

Testimony of Maxwell V. Pritt

***Too Big to Prosecute?: Examining the AI Industry's Mass Ingestion of
Copyrighted Works for AI Training***

**Hearing Before the
Committee on the Judiciary
Subcommittee on Crime and Counterterrorism
United States Senate**

July 16, 2025

Washington, DC

I. Introduction

Chairman Hawley, Ranking Member Durbin, and members of the Subcommittee: thank you for the invitation and opportunity to testify before you today.¹

Today, the Committee considers what is likely the largest domestic piracy of intellectual property in our nation's history. That piracy includes hundreds of terabytes of data and many millions of works, including, for example, at least 12 books authored by members of this subcommittee (three authored by Chairman Hawley alone). The culprits of this unprecedented piracy are, incredibly, some of our nation's largest technology companies. They want vast troves of written text, including a limitless number of copyrighted works, to develop their artificial intelligence models. But these massively profitable tech companies don't want to pay for it. Perhaps they're Bob Dylan fans: "Steal a little and they throw you in jail / Steal a lot and they make you a king."² These decisions to engage in mass piracy were made

¹ I am a Partner at the law firm Boies Schiller Flexner LLP, and am in charge of the firm's San Francisco office. I litigate high-stakes cases for plaintiffs and defendants in courts across the country on issues ranging from antitrust to intellectual property to constitutional and contractual rights. I currently represent authors, artists, and programmers in copyright infringement cases against AI companies including Meta, OpenAI, GitHub, and Midjourney. I am a member of the Judicial Council of California and an Appellate Lawyer Representative for the U.S. Court of Appeals for the Ninth Circuit. I previously served on the California State Bar's Judicial Nominees Evaluation Commission, which vets state judicial candidates, and am Chair Emeritus of the Bay Area Lawyer Chapter of the American Constitution Society. I also taught at Stanford Law School and my alma mater, UC Law San Francisco. After law school, I clerked for Ninth Circuit Chief Judge Mary Murguia (on the district court) and Judge Marsha Berzon.

² Bob Dylan, "Sweetheart Like You" (Columbia Records 1983).

at the highest levels of the tech companies. At Meta, company documents show that Mark Zuckerberg himself made the call. Company documents at Anthropic also show a blatant disregard for our copyright laws, preferring to pirate books to avoid or delay the “legal/practice/business slog,” as Anthropic’s co-founder and CEO Dario Amodei put it.

AI companies now want a pass under a limited exception to infringement—the “fair use” doctrine—that Congress codified in Section 107 of the Copyright Act of 1976. But while these tech companies invoke fair use as a shield in litigation, they know piracy is illegal. As one Meta employee put it: “it’s the piracy (and us knowing and being accomplices) that’s the issue.”³ As others at Meta explained: “if there is media coverage suggesting we have used a dataset we know to be pirated, such as LibGen, this may undermine our negotiating position with regulators on these issues.”⁴

As AI companies scrambled to outpace each other, many of them turned to illegal pirate websites—massive repositories of tens of millions of stolen copyrighted works—to get text for their AI models. By pirating these works for free rather than buying or licensing them from copyright owners, AI companies have built a multibillion-dollar industry generally without paying a single cent to either the creatives whose works are powering their products or the publishers responsible for

³ *Kadrey et. al. v. Meta*, No. 3:23-CV-3417, Pl’s Mot. for Partial Summary Judgment (N.D. Cal., April 30, 2025), Dkt. 574 at 7.

⁴ *See Kadrey et. al. v. Meta*, No. 3:23-CV-3417, Pl’s Opp. to Meta’s Mot. for Summary Judgment (N.D. Cal., April 30, 2025), Dkt. 575 at 37.

introducing and providing those works to the public.⁵ Tech companies’ unapologetic use of these illegal websites has also revived international digital piracy by propping up the for-profit, criminal syndicates that use these illicit marketplaces to violate U.S. copyrights abroad.

