While Congress considers regulation to incentivize innovation and good actors, it is up to the FTC to address abusive practices in the market. We urge you to call upon the FTC to conclude its investigation into OpenAI, make public its order, and establish guardrails for AI services.

While the contested contours of fair use are pending before the Courts, Congress can act decisively by enacting legislation to safeguard the rights of creative artists and urging the FTC to complete its investigation into OpenAI's business practices that place at risk the livelihood of millions of authors, artists, and musicians across the country.

We thank you for your consideration of our views. We ask that this statement from the Center for AI and Digital Policy be included in the hearing record. We would be pleased to provide you and your staff with additional information.

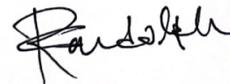
Sincerely yours,



Marc Rotenberg
Executive Director



Merve Hickok
President



Christabel Randolph
Associate Director



Charlotte Lin
Research Assistant



Caroline King
Research Assistant

³¹ A Right to Warn about Advanced Artificial Intelligence, <https://righttowarn.ai>

³² U.S. Senate Judiciary Committee, Subcommittee on Privacy, Technology, and the Law, *Oversight of AI: Insiders' Perspectives* (Sept. 17, 2024), <https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-insiders-perspectives>



Statement of

**Keith Kupferschmid
Chief Executive Officer
Copyright Alliance**

before the

**SENATE COMMITTEE ON THE JUDICIARY
SUBCOMMITTEE ON CRIME AND COUNTERTERRORISM**

July 16, 2025

The Copyright Alliance, on behalf of our membership, submits this statement for the record concerning the hearing titled *Too Big to Prosecute?: Examining the AI Industry's Mass Ingestion of Copyrighted Works for AI Training* before the Senate Judiciary Committee, Subcommittee on Crime and Counterterrorism, on July 16, 2025.

The Copyright Alliance is a non-profit, non-partisan public interest and educational organization that is dedicated to advocating policies that promote and preserve the value of copyright, and to protecting the rights of creators and innovators. We represent the copyright interests of over 2 million individual creators, including established authors and artists, performers and photographers, and software coders and songwriters, as well as a new generation of creators. Some of these creators are career professionals, while others are hobbyists. Some have years of experience, while others are just embarking on their burgeoning careers. Some are critically acclaimed, while others toil in relative obscurity or have limited audiences.

We also represent the copyright interests of over 15,000 organizations in the United States, across the spectrum of copyright disciplines. These include motion picture and television studios, record labels, music publishers, book and journal publishers, newspaper and magazine publishers, video game companies, software and technology companies, visual media companies, sports leagues, radio and television broadcasters, database companies, standard development organizations and many more. While each of these organizations may come to the Copyright Alliance with somewhat different experiences, views, and interests, they all fall under the Copyright Alliance umbrella for one unifying reason—their strong support for the value and importance of copyright and protecting the rights of human creators and copyright owners.

Copyright Alliance members—whether they are an individual creator or an organization, whether they are big or small, or whether they are more traditional creators/copyright owners or a new generation of creators/copyright owners—share at least two things in common: (1) they rely on copyright law to protect their efforts and investments in the creation, reproduction, distribution, and adaptation of works for the public to enjoy, and (2) they are harmed by piracy and are concerned about copyright infringement and piracy-related issues raised by generative AI.

Many Copyright Alliance members and others in the creative industries are already using artificial intelligence as a tool to assist with their content creation, just as they use other technologies, and undoubtedly will use more sophisticated versions of AI as they develop. We therefore support efforts to advance innovation. But all development and use of technology must be done legally and responsibly.

This is consistent with the position the first Trump Administration took that “[t]he AI technologies we develop must also reflect [the] fundamental American values” in “freedom, guarantees of human rights, the rule of law, stability in our institutions, rights to privacy,

respect for intellectual property, and opportunities to all to pursue their dreams.”¹ Abiding by these principles will be critical to making American implementation the gold standard for AI and to protecting and expanding the United States’ dominant global positions in *both* AI and the copyright industries.

