TECHSTRONG RESEARCH

SPECIAL REPORT | SEPTEMBER 2026

Does Copyright Law Lose to National Security?

How Washington adopted the AI industry's argument for uncompensated training, why publishers are being asked to finance America's AI race, and what happens to private property when "strategic necessity" enters the courtroom

ALAN SHIMEL

Founder and CEO, Techstrong Group

TECHSTRONG RESEARCH / SPECIAL REPORT

Does Copyright Law Lose to National Security?

How Washington adopted the AI industry's argument for uncompensated training, why publishers are being asked to finance America's AI race, and what happens to private property when "strategic necessity" enters the courtroom

By Alan Shimel

The administration has connected access to copyrighted training material with American economic strength, military capability and competition with China. That turns a private fair-use dispute into a public-policy question: if privately financed creative work is necessary to a national project, who should bear the cost?

CONTENTS

Shocked, Shocked

As Captain Renault might have put it, I am shocked, shocked to discover that the United States government believes the copyright interests of publishers should take a back seat to the needs of the AI industry.

There is nothing shocking about OpenAI believing it should be allowed to train its models on copyrighted material without paying every copyright owner. Meta and Anthropic have made similar arguments. Model developers have an obvious economic interest in defining the books, articles, photographs, music, software and other material they use as something they are free to consume.

What is new is that the federal government has now entered the courtroom to help make their case.

On September 1, 2026, the Department of Justice filed a statement of interest[1] in the consolidated copyright litigation involving OpenAI, The New York Times and other authors and publishers. The government did not merely argue for a cautious or fact-specific application of fair use. It urged the court to reject broad theories of liability based on the use of copyrighted works in model training and described that training as extraordinarily transformative. It argued that restricting access to training material could slow scientific progress, weaken the American economy and allow foreign competitors, particularly China, to overtake the United States in artificial intelligence.

The government connected a private company's demand for access to copyrighted material with battlefield targeting, intelligence analysis, drones, robotic ships and the balance of global power. Copyright litigation was no longer just copyright litigation. It had become part of the national defense.

That framing changes the question. If OpenAI simply believes that its conduct qualifies as fair use, it can make that argument like any other defendant. If the United States believes access to privately financed creative work is essential to national security, then we are discussing a public need being satisfied with private property.

If the United States needs my property to win the AI race, then the United States can pay the invoice.

A Publisher's Disclosure

> "The web is publicly accessible. It is not ownerless."

I should disclose my position at the outset. I am not a disinterested observer.

I am the founder and CEO of Techstrong Group and editor-in-chief of our media properties. We pay journalists, editors, analysts, producers, designers and contractors to create original reporting, analysis, podcasts, videos and research. We also pay to publish and distribute it. None of that content appears by magic, and none of it is free for us to produce.

To date, Techstrong has never received a dollar from an AI company for using our content to train a model. I cannot say with certainty which companies used it, which articles they used or how often they used them.

That is not because the answer is no. It is because the companies generally do not provide enough information for a publisher like Techstrong to know.

OpenAI describes its training sources in broad categories: publicly available information on the internet, information obtained through partnerships and information provided or generated by users, human trainers and researchers. Its public explanation of model development[2] does not give an individual publisher a reliable way to determine whether a particular article, archive or website was included in a particular model's training data.

That opacity is not a side issue. It sits at the center of the entire copyright debate. A property owner cannot license a use it cannot see, challenge an unlawful copy it cannot identify or verify an assurance that its work was excluded. The AI company controls the dataset, the logs and the evidence. The copyright owner is expected to guess.

This is also different from the bargain that supported the commercial web for decades. Search engines crawled pages, displayed links and sent readers back to the source. Publishers accepted crawling because discovery and referral traffic had value. Generative AI systems can use a publisher's reporting to answer a reader's question without requiring the reader to visit the publisher at all.

The web is publicly accessible. It is not ownerless.

What the Government Is Asking the Court to Do

COPYRIGHT DOCTRINE MEETS THE INFRASTRUCTURE OF AI

THE LEGAL TEST

One dispute, three distinct questions

1
ACQUISITION
How was the work obtained?
Purchase, open-web crawl, paywall bypass, access-control circumvention or pirate library.

2
TRAINING
What computational use occurred?
Pattern analysis and model-weight adjustment; the central fair-use dispute.

3
OUTPUT
What can the system produce?
Protected expression, close substitutes or material that diverts demand from the original market.

The DOJ filing is a statement of interest under 28 U.S.C. Section 517[3], which allows the Justice Department to participate in litigation when the interests of the United States are implicated. It is not legislation. It is not a regulation. It is not a binding decision, and the judge is under no obligation to adopt it.

It is nevertheless a major intervention. According to Reuters[4], this is the first time the federal government has entered one of the major AI copyright cases to advance a position on training. The government's legal arguments could influence not only the claims against OpenAI, but also the emerging rules that will govern the entire American AI industry.

The filing asks the court to separate three questions that are too often treated as one.