The cost-benefit analysis was simple, particularly for tech companies caught flat-footed by OpenAI’s unexpected release of ChatGPT: Expend time and resources to legally acquire the rights to copyrighted books and articles from those who own the rights; or pirate them all for free now from illegal websites and pay litigation damages later—or, even more appealing, pay nothing at all if they can convince the courts to excuse their unprecedented commercial piracy as fair use. The rapid rise of generative AI technology has thus ushered in a new era of domestic piracy on a scale never before seen and by extraordinarily powerful corporate interests seeking a quick profit off the backs of the creative industries that contribute over \$1 trillion in value to the U.S. economy.

I. Illegal Pirate Websites and Peer-to-Peer File Sharing Networks

Digital piracy costs American businesses tens of billions of dollars annually. Each year, countless copyrighted works are made available, downloaded, and distributed from notorious pirate websites such as Library Genesis (“LibGen”) and Z-Library through various methods, including decentralized peer-to-peer file-sharing systems like BitTorrent. BitTorrent is a technological method for downloading and

⁵ Belying the AI companies’ arguments, there is a robust and growing licensing market for AI training data. *See Kadrey v. Meta*, Dkt. 575 at 26.

uploading data over the internet. A key feature of BitTorrent is making available and “sharing” data—and to acquire data fast, you share it both simultaneously (called leeching) and after it is downloaded (called seeding). So as the latest pirated Taylor Swift album or Game of Thrones book is copied to your server, you’re also making another copy and sending it out into the ecosystem to others. In this regard, torrenting is a system where “the *downloaders* of a file *barter* for chunks of it by uploading and downloading them in a tit-for-tat-like manner to prevent parasitic behavior.”⁶

The Judiciary Committee last confronted the specific problem of peer-to-peer file sharing in 2000, when Napster and other music sharing platforms threatened to gut the music industry by offering a platform where users could exchange copyrighted songs for free and without permission of the rightsholders.⁷ When the Ninth Circuit held that Napster’s wide-ranging piracy violated U.S. copyright law,⁸ the company was forced to shut down. Legitimate online music markets proliferated almost immediately: first, iTunes, and later, streaming services like Spotify.⁹ For the next quarter of a century, Congress and the Executive worked with copyright holders to combat global piracy, which the U.S. Chamber of Commerce estimates costs the U.S.

⁶ Johan Pouwelse et al., *The BitTorrent P2P File-Sharing System: Measurements and Analysis*, IPTPS 2005, at 206 (2005) (emphasis in original).

⁷ *Music On The Internet: Is There An Upside to Downloading? Before the S. Comm. on the Judiciary*, 106th Cong. (July 11, 2000).

⁸ *A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004 (9th Cir. 2001).

⁹ 1 Lindey on Entertainment, Publ. & the Arts § 2:28 n. 36 (3d ed. 2024).

economy nearly \$30 billion in lost revenue each year.¹⁰ As of 2017, before AI companies' mass piracy started, eBook piracy cost U.S. publishers \$315 million in annual lost sales.¹¹

The recent discovery of U.S. tech companies' mass piracy creates a seismic shift—from targeting criminal enterprises abroad to combatting piracy here at home. Pirate websites function similarly to Napster, but for books, rather than music. Whenever an individual uses an illegal pirate website in lieu of a legitimate bookseller to acquire a book, that book's author and publisher are directly harmed by the loss of a sale in an otherwise functioning and well-established market for books. A 2016 study indicated that pirated eBooks depress legitimate book sales by as much as 14%.¹²

Furthermore, as the Office of the U.S. Trade Representative documented in its survey of “Notorious Markets for Counterfeiting and Piracy,”¹³ which included LibGen in two recent editions, the harms of piracy are far-reaching, impacting not just vibrant U.S. economies reliant on legitimate markets. In the case of books,

¹⁰ U.S. Chamber of Commerce, *Impacts of Digital Piracy on the U.S. Economy*, (June 15, 2019) <https://www.uschamber.com/technology/data-privacy/impacts-of-digital-piracy-on-the-u-s-economy>.

¹¹ Adam Row, *U.S. Publishers Are Still Losing \$300 Million Annually To Ebook Piracy*, FORBES (July 28, 2019).