The United States will not win the AI race with China if it comes at the expense of good copyright policy. Indeed, not only would undermining copyright law hurt the United States as the global leader in the creative industries, but abiding by good copyright policy will help ensure that users at home and abroad can trust and benefit from American AI technology over implementations from AI developers from other countries. That means using trusted, curated datasets that are the most accurate, providing consumers and businesses in the United States and across the globe the confidence that they can use American AI tools without fear of “hallucinations” or other flaws that can cause harm. Indeed, as we discuss in more detail below, a recent report demonstrating that chatbots already include an alarming amount of Chinese propaganda seeded by the Communist Party demonstrates the harm that comes from relying on indiscriminate scraping of the internet to train AI systems as opposed to licensing reliable content.²

The two copyright issues of most interest to Copyright Alliance members and the copyright community more broadly are online piracy and generative AI. These two issues may appear to be independent from one another, but recent AI copyright infringement lawsuits have shown that they are very much intertwined. Indeed, those lawsuits have demonstrated that some AI companies have leveraged the proliferation of copyrighted works in unauthorized centralized databases to their benefit, for instance, by accessing so-called “shadow libraries,” like Z-Library or Lib-Gen, they copy and ingest pirated works that they use to train their AI systems

¹The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update (emphasis added), <https://trumpwhitehouse.archives.gov/ai/ai-american-values/>.

²See COURTNEY MANNING, MONIQUE SHUM & JOSEY WALDEN, AMERICAN SECURITY PROJECT, EVIDENCE OF CCP CENSORSHIP, PROPAGANDA IN U.S. LLM RESPONSES (June 25, 2025) (stating that “AI-powered chatbots in the United States now regurgitate CCP propaganda in Chinese and English when prompted on certain topics, posing significant ramifications for global AI development and U.S. national security”), https://cdn.prod.website-files.com/67919c3b2972e57c613c2ea2/685b1a27a830fb5b6e7ff511_Sentinel%20Brief%20-%20Evidence%20of%20CCP%20Censorship%20in%20LLM%20Responses.pdf.

instead of licensing them from copyright owners. Therefore, the focus of this hearing, which is to address that overlap, is of significant importance to the Copyright Alliance.

Piracy Harms America's Strong Creative Economy

America's innovation and creative economies continue to be the best in the world. There is a reason for that U.S. dominance—it is U.S. copyright law. Strong and effective copyright protection is critical to and fuels the U.S. economy, culture, trade and employment. A report on the economic impact of copyright by the International Intellectual Property Alliance (IIPA) found that, in 2023, the core copyright industries contributed more than \$2 trillion to the U.S. gross domestic product (GDP) (accounting for 7.66% of the U.S. economy) and employed 11.6 million workers (or 5.43% of the workforce).³ In addition to growing at a rate more than three times that of the rest of the economy, the report notes that the core copyright industries: (1) make up an increasingly large percentage of value added to GDP; (2) create more and better paying jobs than other sectors of the U.S. economy; (3) grow faster than the rest of the U.S. economy; (4) contribute substantially to U.S. foreign sales and exports, outpacing many industry sectors; and (5) make significantly large contributions to the digital economy, which does not even encompass the full scope of the copyright industries' digital activities.⁴ The IIPA report also found that copyright-reliant industries contribute over 63% to the economic value of the digital services sector, which underscores the critical importance of copyright law to America's digital economy.⁵

While the growth of the internet over the last twenty-five years has revolutionized the way that creative works are legally made available and reach their intended audience, it has also facilitated massive amounts of copyright piracy that causes tremendous harm to creators, copyright owners, and consumers. The widespread theft of copyrighted works online is a persistent and evolving problem affecting virtually all types of works and all types of copyright

³ Robert Stoner & Jéssica Dutra, *Copyright Industries in the U.S. Economy: The 2024 Report*, INT'L INTELL. PROP. ALL. (Feb. 2025).

⁴ *Id.* at 21.

⁵ *Id.* at 20.

owners in the digital age, and it undermines the rights of creators, the value of copyright, and our creative economy.

Global online piracy of copyright-protected works results in billions of dollars of economic losses each year, hundreds of thousands of lost jobs, and immeasurable harm to the safety of consumers through the spread of malware, phishing scams, and identity theft.⁶ A study by the Global IP Center found that digital video piracy alone deprives the U.S. economy of a minimum of \$29.2 billion in reduced revenue each year.⁷ This type of piracy not only causes lost revenues to the U.S. creative sectors, it also results in losses to the U.S. economy of between a quarter million and half million jobs and between \$47.5 billion and \$115.3 billion in reduced gross domestic product (GDP) each year.⁸ Piracy also poses a threat to investments in creative works by unjustly enriching bad actors who make no investment and take no risk, at the expense of the creators.