The first is acquisition. How did an AI company obtain a copy of the copyrighted work? Did it purchase a book, crawl an open website, bypass a paywall, violate access controls or download the work from a pirate library?

The second is training. Once the company possessed a copy, did its computational use of that work to identify patterns and adjust model parameters qualify as fair use?

The third is output. Can the completed system reproduce protected expression, generate close substitutes or divert demand from the original market?

There is legal sense in separating those questions. A company could acquire a work illegally even if a later analytical use might otherwise be fair. A training process could be lawful while a particular output infringes. Copyright cases have always depended on purpose, amount, market effect and the specific conduct at issue.

The problem is what the government does after making that distinction. It urges the court to treat model training itself as spectacularly transformative. It argues that copying entire works may favor fair use because the works are not ordinarily presented to the public during training. It discounts theories that generative AI will flood markets with substitutes and reduce demand for human work unless plaintiffs can connect the harm to substantially similar outputs or direct substitution.

The filing argues that licensing could favor the richest technology companies and transfer money to established media organizations. It even warns against subsidizing "old mainstream media." But News Corp and the Financial Times can negotiate with OpenAI. A smaller publisher unable to audit a model or finance a federal lawsuit has much less leverage.

The national-security argument then supplies the weight behind all of it. AI will affect military systems, intelligence capabilities, science and economic competitiveness. China will not wait for American courts or publishers. Therefore, the government argues, an interpretation of copyright that constrains American model development could damage the national interest.

That may be a serious policy concern. It is not one of the four fair-use factors Congress wrote into the Copyright Act.

From an Industry Proposal to Federal Policy

THE POLICY PATH, 2025–2026

From industry proposal to federal policy

The sequence does not prove coordination. It shows how quickly the industry's strategic framing became the government's public position.

DATE	ACTOR	MILESTONE
JAN 2025	White House	Executive order directs removal of barriers to U.S. AI leadership.
MAR 2025	OpenAI	AI Action Plan proposal links broad training freedom to competition with China.
MAY 2025	Copyright Office	Training report rejects categorical answers and emphasizes case-specific fair use.
JUL 2025	White House	AI Action Plan prioritizes innovation, infrastructure and international leadership.
MAR 2026	White House	National framework states the administration's view that model training is lawful.
JUL 2026	Federal court	Final approval granted to Anthropic's $1.5 billion authors' settlement.
SEP 2026	Department of Justice	Statement of interest urges broad protection for model training.

The DOJ filing did not arrive without a policy history.

On January 23, 2025, President Trump issued an executive order[5] directing federal officials to remove barriers to American leadership in artificial intelligence. Less than two months later, OpenAI submitted its proposals for the U.S. AI Action Plan[6]. Among them was a call for a copyright strategy that would preserve the ability of American AI models to learn from copyrighted material. OpenAI explicitly connected that freedom to competition with China.

There is nothing improper about a company advocating for policies that benefit its business. Every significant industry in Washington presents its interests as aligned with the national interest. The government's responsibility is to decide where the industry's interest ends and the public interest begins.

In May 2025, the U.S. Copyright Office issued the prepublication version of its report on generative AI training[7]. The report did not accept the categorical position that all AI training is infringing, but it did not endorse the industry's categorical position either. It concluded that fair use depends on the works involved, the sources of the copies, the purpose of the model and the controls placed on outputs. It warned that using vast collections of copyrighted expression to produce commercial content that competes in the same markets could exceed the boundaries of fair use, particularly when the works were obtained illegally.

Register of Copyrights Shira Perlmutter was notified of her removal the following day. The administration's authority to remove her became the subject of litigation because the Copyright Office sits within the Library of Congress and advises Congress. The timing is documented. The motive remains disputed and should not be stated as proven. In September 2025, a federal appeals court blocked the administration from interfering with Perlmutter's work, and in June 2026 the Supreme Court declined to lift that protection while the litigation continued. Reuters covered the removal[8] and the Supreme Court action[9].

The administration continued building a national policy around rapid AI development. Its July 2025 AI Action Plan[10] emphasized innovation, infrastructure and international leadership. Its March 2026 National Policy Framework for Artificial Intelligence[11] went further. The document stated that the administration believed training AI models on copyrighted material did not violate copyright law. It acknowledged that copyright owners disagreed and said the courts should resolve the question. It also cautioned Congress against taking actions that would interfere with the judiciary's resolution of fair use.

The framework left room for Congress to facilitate licensing systems or collective-rights organizations, but argued that Congress should not determine when licenses are legally required. Creators would receive protection from infringing outputs, while the basic use of their works for training would be left to the courts under an administration-backed theory of broad fair use.

Then came the DOJ statement of interest. The progression is difficult to miss. The industry requested a broad freedom to train. The White House adopted the same strategic framing. The administration advised Congress not to legislate the core licensing question. DOJ then entered private litigation to encourage the courts to adopt the industry's preferred interpretation.

That chronology does not prove that OpenAI dictated the government's filing. It does establish that an industry's policy request matured into federal policy with remarkable efficiency.