¹² Imke Reimers, *Can Private Copyright Protection Be Effective? Evidence from Book Publishing* (2016), 59 J. L. & ECON. 411, 414 (2016) <https://doi.org/10.1086/687521>.

¹³ See Office of the U.S. Trade Representative, *USTR Releases 2024 Review of Notorious Markets for Counterfeiting and Privacy* (Jan. 8, 2025) <https://ustr.gov/about/policy-offices/press-office/ustr-archives/2007-2024-press-releases/ustr-releases-2024-review-notorious-markets-counterfeiting-and-piracy>.

internet piracy harms artists, graphic designers, bookstores, publishers, printing presses, copy editors, and others working in creative economies. This rampant piracy undermines critical U.S. comparative advantages by hampering innovation and creativity.¹⁴

Since 2015, federal courts have consistently held that pirate websites such as LibGen violate U.S. copyright law.¹⁵ Pirate websites are also routinely targeted by government enforcement actions. Authorities regularly shut down their domains and even prosecute the perpetrators.¹⁶ It is hard to reconcile from any good-faith policy or legal perspective how illegal websites that traffic in piracy and are permanently enjoined by federal courts can simultaneously exist as legitimate sources for AI companies.

II. Fair Use & Trends in AI Litigation

The last few years have demonstrated both the potential and pitfalls of generative AI. While there is no evidence yet to support the companies' far-reaching marketing claims such as curing disease, the data analysis capabilities of the machines are impressive. In addition to large language models ("LLMs"), image,

¹⁴ *Digital Copyright Piracy: Protecting American Consumers, Workers, and Creators, Hearing Before the Subcomm. on Courts, Intel. Prop., and the Internet of the H. Comm. on the Judiciary* 118th Cong. 68 (December 13, 2023) (Statement of the Association of American Publishers, at 3).

¹⁵ *Elsiever Inc. v. www.Sci-Hub.org*, 2015 WL 6657363 (S.D.N.Y. 2015); *see also Cengage Learning, Inc. v. Does 1-50 d/b/a Library Genesis*, Case No. 23-cv-8136-CM, Dkt. 36 at 2-3 (S.D.N.Y. Sept. 24, 2024) (granting permanent injunction against LibGen for copyright infringement).

¹⁶ *See* Indictment, *United States v. Napolsky et al.*, No. 1:22-CR-525 (E.D.N.Y. Nov. 16, 2022), Dkt. 4 (criminal indictment of Z-Library perpetrators).

video, and music generators are beginning to proliferate. But the more we learn about generative AI technology, the more it has become clear that many AI companies have cut corners in their race to be the biggest and the best. Under competitive pressure, AI companies opted to play fast and loose with our intellectual property because they know how valuable it is but think they can get away with not paying for it by claiming “fair use.”

Training data is the raw material that powers the AI industry. AI-generated outputs are a direct product of the data on which AI models train. Training on scientific articles, like medical journals, is what allows LLMs to analyze medical issues. Training on software code is what allows LLMs to generate code for programmers. And training on literary works is what allows LLMs to write creatively.

AI companies have long recognized the value of copyrighted books and long-form text as training data for LLMs. But instead of buying and licensing books from authors or publishers, massive tech companies like Meta, OpenAI, and others abandoned efforts to acquire books through legitimate means and instead turned to piracy, sometimes using peer-to-peer file sharing networks to acquire books and, in the case of Meta, even copying and distributing vast amounts of pirated data to others.

A. Exemplar Case: Meta’s Mass Piracy

Much of Meta’s exploitation of pirated copyrighted works is now public through the efforts of my firm and our co-counsel on behalf of authors. From the very early

days of its GenAI program, Meta concluded that training its LLM, called Llama, on books would improve its performance. Meta also believed its competitors were obtaining superior results due to their use of pirated datasets to train their LLMs.¹⁷ In response, Meta devised a simple strategy to catch up: acquire more books—and fast.¹⁸ In October 2022, Meta’s AI team began seeking in-house legal approval for “pure exploration work” regarding the performance benefits that could be achieved by training Llama on books and articles obtained from LibGen.¹⁹ Meta initially planned to use LibGen only to test its value and then set up licensing agreements.²⁰ When Meta’s legal team approved the plan, Meta downloaded 2.2 million books from LibGen via torrenting.²¹ Unsurprisingly, books proved valuable and improved model performance.