In an article outlining the effects of online piracy, one of today’s panelists, Professor Michael Smith, explains that digital piracy harms creators by reducing their ability to commercialize their creative efforts.⁹ He points to a broad consensus in peer-reviewed academic literature that confirms “that online piracy does exactly what one would expect: it makes it harder for creators and rights owners to make a fair market return on their investments in content creation and dissemination.”¹⁰ In addition to the harms caused to copyright owners, the article summarizes the harms caused to society by reducing creators’ economic incentives to invest in creative output. It explains that economic theory reinforces the Constitutional principle¹¹ that the public

⁶ Impacts of Digital Video Piracy on the U.S. Economy, GLOBAL INNOVATION POLICY CENTER (June 2019), available at: <https://www.uschamber.com/technology/data-privacy/impacts-of-digital-piracy-on-the-u-s-economy>.

⁷ *Id.*

⁸ *Id.* at 14.

⁹ Michael D. Smith, *What the Online Piracy Data Tells Us About Copyright Policymaking*, HUDSON INSTITUTE (April 12, 2023), <https://www.hudson.org/intellectual-property/what-online-piracy-data-tells-us-about-copyright-policymaking>.

¹⁰ *Id.*

¹¹ U.S. Const. art. I, § 8, cl. 8.

interest is promoted by ensuring creators can pursue their own private interests, and that reduced incentives “cause significant problems for both creators and the broader society that benefits from their talents.”¹² Finally, the article cites to significant empirical evidence in the academic literature that the losses in revenues that result from online piracy has harmed consumers by reducing both the quantity and quality of creative output that would have occurred absent piracy.¹³

AI Companies’ Mass Ingestion of Pirated Works from Illicit Sources to Train Their AI Systems is Extremely Harmful to the U.S. Creative Economy

AI companies obtain copyrighted works and other content they use to train their AI systems in many different ways—some legal and some illegal. For example, some companies legally use proprietary materials that they have created themselves or otherwise own.¹⁴ Some may train on material that is in the public domain or not protected by copyright. But the most common example of legal training is through licensing deals that have been and continue to be struck between copyright owners (or their representatives) and AI developers.¹⁵ The robust and ever-expanding licensing-for-training market spans the spectrum of creative works and generative AI models. Many copyright owners are in the unique position of being able to organize, curate, and apply metadata to high quality works and many AI companies recognize the value in clean, ethically sourced, liability free works over indiscriminately scraped masses of material. *The*

¹² Smith, *supra* note 6.

¹³ *Id.*

¹⁴ *Adobe unveils AI video generator trained on licensed content*, Benj Edwards, ARS TECHNICA (Oct. 14, 2024), <https://arstechnica.com/ai/2024/10/adobe-unveils-ai-video-generator-trained-on-licensed-content/>.

¹⁵ There is already high demand for corpora of copyrighted works for ingestion by AI systems, and copyright owners are offering and entering into various licensing agreements. Publishers and copyright owners of scientific and research works such as Elsevier, JSTOR, the Copyright Clearance Center (and many others) have either offered or entered into licensing agreements that allow for text and data mining (TDM) or other generative AI uses. Getty Images has struck several licensing deals with generative AI companies for use of portions of its catalog of stock images for “training.” Multiple news organization, including NewsCorp, the Associated Press, the Atlantic, the New York Times, and the Financial Times, have reached deals with various AI developers. The list goes on and on—with new licensing deals being announced almost daily. See Copyright Alliance, “Generative AI Licensing Isn’t Just Possible, It’s Essential,” available at <https://copyrightalliance.org/generative-ai-licensing/> (citing original sources and media coverage). See <https://copyrightalliance.org/artificial-intelligence-copyright/licensing/>.

current licensing landscape is clear evidence of a free-market licensing system that is working and must be fostered and not disrupted by illegal activity by some bad actors.

In some cases, licenses for AI training are direct, voluntary agreements reached between an AI developer and copyright owner. There are also a growing number of voluntary collecting rights organizations, including Protege, Created by Humans, and ProRata.ai, that voluntarily license the works of many copyright owners to AI companies. Many of these organizations are small companies created within the last several years for the purpose of making voluntary licensing of copyrighted works for training much easier for AI companies and creators by allowing an AI company to voluntarily license many copyrighted works from many creators at one time, through one license agreement. The collecting rights organizations model is not a new one. It has been around for many decades and proven to be successful in many other areas of copyright licensing.