The Law Congress Did Not Pass

Congress has already spoken about fair use. Section 107 of the Copyright Act[12] instructs courts to consider the purpose and character of a use, the nature of the copyrighted work, the amount used and the effect on the potential market. It also makes clear that the analysis is case-specific. No single factor automatically controls every dispute.

Congress has not enacted a blanket exemption for AI training. It has not created a national-security exception to copyright. It has not authorized commercial AI companies to take training material without payment. It has not added competition with China as a fifth fair-use factor.

The executive branch is entitled to tell a court how it believes federal law should be interpreted. That is not the same as rewriting the Copyright Act.

The institutional concern is more subtle. The administration has adopted the AI industry's legal theory as policy, discouraged Congress from deciding when licenses should be required and now asked the judiciary to establish a rule with nationwide economic consequences. If the courts accept that position, the country could acquire the functional equivalent of an AI-training exemption without Congress ever debating the price, conditions, transparency requirements or distributional effects of that exemption.

Members of Congress have proposed narrower measures. The bipartisan TRAIN Act[13] would give copyright owners an administrative subpoena process to determine whether their works were used to train a generative AI model when they have a good-faith basis for believing so. The CLEAR Act[14] would require model developers to file detailed notices about copyrighted training materials with the Copyright Office and would create a public database.

Neither bill has become law. Neither decides that every training use requires a license. Both recognize the obvious starting point: Copyright owners cannot exercise whatever rights they possess if they are not allowed to know whether their property was used.

Congress is also considering the Deterring American AI Model Theft Act[15]. That proposal treats unauthorized extraction of the capabilities of privately owned American models as a threat to intellectual property and national security. It calls for government cooperation with model owners to identify and deter attempts to use their systems to train or improve competing models while evading contractual, technical or geographic restrictions.

Put the bills next to one another and the contradiction almost writes itself. When authors want to discover whether an AI company used their books, Congress considers giving them a subpoena. When an AI company fears that a foreign competitor used its outputs to improve another model, Congress considers mobilizing the government to deter model theft.

Apparently, intellectual property becomes strategically important after it enters the model.

Private Property by Judicial Redefinition

Copyright is property, but it grants limited rights subject to exceptions such as fair use. A newspaper does not own the facts it reports, and an author does not own an idea simply because it appeared in a book. Copyright protects original expression.

That makes the property argument more complicated than accusing the government of seizing a printing press. If a court determines that a particular training use is fair, the owner did not possess an enforceable right to exclude that use. Without a recognized right, there is nothing for the government to take under the Fifth Amendment.

The administration's approach is therefore not a formal confiscation. It is an effort to persuade courts to draw the boundary of the property right in a place that gives AI developers broad freedom to copy. The legal label is fair use. The economic effect may still be the uncompensated transfer of value from creators to model companies.

Recent cases have not produced a universal rule.

In Bartz v. Anthropic, Judge William Alsup held that using lawfully acquired books to train Claude was "exceedingly transformative" and qualified as fair use. He compared the process to a reader learning from books rather than distributing substitutes for them. But he treated Anthropic's alleged downloading of millions of books from pirate libraries as a separate matter that could not be excused merely because the company later wanted to use those copies for training. The June 2025 ruling[16] gave Anthropic a major victory on training and a potentially enormous problem on acquisition.

In Kadrey v. Meta, Meta prevailed because the named plaintiffs failed to develop sufficient evidence that Meta's models would dilute the market for their books. Judge Vince Chhabria made clear that the result was tied to the record before him. His decision[17] warned that generative systems capable of producing large volumes of competing material could threaten authors' markets and incentives, and that a stronger evidentiary showing could lead to a different outcome.

In Thomson Reuters v. Ross Intelligence, a federal court rejected fair use where Ross used protected Westlaw headnotes to help build a competing legal-research service. The court found that the use was commercial, insufficiently transformative and directed at the same market. The February 2025 decision[18] did not involve a modern generative language model, but it is a reminder that courts do not regard every form of machine learning as legally interchangeable.

The New York Times litigation against OpenAI has not yet produced a final answer on the central training questions. A 2025 ruling allowed key direct and contributory infringement claims to proceed while dismissing some claims involving outputs that were not sufficiently similar to the Times articles. The litigation is moving toward decisions based on a much larger record.

The law is unsettled because the facts matter. Where did the copies come from? Can the model reproduce or substitute for the work? Is there a licensing market? Those are traditional fair-use questions.

"We need to beat China" is a policy argument. It should not become a judicial shortcut around all of them.

The $1.5 Billion Asterisk

THE $1.5 BILLION ASTERISK

What the Anthropic settlement decided — and what it didn't

WHAT IT DID	WHAT IT DID NOT DO
Resolved claims involving books allegedly obtained from pirate libraries and retained in a central research collection.	Establish that every use of a copyrighted book for AI training requires a license.
Avoided a damages trial and produced the largest known U.S. copyright settlement.	Reverse the court's finding that training on lawfully acquired books was fair use.
Created a recovery path for roughly 500,000 eligible titles.	Constitute an admission that every allegation in the authors' case was true.