In early 2023, Meta scrambled to gather more text data for subsequent iterations of the Llama models.²² Managers stressed the need to get as many books as possible as quickly as possible.²³ Over the next couple of months, Meta briefly

¹⁷ *Kadrey v. Meta*, Dkt. 574 at 6.

¹⁷ *Id.*

¹⁸ *Id.*

¹⁹ *Id.*

²⁰ *Id.*

²¹ *Id.* at 6-7.

²² *Id.* at 7.

²³ *Id.* at 8.

explored licensing books.²⁴ Licensing budgets as high as \$200 million were floated.²⁵ But then, in April 2023, Meta’s short-lived licensing efforts came to a halt.²⁶ Meta executives instructed Meta’s business development team—the team tasked with licensing—to stop all text licensing efforts.²⁷ And Meta instead resorted to using pirated copyrighted works from LibGen, which it continued to copy and use as a key component of its commercially available Llama models.²⁸ To conceal its actions, Meta’s employees started referring to these pirated datasets as “external” or “publicly available” instead of “pirated.”²⁹ Then, in May 2023, Meta employees torrented even more copyrighted works from LibGen, this time using LibGen’s contents to cross-reference against major publishers’ catalogues to determine whether licensing efforts would be necessary at all.³⁰ This “gap approach” showed that Meta viewed pirated books simply as a free substitute for licensed books. And why license if it could pirate for free?

After discovering that LibGen contained most of the books it needed, Meta’s executives greenlighted its use as training data for Llama. Documents uncovered during litigation confirmed that after an escalation to Meta CEO Mark Zuckerberg,

²⁴ *Id.*

²⁵ *Id.*

²⁶ *Id.* at 8.

²⁷ *Id.* at 8-9.

²⁸ *Id.* at 9.

²⁹ *Id.*

³⁰ *Id.*

the GenAI team was “approved to use LibGen for Llama 3.”³¹ That internal memorandum, which documented Mr. Zuckerberg’s approval, admits that the approval came despite the fact that LibGen is “a dataset we know to be pirated.”³² Notably, that memo also included in-house counsel, who apparently advised that “in **no case** would we disclose publicly that we had trained on libgen,” because “if there is media coverage suggesting we have used a dataset we know to be pirated, such as LibGen, this may undermine our negotiating position with regulators on these issues.”³³

When an August 2023 exposé published in *The Atlantic* revealed broadly that Meta had been training on copyrighted works without permission, one employee worried that the public could realize that Meta was continuing to use pirated data for training: “It’s the piracy (and us knowing and being accomplices) that’s the issue,” she remarked.

Meta went to great lengths to conceal its use of illegal websites and pirated copyrighted works. Senior leadership and engineers knew, but it was only conveyed to others on a “‘need to know’ basis.”³⁴ Indeed, several employees expressed strong reservations. One Meta researcher commented, “I feel that using pirated material should be beyond our ethical threshold.”³⁵ Another referred to LibGen as an “illegal

³¹ *Id.* at 10.

³² *Id.*

³³ *Id.* (Ex. 61) (emphasis in original).

³⁴ *Id.* at 7.

³⁵ *Id.* at 10.

pirated website[]” and expressed that “it should not go in the training of the published model.”³⁶

But as 2023 progressed and Meta’s models increased in size, Meta’s hunger for high-quality data increased. In late 2023, as development started for Llama 4, Meta’s engineers explored the use of Anna’s Archive, an aggregator of many illegal pirated datasets. Meta engineers confirmed that Anna’s Archive contained substantially all of LibGen, nearly all of Z-Library, and over two-thirds of an additional database called Internet Archive.³⁷ So Meta began downloading and processing copyrighted works from the Anna’s Archive aggregation of illegal pirate websites, despite employees calling it a “pretty shady website” that “won’t be popular with the lawyers.”³⁸

To speed up its acquisition of terabytes upon terabytes of pirated works—and ultimately to keep pace with its pirate website-using competitors—Meta resorted to torrenting, the peer-to-peer file sharing protocol discussed above. This protocol optimizes speed by simultaneously making available and transmitting content being downloaded to others.³⁹ Meta’s engineers were well aware of the legal risks posed by torrenting data, but they decided to do it anyway, apparently with the approval of Meta’s in-house counsel.⁴⁰ One employee told others: “Btw, it would not be trivial to

³⁶ *Id.*

³⁷ *Id.* at 11.