While some AI companies claim that licensing works for training is impossible due to the large amount of material they desire to train a model, that is a self-serving argument that ignores the many different voluntary licensing solutions that have already been (and continue to be) established to meet market demand. A judge recently made this point in a generative AI infringement case, explaining that:

“[T]he suggestion that adverse copyright rulings would stop this technology in its tracks is ridiculous. These products are expected to generate billions, even trillions, of dollars for the companies that are developing them. If using copyrighted works to train the models is as necessary as the companies say, they will figure out a way to compensate copyright holders for it.”¹⁶

Just as with other technological advancements like the internet or streaming that temporarily disrupted the copyright marketplace, new and robust voluntary licensing market solutions

¹⁶ *Kadrey v. Meta Platforms, Inc.*, 3:23-cv-03417, Dkt. 601 at 3 (N.D. Cal.).

support the development of generative AI that benefits both copyright owners and AI companies.

Unfortunately, AI companies do not always license their training materials and instead choose to cut corners by opting for quicker, less expensive, illicit routes. Indeed, using illicitly sourced material for AI training is a problem identified in numerous lawsuits brought by copyright owners against AI companies for the unauthorized mass ingestion of copyrighted works to train their generative AI models. In these lawsuits, the defendant AI companies have not denied their use of such illicit source materials. It is now common knowledge that some AI companies routinely engage in indiscriminate mass ingestion of copyrighted works for the training of their AI systems, which inevitably implicates either or both of copyright owners' reproduction rights by copying works from piratical websites and services (where copies of those works were made available without the owner's authorization) and/or violating the anticircumvention provisions in copyright law by intentionally breaching firewalls and other technical measures for the purpose of illegally downloading copyrighted works that were only intended for authorized consumer use. While some AI datasets involve more discriminating curation, many of the most popular repositories used for large language model (LLM) training contain hundreds of thousands, and in some cases millions, of copies of pirated works that have been knowingly ingested from illicit sources.¹⁷ We want U.S. AI companies to dominate over their foreign counterparts, but engaging in that type of large scale, commercial, and harmful activity is not the way to establish American AI dominance.

There are many different illicit ways some AI developers get pirated material for training. The list below is nonexclusive and includes methods that often overlap with one another.

- *Using pirated works sourced from shadow libraries:* So-called shadow libraries are online repositories containing massive amounts of pirated works, including books, research papers, and other literary works. Many shadow libraries have repeatedly been

¹⁷ *The Unbelievable Scale of AI's Pirated-Books Problem*, Alex Reisner, THE ATLANTIC (Mar. 20, 2025), <https://www.theatlantic.com/technology/archive/2025/03/libgen-meta-openai/682093/>.

found to be illegal and against the public interest, and have had their domains shut down and operators arrested.¹⁸ In fact, illicit services like Library Genesis (aka LibGen), Z-Library, Sci-Hub, and Bibliotik have already been indicted for criminal copyright infringement or listed in the Office of the U.S. Trade Representative’s annual review of Notorious Markets for Counterfeiting and Piracy.¹⁹ Despite this, many AI developers continue to seek out and use these criminal enterprises as a source to build their AI models.

- *Scraping copyrighted works from websites*: Using bots to indiscriminately scrape the internet for copyrighted works that can be used as training material inevitably involves scraping pirated works off illicit websites, whether made available through the above-mentioned shadow libraries or any other online repositories of pirated works. Many AI companies have used and continue to use massive datasets compiled by organizations like Common Crawl, which uses automated software programs to systematically crawl and copy, or “scrape,” the entire internet. Additionally, some web crawlers are designed to bypass firewalls and access legitimate websites by simulating human behavior, further obstructing copyright owners’ protection efforts.
- *Peer-to-peer (P2P) torrenting*: P2P torrenting is a process used to download large amounts of material while often simultaneously redistributing them (through a process known as “seeding” or “leeching”) to other P2P users. In some ongoing lawsuits, AI companies have been shown to have knowingly availed themselves of P2P networks that are notorious hotbeds of digital piracy and copied an untold number of pirated copies of copyrighted works using torrent services. In at least some cases, it is alleged

¹⁸ Brief for the International Association of Scientific, Technical, and Medical Publishers as Amicus Curiae Supporting Plaintiffs at 2, *Kadrey v. Meta Platforms, Inc.*, 3:23-cv-03417, (N.D. Cal.).