The Anthropic litigation gives us the clearest test of the difference between training and acquisition.

Anthropic agreed to pay $1.5 billion to settle the authors' class action, and a federal judge granted final approval in July 2026. It is the largest known copyright settlement in the United States. More than 91 percent of authors and publishers covered by the settlement submitted claims, according to Anthropic, and roughly 500,000 eligible titles were included. Reuters reported on the final approval[19].

The settlement is sometimes described as Anthropic paying authors because it trained Claude on their books. That is incomplete. The court had already ruled that the training use of lawfully acquired books was fair. The remaining exposure involved copies allegedly obtained from pirate libraries and retained in a central research collection. Anthropic settled those claims rather than proceed to a damages trial. As with any settlement, it did not constitute an admission that every allegation was true.

The resulting rule sounds sensible: Training may be fair use, but stealing the copies is not.

It becomes considerably less reassuring when applied to a small publisher. How would Techstrong know whether an AI company crawled an openly accessible page, copied material through a licensed database, circumvented a restriction or obtained an archive from somewhere else? How could we distinguish a permitted copy from a pirated one when the company holding all of the evidence will not disclose its dataset?

The largest authors' groups and publishers can retain sophisticated counsel, compel discovery and sustain years of federal litigation. Most creators cannot. The nominal rule may be the same for everyone, but the ability to enforce it is radically unequal.

There may not be one set of rules for Anthropic and another for OpenAI. There may be one set of enforceable rights for copyright owners who can afford to sue and another for everyone else.

Name That Tune

Books are not Anthropic's only copyright problem.

Universal Music Publishing Group, Concord Music Group and ABKCO have pursued Anthropic over song lyrics since 2023. In January 2026, those publishers brought a much larger action involving more than 20,000 musical works and potential statutory damages exceeding $3 billion. The publishers allege that Anthropic copied and retained lyrics, sheet music and compositions without permission. Reuters described the new case[20] as potentially one of the largest non-class-action copyright cases filed in the United States.

Sony Music Publishing and Warner Chappell filed another federal action in August 2026. Their complaint alleges that Anthropic obtained protected lyrics and sheet music through torrents, scraping, pirate sites and other unauthorized sources. It also alleges that Claude can reproduce lyrics verbatim and generate purportedly new lyrics that compete with the publishers' works. The defendants include Anthropic as well as CEO Dario Amodei and co-founder Benjamin Mann for some of the torrent-related claims. Anthropic disputes the allegations and says it will defend its training practices as fair use. Reuters reported the filing[21], and the complaint docket is available through CourtListener[22].

These are pending allegations, not judicial findings. They nevertheless put all three stages of the AI copyright problem into a single case. The publishers challenge how Anthropic acquired the works, what it did with them during training and what Claude can produce after training.

That matters because the DOJ position is built around compartmentalization. Acquisition can be litigated separately. Training should generally be treated as transformative. Outputs can be evaluated one at a time

for substantial similarity. Each proposition has legal logic when viewed alone. Together they can create an obstacle course that only the largest rights holders can navigate.

A publisher first has to discover that its work was used. It then has to identify the source of the copy. It may have to show that an access restriction was circumvented or that a pirate dataset was used. It must separately address the training defense. If it challenges outputs, it may have to document substantially similar passages or direct market substitution. At every stage, most of the technical evidence resides with the defendant.

If the government believes illegal acquisition remains actionable, then it should support enough transparency to make that right enforceable. Otherwise, the acquisition distinction is a locked courthouse door with a sign telling creators that they are welcome to enter if they can find the key.

Public When It Needs an Exemption, Private When the Bill Arrives

A NATIONAL PROJECT BUILT ON PRIVATELY FINANCED CREATIVE WORK

> "OpenAI is public when it needs an exemption and private when the bill arrives."

The national-security framing invites a question that ordinary fair-use analysis does not answer. When the government needs private property for a public purpose, why should the owner alone bear the cost?

The Takings Clause of the Fifth Amendment prevents the government from taking private property for public use without just compensation. The Supreme Court has described the basic principle as preventing the government from forcing a small number of people to carry public burdens that should be borne by the public as a whole. Copyright has its own government-use provision. Under 28 U.S.C. Section 1498(b)[23], when the United States infringes a copyright, or when an authorized contractor acts for the government, the owner can seek reasonable and entire compensation in the Court of Federal Claims.

That does not mean publishers currently possess a winning takings claim against OpenAI or the government. OpenAI is a private company. DOJ has not ordered it to use any particular work. If training is fair use, there is no infringement to compensate. Eminent domain is therefore better understood here as an analogy and a possible legislative model, not a description of the present case.

DOJ nevertheless goes out of its way to protect the government from the compensation question. In footnote two of its statement, the United States expressly says it is not arguing that OpenAI's challenged activities were authorized by or performed for the government under Section 1498.