³⁸ *Id.*

³⁹ *Id.* at 11-12.

⁴⁰ *Id.* at 12.

download libgen if everything is in torrents,” sharing a link to a Quora webpage asking, “What is the probability of getting arrested for using torrents in the USA?”⁴¹ Meta opted to hedge some of this legal risk—and the risk of law enforcement or public discovery generally—by using Amazon Web Services (“AWS”) for its torrenting activities, a deviation from Meta’s usual practice.⁴² When an engineer asked why Meta infrastructure could not be used, he was told it was to avoid risk of tracing the torrenting to Meta’s servers.⁴³ And, although Meta could have changed the default settings on its torrent client to completely prevent uploading, it did not do so, presumably because changing those settings would decrease the speed of Meta’s downloads.⁴⁴

Meta’s embrace of data piracy increased exponentially over time. Between April and July 2024 alone, Meta downloaded over **134 terabytes** from pirate websites through Anna’s Archive and uploaded over **40 terabytes** of pirated data to third parties.⁴⁵

Meta has attempted in litigation to justify its piracy by arguing it had no choice but to pirate the books because legitimate acquisition was prohibitively expensive or time-intensive. But ability-to-pay was never an obstacle for a company like Meta. This massive piracy occurred at the hands of one of the world's richest companies,

⁴¹ *Id.*

⁴² *Id.* at 14.

⁴³ *Id.* at 14-15.

⁴⁴ *Id.* at 14

⁴⁵ *Kadrey v. Meta*, Dkt. 562-50 at 4.

which has invested huge sums of money into other aspects of its AI program, including data centers and talent. Meta plans to spend hundreds of billions of dollars to build AI data centers⁴⁶ and reportedly has offered gargantuan recruitment bonuses to individual data scientists,⁴⁷ yet Meta has not paid a single cent for the copyrighted works it pirated.

B. Other U.S. Tech Companies' Piracy

Meta is far from alone in its conduct. Piracy has become endemic to the GenAI industry. Anthropic, the company behind the LLM known as Claude, also pirated millions of books to build its LLM. A recent decision in *Bartz v. Anthropic* noted that the company downloaded over seven million pirated copies of books from LibGen and another illegal website called Pirate Library Mirror and paid nothing.⁴⁸ OpenAI used LibGen too.

By downloading millions upon millions of books and other copyrighted works from pirate websites and then using those unauthorized copies for their AI products, these companies have committed copyright infringement on a massive scale. These historic acts of domestic piracy have deprived the U.S. creative industry of billions of dollars. Further, Meta and any other company that used a torrenting network like BitTorrent to source pirated works has perpetuated the copyright infringement there

⁴⁶ Jaspreet Singh and Aditya Soni, *Meta's Zuckerberg pledges hundreds of billions for AI data centers in superintelligence push*, REUTERS (July 14, 2025).

⁴⁷ *Sam Altman says Meta offered \$100 million bonuses to OpenAI employees*, REUTERS (June 18, 2025).

⁴⁸ *Bartz v. Anthropic*, No. 3:24-CV-5417 (N.D. Cal. June 23, 2025), Dkt. 231 at 18.

on both a massive and exponential scale by making and distributing additional copies of pirated works to others on the network, who in turn could distribute those works to other network participants, and so on. In fact, one major pirate websites boasts that AI companies saved the illicit digital books market: “Not too long ago, ‘shadow-libraries’ were dying. Sci-Hub, the massive illegal archive of academic papers, had stopped taking in new works, due to lawsuits. ‘Z-Library’, the largest illegal library of books, saw its alleged creators arrested on criminal copyright charges. They incredibly managed to escape their arrest, but their library is no less under threat . . . Then came AI. Virtually all major companies building LLMs contacted us to train on our data.”⁴⁹

III. Piracy Is Incompatible with “Fair Use”

Meta and other AI companies now seek to defend their massive, systemic infringement in court. In defense, the AI companies argue their infringement was “fair use.”