¹⁹ See generally OFFICE OF THE U.S. TRADE REPRESENTATIVE, *2024 Review of Notorious Markets for Counterfeiting and Piracy*, (Jan. 8, 2025) [https://ustr.gov/sites/default/files/2024%20Review%20of%20Notorious%20Markets%20of%20Counterfeiting%20and%20Piracy%20\(final\).pdf](https://ustr.gov/sites/default/files/2024%20Review%20of%20Notorious%20Markets%20of%20Counterfeiting%20and%20Piracy%20(final).pdf); see also OFFICE OF THE U.S. TRADE REPRESENTATIVE, *2023 Review of Notorious Markets for Counterfeiting and Piracy*, at 27 & 30 (Jan. 30, 2024), https://ustr.gov/sites/default/files/2023_Review_of_Notorious_Markets_for_Counterfeiting_and_Piracy_Notorious_Markets_List_final.pdf.

that an AI company has not only downloaded but uploaded the works through seeding or leeching. Such redistribution of massive amounts of copyrighted works is incredibly harmful and possibly criminal.

The harm online piracy already causes to copyright owners is no doubt exacerbated when pirated works are reproduced, and in some instances redistributed, by AI companies that employ some or all of the mechanisms above to amass training material. Specifically, the harm caused by illicit shadow libraries, which is already massive, is compounded by AI companies that access the pirated works through torrenting because it enables further downstream dissemination. It also contributes to the distribution of illegitimate copies of scientific and medical works that contain misinformation or have been retracted or updated.

Mass Ingestion of Pirated Works for AI Training Should Weigh Heavily Against Fair Use

In the most recent report in its ongoing study of the copyright issues raised by generative AI, the U.S. Copyright Office addresses the mass ingestion of copyrighted works for training by AI developers. Ultimately, the Office concludes that, “the knowing use of a dataset that consists of pirated or illegally accessed works should weigh against fair use.”²⁰ Judges in two generative AI infringement cases in the Northern District of California recently issued orders on fair use related to the use of pirated works for AI training.²¹ Similar to the Copyright Office, both Judges expressed serious concerns over the defendants AI companies’ mass ingestion of pirated works. In one of the cases, *Bartz v. Anthropic*, the Judge’s order describes how Anthropic downloaded over 7 million copies of pirated books and concludes, “[p]iracy of otherwise available copies is inherently, irredeemably infringing even if the pirated copies are immediately used for the transformative use and immediately discarded.”²² While the Judge in the other case, *Kadrey v. Meta*, explained that “[i]n the vast majority of cases, this sort of peer-to-peer file-sharing will

²⁰ *Id.*

²¹ *Bartz v. Anthropic PBC*, 3:24-cv-05417, Dkt. 231 (N.D. Cal.); *Kadrey v. Meta Platforms, Inc.*, 3:23-cv-03417, Dkt. 598 (N.D. Cal.).

²² *Bartz*, at 19.

constitute copyright infringement.”²³

These two court cases and many other pending AI infringement cases are distinguished from past fair use cases that are relied on by AI companies. Those past cases involved the use of one or many *legitimate* copies, and that use was found to be fair use because the copy was not otherwise available. In contrast, many of the pending AI training infringement cases involve the use of *millions of pirated* copies. Never has a court confronted the use of pirated copies on such a massive scale. Moreover, copyrighted works that AI companies wish to use for training are largely available through legitimate online sources, which is in stark contrast to past cases where works were not licensable or otherwise available. Sourcing mass amounts of pirated copyrighted works is not necessary for the development of generative AI models, it is simply the least expensive and fastest way for some AI companies to get the vast quantity of works they want.

In sum, the major distinction between these AI cases and past fair use cases is (i) the sheer scale of what is being ingested by AI companies, (ii) the fact that AI companies are using pirated copies, not legitimately made copies, and (iii) the sheer scale of competing outputs that are generated by using these pirated works. In these AI infringement cases, there are millions of pirated works being illegally copied and used to produce millions of outputs that may compete with and harm the market for the original works. The fact that AI companies intentionally use these illicit sites and services does not presumptively disqualify an AI developer from claiming fair use is antithetical to the foundations of our copyright system and the rights it guarantees.