Read the filing as a whole and the position is remarkable. OpenAI's access to copyrighted material serves scientific progress, economic prosperity and national defense. An adverse copyright ruling could weaken American weapons, intelligence and global competitiveness. But OpenAI is still acting privately when the question turns to who should pay.

OpenAI is public when it needs an exemption and private when the bill arrives.

The legal response is that fair use is not an exemption invented for OpenAI and that public benefits have always been relevant to the first fair-use factor. Fair enough. But the more heavily the administration relies on national necessity, the less convincing it becomes to place the entire cost on individual copyright owners.

National security has justified government procurement, compulsory access, requisition and eminent domain throughout American history. Those mechanisms do not always give property owners everything they want, but they at least recognize that a public project is being built with something of value. The property does not cease to exist because the government's purpose is important.

If AI training is truly a national project, Congress can authorize defined uses and provide compensation. It can establish a collective license, create a fund or use a government-authorization framework for specific national-security models. What it should not do is obtain the public benefit of nationalization while leaving all costs and risks with the private owners.

Too strategic to regulate. Too private to compensate.

The Copyright Office Warning

The Copyright Office report offers a more credible starting point because it refuses the easy answers offered by both sides.

It does not say that every act of model training requires permission. Research, analysis and uses that do not produce substitutes may qualify as fair use. Copyright does not protect facts, general concepts or styles. Models can learn relationships and patterns without storing a human-readable copy of every source in the way a conventional library does.

The report also refuses to pretend that creative works are merely raw data. They contain protected expression created through human effort and financial investment. Copying them at different stages of model development can implicate exclusive rights. A commercial system trained on enormous quantities of copyrighted work and designed to produce competing expressive content may present a very different fair-use case from a research system that detects patterns without serving the original market.

The Office found that voluntary licensing was already developing and recommended allowing those markets additional time to mature. It also identified collective licensing as a possible response where individual transactions are impractical. A collective system could aggregate rights across thousands of smaller publishers and creators, give model developers a scalable path to permission and distribute revenue beyond the handful of companies powerful enough to negotiate direct deals.

The report was especially skeptical of making opt-out the default. Copyright normally begins with the owner's right to authorize certain uses. A technical instruction in robots.txt is a crude substitute. A publisher may want Google to index an article and readers to share it while declining to provide the same article for commercial model training. Those are different uses and should not require withdrawing from the open web entirely.

DOJ's filing attacks the Copyright Office analysis and argues that it deserves no deference. It even notes that Perlmutter is challenging her removal. That is legally relevant background, but it also highlights the institutional conflict. The expert office Congress created to advise it produced a nuanced report inconvenient to the administration's policy. DOJ is now urging a court to reject the report in favor of a much broader industry-friendly rule.

The Publisher Pays Twice

THE PUBLISHER FUNDS THE INPUT WHILE THE MODEL CAPTURES THE ANSWER

The economic bargain supporting digital publishing was already under severe pressure. Advertising moved to platforms, social networks reduced referrals and search algorithms changed without notice. Publishers adapted because adaptation is part of the business.

Generative AI alters the bargain at a more fundamental level. A model can use reporting to formulate an answer while eliminating the reader's reason to visit the source. The publisher pays to discover, verify, write and edit the information. The AI system uses that information to serve its own user. The publisher may receive neither a licensing payment nor a click.

A July 2025 Pew Research Center study[24] found that users clicked a conventional search result in 8 percent of visits when an AI summary appeared, compared with 15 percent when no AI summary appeared. Only 1 percent clicked a source link contained in the AI summary. Users were also more likely to end the browsing session after receiving the summary.

One study does not establish the effect on every publisher. It confirms the incentive: The better the AI answer, the less reason a user has to consult the reporting that made it possible.

Cloudflare has described the collapse of the old crawl-for-referral exchange in even sharper terms. The company operates a vast network and sees both crawler requests and referrals across its customers. It has introduced Pay Per Crawl[25] and AI Crawl Control[26] to let publishers block AI crawlers, request payment or express different preferences for different uses. Cloudflare's figures are its own network data and should be read accordingly, but its product strategy reflects a market need: Publishers want more than the binary choice between surrendering their content and disappearing from the web.

AI companies respond that models do not contain a conventional database of copied articles. Training changes numerical weights by learning statistical relationships. That technical distinction matters legally, but it does not settle the economics. A refinery transforms crude oil. That does not make the oil free.

We also know that publisher content has market value because AI companies already pay for some of it. OpenAI has signed agreements with News Corp[27], the Financial Times[28], Axel Springer, The Associated Press, Hearst, Condé Nast, Vox Media, The Atlantic, Time and others. Its publisher-partner list[29] is evidence that high-quality, current content improves AI products and is worth licensing.

Those agreements are preferable to taking the content without permission, but they expose a bargaining divide. Large publishers bring recognizable brands, extensive archives and enough legal leverage to make a negotiation worthwhile. An independent technology publisher cannot sit across the table from a company valued in the hundreds of billions of dollars as an equal party.