Fair use is an *exception* that allows for *limited* use of copyrighted material without permission from the copyright owner under certain conditions. The doctrine is meant to balance the rights of creators with the public interest in freedom of expression and access to information. For example, a musician who creates a parody of a song usually will not be liable for copyright infringement because the fair use doctrine protects her. Why? Because parodies are considered sufficiently

⁴⁹ Anna’s Archive, *Copyright reform is necessary for national security* (January 31, 2025), <https://annas-archive.org/blog/ai-copyright.html>.

“transformative” (among other reasons, a parody cannot exist without using the copyrighted work it parodies) that they do not operate as a substitute for the original and in so doing, do not harm the market for the original. While courts analyze fair use under four non-exhaustive factors, the analysis often focuses on questions of the purpose of the copying, whether it is transformative, and “market substitution” for actual or potential markets of *each use of each copy* the alleged infringer makes of the work.

But internet piracy is the antithesis of fair use. Downloading millions of pirated works from known illegal databases created by foreign actors to avoid compensating American creators and rightsholders is not “transformative”: it does nothing to create new expression or facilitate meaningful dialogue. Internet piracy also functions as pure market substitution—taking for free what otherwise must be purchased.

For over a century, courts have held that unmitigated piracy of copyrighted works, *i.e.*, the duplication of entire works to avoid compensating rightsholders, is not fair use.⁵⁰ That through-line has been applied to digital piracy—the illegality of downloading and sharing copyrighted material has been well-established since the

⁵⁰ See, e.g., *Folsom v. Marsh*, 9 F. Cas. 342, 342-45 (C.C.D. Mass. 1841) (explaining “it is as clear, that if [defendant] thus cites the most important parts of the work, with a view, not to criticise, but to supersede the use of the original work, and substitute the review for it, such a use will be deemed in law a piracy”) (Story, J.); see, e.g., *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 550 (1985) (“As Justice Story’s hypothetical illustrates, the fair use doctrine has always precluded a use that ‘supersede[s] the use of the original.’”).

days of Napster.⁵¹ The uncontroversial premise behind this conclusion is that for fair use to apply, the work that was copied must have been lawfully obtained in the first place.⁵²

Whatever the merits of generative AI, stealing copyrighted works off the internet for one's own benefit has always been unlawful.⁵³ While everyone expected these historic AI copyright cases to test the boundaries of the “fair use” defense with respect to training AI models on copyrighted works, no one expected that billion- and trillion-dollar companies would be arguing—and courts would be contemplating—that mass piracy for commercial use could also fall within the ambit of the fair use defense.

As the U.S. Copyright Office recently explained in a seminal report on copyright and AI, “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.”⁵⁴ Deeming rampant digital piracy fair use would mark the first time in

⁵¹ See, e.g., *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, at 919 (2005) (“[O]ne who distributes a device with the object of promoting its use to infringe copyright . . . is liable for the resulting acts of infringement by third parties.”); *see also United States v. Slater*, 348 F.3d 666, 669 (7th Cir. 2003).

⁵² *Atari Games Corp. v. Nintendo of Am., Inc.*, 975 F.2d 832, 843 (Fed. Cir. 1992) (“To invoke the fair use exception, an individual must possess an authorized copy of a literary work.”) (emphasis added).

⁵³ *UMG Recordings, Inc. v. MP3.Com, Inc.*, 2000 WL 710056, at *1 (S.D.N.Y. June 1, 2000) (the “mere fact” that copyright infringement is “clothed in the exotic webbing of” a new technology “does not disguise its illegality”) (Rakoff, J.).