Strong Enforcement of Copyright Laws is Integral to Our National Security

Some AI companies claim that limiting the use of copyrighted material for AI training amounts to a national security threat because it could impede AI innovation in the United States, potentially granting geopolitical competitors like China a technological advantage. Those claims could not be more wrong.

Promoting strong copyright laws and the Constitutional guarantees of the protection of human

²³ *Kadrey*, at 20.

creators is critical to maintaining American AI dominance, which in turn will diminish national security threats. American culture and arts are key tools for spreading American values and advancing democracy and freedom across the globe. The global popularity of American books, movies, music, and art provides vehicles by which we are able to project American values to the rest of the world. Tearing down American creativity leaves us and the rest of the world vulnerable to Chinese propaganda and is an obvious and dangerous step in the wrong direction.

Before suggesting an unwarranted upheaval to U.S. copyright law for the purposes of national security, AI companies should clean their own house. For example, it has been widely reported that China has been covertly using “ChatGPT to spread propaganda, manipulate social media engagement, and target journalists and politicians in a coordinated AI-powered influence campaign.”²⁴ And ChatGPT is not alone. A report issued at the end of June by the American Security Project concluded that:

“The Chinese Communist Party’s aggressive censorship laws and disinformation campaigns have resulted in a proliferation of propaganda and censorship across the global AI data marketplace. AI-powered chatbots in the United States now regurgitate CCP propaganda in Chinese and English when prompted on certain topics, posing significant ramifications for global AI development and U.S. national security.”²⁵

If AI companies want to talk about national security, the conversation must start with this very real threat to our national security, instead of the contrived threat of copyright law.

Additional Final Points

It is important to understand that copyright owners’ rights are implicated when their works are

²⁴ *OpenAI Reveals China Covertly Used ChatGPT To Spread Propaganda, Manipulate Social Media Engagement, And Target Journalists And Politicians In A Coordinated AI-Powered Influence Campaign*, Ezza Ijaz, WCCFTECH (June 6, 2025), <https://wccftech.com/openai-reveals-china-covertly-used-chatgpt-to-spread-propaganda-manipulate-social-media-engagement-and-target-journalists-and-politicians-in-a-coordinated-ai-powered-influence-campaign/>.

²⁵ *Evidence of CCP Censorship, Propaganda in U.S. LLM Responses*, Courtney Manning, Monique Shum, and Josey Walden, THE AMERICAN SECURITY PROJECT (June 25, 2025), <https://ai.americansecurityproject.org/research/ccp-censorship-in-llm-responses>.

reproduced to create datasets or ingested by generative AI systems, regardless of whether the AI systems generate infringing output or distribute infringing copies to end users. Some AI companies like to argue that there is no harm to copyright owners if their models do not generate infringing outputs, but that is only half the story. The mass ingestion of pirated works indisputably implicates a copyright owner's right of reproduction, which is the foremost right guaranteed by Section 106 of the Copyright Act. AI developers' choice to ingest massive amounts of pirated works to train their AI systems instead of entering into licensing agreements with copyright owners may also harm the copyright owners whose works have been ingested by negatively impact their ability to commercialize their works, recoup investments, and the incentivization to create new works.

Finally, a separate issue we would like to bring to the attention of the Subcommittee is legislation intended to block access to large-scale commercial foreign piracy sites. The Copyright Alliance strongly supports the establishment of a judicial blocking system that would allow copyright owners to petition a federal court to issue an order requiring a U.S. internet service provider (ISP) to prevent foreign-based websites and services that have a primary purpose of providing access to infringing material from providing pirated content to U.S. consumers. Over 50 nations across the globe have established similar systems that have been proven to effectively curb piracy and promote legitimate services without harming free expression or breaking the internet.²⁶ Efforts to establish a judicial blocking system have bipartisan and bicameral support, and we hope to soon see the introduction of legislation that will help copyright owners combat harmful foreign-based piracy.

²⁶ *Blocking Access to Foreign Pirate Sites: A Long-Overdue Task for Congress*, Rodrigo Balbontin, INFORMATION TECHNOLOGY & INNOVATION FOUNDATION (June 9, 2025), <https://itif.org/publications/2025/06/09/blocking-access-to-foreign-pirate-sites-a-long-overdue-task-for-congress/>.