DOJ argues that mandatory licensing could entrench the largest AI companies because only they could afford the bills. That is a legitimate concern. But declaring training categorically free does not create a competitive market. It transfers the value to every model developer while denying compensation to every publisher that lacks the leverage to negotiate a voluntary deal.

The alternative is not billions of individual negotiations. Collective licensing exists because some markets contain too many works and owners for direct licensing. It provides broad access while distributing compensation. The Copyright Clearance Center has already introduced an AI systems training license[30], although no current system resolves the entire market.

The government is presenting a false choice between zero-cost training and a licensing maze that stops American AI. A functioning policy can protect innovation while recognizing that the people who produce the raw material are part of the same American economy.

Train as We Say, Not as We Do

THE ONE-WAY PROPERTY REGIME

The vocabulary changes with the owner

PUBLISHER INPUTS	AI MODEL OUTPUTS
Human-created work is described as training material.	Model output is described as proprietary capability.
Mass copying is characterized as learning.	Mass extraction is characterized as theft or free-riding.
Transparency is resisted as burdensome or confidential.	Access controls, contracts and account monitoring are actively enforced.
National security supports broad access.	National security supports restricting foreign access.

The industry's reaction to Chinese model distillation reveals how quickly its vocabulary changes when the intellectual investment belongs to an AI company.

Distillation can be a legitimate method for teaching a smaller model to reproduce some capabilities of a larger one. It can also involve a competitor sending large numbers of queries to a closed model and using the responses to train a rival system. OpenAI and Anthropic argue that several Chinese companies crossed that line through deceptive accounts, evasion of geographic restrictions and large-scale extraction.

Anthropic reported in February 2026 that DeepSeek, Moonshot AI and MiniMax used approximately 24,000 fraudulent accounts to generate more than 16 million exchanges with Claude as part of alleged distillation attacks[31]. Anthropic said the campaigns allowed competitors to acquire capabilities at a fraction of the time and expense required to develop them independently.

OpenAI similarly accused DeepSeek of using OpenAI model outputs to improve competing systems and warned lawmakers about foreign companies attempting to free-ride on American research. OpenAI's own terms of use[32] prohibit automatically extracting output and using it to develop models that compete with OpenAI.

The legal cases are not identical. Distillation may violate contracts and terms of service. It may involve fake accounts, circumvention, trade secrets or export controls. Copyright in model outputs can be uncertain, particularly when outputs lack sufficient human authorship. Those differences should be stated clearly.

The economic principle is harder to distinguish. An organization spends billions of dollars creating something valuable. Another organization uses the resulting material without permission to accelerate a competing product. The original investor describes the practice as free-riding that undermines innovation.

That description will sound familiar to every author, musician and publisher involved in the copyright cases.

When a model developer copies millions of human-created works, the company calls it learning. When another company learns from millions of model outputs, the company calls it theft. When unrestricted access benefits

American AI companies, it advances national security. When the same logic benefits Chinese competitors, it threatens national security.

You cannot have it both ways. At a minimum, the industry needs a consistent theory explaining why consent, compensation and competitive substitution matter on one side of the model but not the other.

Money, Access and AI Exceptionalism

No examination of federal AI policy would be complete without following the political money. That does not mean treating every campaign contribution as a bribe. It means recognizing that access and influence are distributed through a political system in which money matters.

Sam Altman announced a personal $1 million contribution to President Trump's inaugural fund and said Trump would lead the country into the age of artificial intelligence. Reuters reported the contribution[33] alongside donations from several other technology leaders and companies.

OpenAI co-founder Greg Brockman and his wife, Anna, subsequently contributed a combined $25 million to the Trump-aligned MAGA Inc. super PAC, according to federal filings reported by Reuters[34]. They also committed substantial money to Leading the Future, a pro-AI political network supporting candidates considered friendly to the industry. OpenAI has said these are personal donations rather than corporate contributions.

The major AI companies have also expanded their federal lobbying. Their filings identify copyright, regulation, infrastructure and other AI policies among the issues on which they seek to influence the government. Again, that is legal and unsurprising. A strategically important industry facing existential policy decisions would be foolish not to make its case in Washington.

The concern arises from the combination of money, access and doctrine. The companies tell the administration that their commercial success is inseparable from American national security. The administration embraces that framing. Policy obstacles are recast as threats to national competitiveness. Private companies become national champions without becoming public utilities, government contractors or entities with public compensation obligations.

We do not have evidence that a particular contribution purchased the DOJ statement. We should not imply that we do. We do have enough evidence to ask whether a politically favored industry is receiving an extraordinary degree of deference at the expense of a less concentrated group of authors, musicians, journalists and independent publishers.

The appearance problem is compounded by DOJ's choice of language. A government that worries licensing will subsidize "old mainstream media" does not sound like a neutral defender of copyright's constitutional balance. It sounds like an administration bringing its political resentments into a dispute over property rights.

The Strongest Case for the AI Companies

> "Difficulty negotiating licenses does not establish that the price should be zero."