⁵⁴ United States Copyright Office, Report on Copyright and Artificial Intelligence: Part 3 (May 9, 2025) (pre-publication version) (emphasis added).

history that piracy and trafficking in stolen goods are given a pass under U.S. copyright law. As Judge Alsup put it: “There is no carveout [] from the Copyright Act for AI companies.”⁵⁵

IV. AI Companies’ Piracy Is a Bipartisan Issue

Tech companies’ mass digital piracy affects everyone irrespective of political persuasion. This is a bipartisan issue. For example, because Meta torrented LibGen and Anna’s Archive, we know it pirated the following books written by members of this very Subcommittee:

- “The Tyranny of Big Tech,” by Senator Josh Hawley
- “Antitrust: Taking on Monopoly Power from the Gilded Age to the Digital Age,” by Senator Amy Klobuchar
- “United: Thoughts on Finding Common Ground and Advancing the Common Good,” by Senator Cory Booker
- “A Time for Truth: Reigniting the Promise of America,” by Senator Ted Cruz
- “Life Equity: Realize Your True Value and Pursue Your Passions at Any Stage in Life,” by Senator Marsha Blackburn
- “Manhood: The Masculine Virtues America Needs,” by Senator Josh Hawley
- “Theodore Roosevelt: Preacher of Righteousness,” by Senator Josh Hawley
- “Justice Corrupted: How the Left Weaponized Our Legal System,” by Senator Ted Cruz

⁵⁵ *Bartz v. Anthropic*, Dkt. 231 at 14.

- “One Vote Away: How a Single Supreme Court Seat Can Change History,” by Senator Ted Cruz
- “Unwoke: How to Defeat Cultural Marxism in America,” by Senator Ted Cruz
- “The Joy of Politics: Surviving Cancer, a Campaign, a Pandemic, an Insurrection, and Life's Other Unexpected Curveballs,” by Senator Amy Klobuchar
- “The Senator Next Door: A Memoir from the Heartland,” by Senator Amy Klobuchar

Meta also pirated books written by every President and Vice President in the 21st century, including:

- “The Art of the Deal,” by President Donald Trump
- “Hillbilly Elegy: A Memoir of a Family and Culture in Crisis,” by Vice President J.D. Vance
- “Promise Me, Dad: A Year of Hope, Hardship, and Purpose,” by President Joseph Biden
- “The Truths We Hold: An American Journey,” by Vice President Kamala Harris
- “So Help Me God,” by Vice President Mike Pence
- “A Promised Land,” by President Barack Obama
- “Decision Points,” by President George W. Bush

- In My Time: A Personal and Political Memoir,” by Vice President Dick Cheney

Meta did not pay for its use of any of these books. Moreover, Meta’s indiscriminate torrenting means that countless copies of these books were made and distributed by Meta to other internet pirates—each of which constituted yet another lost sale.

V. Conclusion

In conclusion, I would like to address a common refrain stated publicly by some AI companies—that in order to keep pace with China, the United States has no choice but to excuse widespread digital piracy as fair use. That premise is simply untrue. Protecting intellectual property rights is a core American value and has been one since our founding. As Senator Leahy, who once chaired the Judiciary Committee, said: “the intellectual property generated in our country is the envy of the rest of the world.”⁵⁶ U.S. intellectual property law has always fostered innovation, not hindered it.

This nation’s commitment to protecting intellectual property—including U.S. tech companies’ IP—has made us more competitive, not less. The threat of innovation in foreign countries has never been grounds for this country to excuse rampant lawbreaking and abandon fundamental principles of the rule of law. Nor should it be now.

⁵⁶ *Music On The Internet: Is There An Upside to Downloading? Before the S. Comm. on the Judiciary*, 106th Cong. (July 11, 2000). See also U.S. Chamber of Commerce, 2023 International IP Index.

To be clear, a policy of responsible AI simply requires adherence to our nation's most foundational legal principles and the existing laws governing intellectual property rights. No one is above the law—and certainly not billion- and trillion-dollar tech companies. Congress can—and must—promote AI growth and innovation to secure U.S. dominance while also promoting the progress of science and the arts. To maintain U.S. dominance across cultural output in the arts and sciences, Congress must protect our creative industries by, at the very least, aiding enforcement of the copyright laws as they already exist on the books against IP pirates, without favoring those who have shown they are most capable of infringing copyrights on a massive scale.