The AI companies are not wrong about everything. Their practical problem is real.

Frontier models learn from immense and varied collections. Identifying every rights holder and negotiating an individual license for every work could be impossible. Transaction costs might exceed the value of many uses. A licensing requirement that only OpenAI, Google, Microsoft, Meta, Amazon and Anthropic can afford could lock smaller developers and open-source projects out of the market. Copyright must not be extended to facts, ideas, methods or styles merely because a computer learned them from protected expression.

There is also a real national-security competition. Advanced AI will affect military planning, intelligence, cybersecurity, biology, energy, manufacturing and economic productivity. China will not design its policy around the comfort of American copyright owners. Rules that prevent American models from learning from large bodies of knowledge while foreign rivals proceed without restraint could weaken the United States.

Courts have repeatedly had to apply old statutes to new technologies. Search indexing, reverse engineering and digitization all generated copyright battles. Requiring permission for every computational encounter with a work could create rights Congress never intended.

Those are substantial arguments. They do not require us to accept every demand made in their name.

Difficulty negotiating licenses does not establish that the price should be zero. Competition with China does not explain why training sources must remain secret. Transformative use does not excuse piracy. Protecting ideas and facts does not require eliminating the market analysis for protected expression. The risk of entrenching large AI companies does not justify strengthening them with free access to assets financed by smaller businesses.

The debate is routinely framed as a choice between innovation and copyright. It is not. Copyright itself is an innovation policy. The Constitution authorizes exclusive rights to promote progress by giving creators an incentive to create. AI can produce extraordinary public benefits, but models will not have much to learn from if the businesses producing reliable human knowledge cannot survive.

A Compensation System Worth Debating

POLICY DESIGN

A compensation system worth debating

OPTION	STRENGTH	LIMIT	DISTRIBUTIONAL EFFECT
Direct licensing	High precision	Poor at internet scale	Favors large rights holders and large model developers
Collective licensing	Broad category access	Requires representative institutions	Can include smaller publishers and creators
Compensation fund	Scalable and predictable	Less precise allocation	Spreads cost across commercial developers
Transparency rules	Makes rights enforceable	Does not itself set a price	Foundational to every other mechanism
Section 1498-style use	Fits defined government needs	Limited to authorized public purpose	Places national-defense costs on the nation

The choice is not total prohibition or unlimited fair use.

Congress could begin with transparency. Model developers could disclose sufficiently detailed summaries of training sources, with confidential audit procedures for information that genuinely constitutes a trade secret. The European Union already requires providers of general-purpose AI models to publish summaries of training content under Article 53 of the AI Act, using a mandatory Commission template[35]. Transparency does not determine whether a use is fair. It gives copyright owners enough information to participate in the determination.

Congress could encourage collective licensing. A model developer would not need to negotiate article by article or song by song. It could license defined categories through an organization representing many owners. Extended collective licensing could cover an entire class while preserving appropriate rights to opt out. The details would be difficult, particularly when ownership records are incomplete, but difficult is not the same as impossible.

A compensation fund could spread costs across commercial developers according to revenue, compute or deployment. It would be less precise than direct licensing but more honest than declaring every input free.

Lawmakers could distinguish nonprofit research from commercial development. They could condition safe harbors on lawful acquisition, dataset records, reasonable output protections and cooperation with audits. They could impose stronger remedies for torrenting, circumventing paywalls and removing copyright-management information.

For models genuinely developed at the direction of the government for national-security purposes, Congress could use a Section 1498-style process. The government would obtain the access it considers necessary, and the owners would receive a defined route to compensation. That would put the cost of national defense where it belongs: on the nation, not on whichever authors and publishers happened to supply the training material.

No mechanism will fit books, software, journalism, music, photography and video equally well. Congress should hear from developers, researchers, creators, publishers, libraries and the public before drawing the lines.

What Congress should not do is leave the entire bargain to private lawsuits in which a few well-funded plaintiffs try to reconstruct secret datasets after the fact. Nor should it allow the executive branch to turn one industry's desired interpretation into national policy while telling lawmakers to stay out of the central licensing question.

Send Washington the Invoice

> "National security can justify taking or licensing private property. It should not make that property disappear."

I do not want to stop AI development. Techstrong covers artificial intelligence every day. We use AI tools. We understand their value, their momentum and their potential. The success of American AI matters to our readers, our industry and our country.

That does not require pretending its raw materials came from nowhere.

The articles, books, songs, photographs and software used to build these systems exist because someone invested time, judgment, talent and money in creating them. AI companies rightly demand recognition for the billions they invest in models, chips, data centers, researchers and energy. They become much less interested in investment and ownership when the asset belongs to someone upstream.

The United States may conclude that access to copyrighted work is necessary for American AI leadership. It may decide that billions of individual licenses are impractical. It may decide that certain training uses are so transformative and socially valuable that they deserve broad protection.

Those are consequential public choices. Congress should make them openly. The rules should identify what uses are permitted, require enough transparency to enforce the remaining rights and provide a workable compensation system when private creative property is being used to advance a national project.

The same principle should apply on both sides of the model. If consent, investment and competitive appropriation matter when DeepSeek learns from OpenAI or Anthropic, they cannot become irrelevant when OpenAI or Anthropic learns from everyone else. If American intellectual property deserves government protection, that category must include the people who wrote the material on which American AI was built.

National security can justify taking or licensing private property. It should not make that property disappear.

If the United States needs privately financed creative work to win the AI race, then the United States and the companies selected to win it should be prepared to pay the invoice.

Sources and linked documents

Source markers in the report are clickable. The complete linked-document index follows.

1. CourtListener. https://storage.courtlistener.com/recap/gov.uscourts.nysd.641355/gov.uscourts.nysd.641355.316.0.pdf
2. OpenAI Help Center. https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed
3. U.S. House Office of the Law Revision Counsel. https://uscode.house.gov/view.xhtml?edition=prelim&num=0&req=granuleid%3AUSC-prelim-title28-section517
4. Reuters. https://www.reuters.com/legal/litigation/us-government-backs-openai-new-york-times-copyright-case-2026-09-02/
5. The White House. https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/
6. OpenAI. https://openai.com/global-affairs/openai-proposals-for-the-us-ai-action-plan/
7. U.S. Copyright Office. https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf
8. Reuters. https://www.reuters.com/legal/government/trump-fires-head-us-copyright-office-2025-05-12/
9. Reuters. https://www.reuters.com/legal/government/us-supreme-court-wont-let-trump-remove-top-copyright-official-now-2026-06-30/
10. The White House. https://www.whitehouse.gov/releases/2025/07/white-house-unveils-americas-ai-action-plan/
11. The White House. https://www.whitehouse.gov/wp-content/uploads/2026/03/03.20.26-National-Policy-Framework-for-Artificial-Intelligence-Legislative-Recommendations.pdf
12. U.S. Copyright Office. https://www.copyright.gov/title17/92chap1.html#107
13. GovInfo. https://www.govinfo.gov/content/pkg/BILLS-119hr7209ih/xhtml/BILLS-119hr7209ih.html
14. GovInfo. https://www.govinfo.gov/content/pkg/BILLS-119s3813is/html/BILLS-119s3813is.htm
15. GovInfo. https://www.govinfo.gov/content/pkg/BILLS-119hr8283ih/html/BILLS-119hr8283ih.htm
16. Copyright Alliance. https://copyrightalliance.org/wp-content/uploads/2025/06/Bartz-v.-Anthropic-Order.pdf
17. Justia. https://law.justia.com/cases/federal/district-courts/california/candce/3%3A2023cv03417/415175/598/
18. Reuters. https://www.reuters.com/legal/thomson-reuters-wins-ai-copyright-fair-use-ruling-against-one-time-competitor-2025-02-11/
19. Reuters. https://www.reuters.com/world/us-judge-approves-anthropics-15-billion-settlement-copyright-lawsuit-2026-07-20/
20. Reuters. https://www.reuters.com/legal/litigation/anthropic-faces-new-music-publisher-lawsuit-over-alleged-piracy-2026-01-28/
21. Reuters. https://www.reuters.com/legal/government/sony-warner-music-sue-anthropic-over-songs-used-ai-training-2026-08-31/
22. CourtListener. https://www.courtlistener.com/docket/74720572/sony-music-publishing-us-llc-v-anthropic-pbc/
23. GovInfo. https://www.govinfo.gov/link/uscode/28/1498
24. Pew Research Center. https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/
25. Cloudflare. https://blog.cloudflare.com/introducing-pay-per-crawl/
26. Cloudflare. https://blog.cloudflare.com/introducing-ai-crawl-control/
27. Reuters. https://www.reuters.com/technology/sam-altmans-openai-signs-content-agreement-with-news-corp-2024-05-22/
28. Reuters. https://www.reuters.com/technology/financial-times-openai-sign-content-licensing-partnership-2024-04-29/
29. OpenAI. https://openai.com/index/introducing-chatgpt-search/
30. Copyright Clearance Center. https://www.copyright.com/media-press-releases/ccc-announces-ai-systems-training-license-for-the-external-use-of-copyrighted-works-coming-soon/
31. Anthropic. https://www.anthropic.com/news/detecting-and-preventing-distillation-attacks
32. OpenAI. https://openai.com/policies/row-terms-of-use/
33. Reuters. https://www.reuters.com/world/us/corporate-america-pledges-donations-trump-inauguration-2024-12-17/
34. Reuters. https://www.reuters.com/world/us/trump-aligned-maga-inc-super-pac-enters-2026-with-300-million-stockpile-2026-01-03/
35. European Commission. https://digital-strategy.ec.europa.eu/en/library/explanatory-notice-and-template-public-summary-training-content-general-purpose-ai-models

About the author

Alan Shimel is the founder and CEO of Techstrong Group and editor-in-chief of its media properties.