Senate Judiciary Subcommittee on Crime and Counterterrorism
Too Big to Prosecute?: Examining the AI Industry's Mass Ingestion of Copyrighted
Works for AI
July 16, 2025
Questions for the Record
Senator Amy Klobuchar

<u>For Mr. Baldacci:</u>

According to the Pew Research Center, newspaper advertising revenue has plummeted from $37 billion to $9 billion in recent years. Much of this decline is the result of online platforms finding ways to capture that ad revenue.

- In your written testimony, you note that the median income for authors has experienced a similarly dramatic drop, falling 42% in a decade. How does generative AI's use of author content further threaten the livelihood of authors?

The unlicensed use of books and journalism to train generative AI is truly an existential threat to the future of the already precarious writing profession; and that is because it attacks our potential earnings and the incentives to write in several ways at once: 1) it allows users, with just a little prompting, to quickly generate copycat books, books in an author's style and other infringing works that serve as market substitutes; 2) floods the market with cheaply priced AI-generated books, devaluing all books; 3) robs authors of licensing income both for the training and downstream uses; and 4) promotes piracy.

With most authors still striving to make up the losses of the last decade plus—authors' mean writing income in 2022 was only $20,000 (and only half of that from books) according to the Authors Guild's most recent author earnings survey—it does not take much more income loss before many can no longer afford to write. That in turn will impact the breadth of books that are written and published and prevent many great stories, histories, and ideas from seeing the light of day. It also will impact publishers' ability to invest in a broad assortment of new books, forcing them to focus on proven bestsellers, such as celebrity books and books by already established brand name writers, making it even harder for talented, up-and-coming authors writers to be discovered and earn any meaningful income. Those would-be authors will have less incentive to educate and train themselves and become writers, which will impoverish our nation's literature—directly upending the goals that copyright has existed to advance since the framing of the Constitution. What we're witnessing is theft of authors' works by some of the largest companies in the world to develop technologies that will grow their already enormous

profits. Unless guardrails are placed and these are technologies regulated, it will further accelerate the ongoing transfer of wealth from middle-class creators to the tech industry, and to the detriment of us all.

1. Outputs that Serve as Substitutes for our Books

Once trained on a book, an AI model can produce outputs that directly infringe a work, like summaries, excerpts, and derivative works such as video and audio versions of their work, as well as outputs that emulate a particular author's body of work. While AI companies have technologies that allow them to prevent certain types of outputs, they will not do so unless there are legal consequences. I understand that none of the major LLMs today use technologies to prevent requested outputs "in the style of" an author or that emulate an author. As I said in my testimony, ChatGPT was able to write a plot summary of a "David Baldacci novel" that included "my plot lines and twists, character names, narratives, and every other element from my work, as well as my writing style ripped from my copyrightable expression." If "fair use" is allowed to trounce authors' copyrights, AI companies will have no incentive to prevent output uses that unfairly compete with human authored books. This is why licensing is so important. It not only protects copyright, but gives authors and publishers important controls on downstream and output uses of the work.

As also I noted in my testimony, it is now common to see an AI-generated "summary" or other knockoff of an anticipated bestseller appear on Amazon on the day of or even before the release of the real book. These are clearly intended to confuse consumers and unfairly capitalize on the success of human authors by diverting sales away from the books into which they have invested years of their lives.

This means that when trained on authors' works without permission, not only do authors not get paid for that use of their work, but they also lose sales of their books to AI-generated books that infringe or otherwise serve as direct substitutes.

2. Flooding the Market

The inevitable flood of AI-generated copycat books will dilute the market for all human-authored books, and I understand that we are already seeing evidence of this dilution in the markets for certain genres, like romance and other genre fiction. That means more and more aspiring writers competing for fewer and fewer writing jobs.[1] As a result, we will have far fewer talented people who can devote themselves to writing as a profession

---

[1] *See* Pranshu Verma and Gerritt de Vynck, ChatGPT took their jobs. Now they walk dogs and fix air conditioners, Washington Post (June 2, 2023), https://www.washingtonpost.com/technology/2023/06/02/ai-taking-jobs/.

and, inevitably, fewer great books will be the unfortunate result. This will irrevocably diminish our culture and society. After all, as the Supreme Court has long recognized, copyright exists not just to "secure a fair return for an author's creative labor," but ultimately "to stimulate artistic creativity for the general public good."[2]

As a federal court put it just a few weeks ago:

> Generative AI has the potential to flood the market with endless amounts of images, songs, articles, books, and more. People can prompt generative AI models to produce these outputs using a tiny fraction of the time and creativity that would otherwise be required. So by training generative AI models with copyrighted works, companies are creating something that often will dramatically undermine the market for those works, and thus dramatically undermine the incentive for human beings to create things the old-fashioned way.[3]

3. Loss of Licensing Income

When AI companies simply help themselves to books for AI training without permission or compensation, it deprives authors of a valuable and rapidly growing licensing market. In the recent past, when new technologies come into existence, such as e-books, it has led to a decline in authors' book earnings, making it imperative to find other sources of income. Licensing markets are a critically important piece of this. A well-functioning copyright system would allow authors to license their books to AI companies for training and thereby obtain a valuable income stream to mitigate losses from market dilution and copycat books.

Moreover, it is only fair that authors should share in the economic rewards of generative AI, since those systems, which have generated billions for tech companies, depend on our works. As the evidence in the current class action lawsuits against AI companies for the unlicensed use of books, the large AI companies all risked pirating millions of books to train their AI precisely because they understood just how essential books are to the quality of their LLMs. In other words, they are directly profiting off our talents without paying us anything. This result is in direct contravention of U.S. copyright laws, both their letter and spirit.

---

[2] *Twentieth Cent. Music Corp. v. Aiken*, 422 U.S. 151, 155 (1975) (internal quotation marks omitted).
[3] Order Denying the Plaintiffs' Motion for Partial Summary Judgment and Granting Meta's Cross Motion for Partial Summary Judgment, at 1-2, *Kadrey v. Meta*, No. 23-cv-03417, doc. no. 598 (N.D. Cal. June 25, 2025).

Just as these companies are expected to pay for the electricity and infrastructure needed to run their data centers, so too should they have to pay for the vital creative works that are the *sine qua non* of their products.

Indeed, without our books they cannot build their AI platforms. Simply because books have become valuable to the AI community in a way that was not foreseen previously does not give them the right to steal our work. Oil was not that valuable until the invention of the combustion engine. When that happened, no one thought it was okay for the oil companies to steal the very product they needed to build their businesses. The situation is not entirely analogous because, unlike oil, creative works are particularly vulnerable to cheap copying and piracy. That is why the founders, in Article I of the Constitution, gave additional protections to copyright holders.

Yet if AI companies can simply take all the books they need for free, these essential, already-developing licensing markets will evaporate, leaving authors in the grossly unfair position of losing work to AI systems that their own books were used to create, without a penny of compensation.

4. Incentives to Pirate Books

Last, allowing AI companies to continue copying books *en masse* from pirate websites will only incentivize more piracy—already a significant drain on authors' livelihoods. Pirate sites have a ready user base in the richest companies in the world with an insatiable demand for books. It doesn't take an economist to know that, if these companies are free to get their books from such sites, new pirates will emerge and find ways to get their own piece of the action.[4] As Professor Smith explained in his testimony, we know that digital piracy significantly harms authors' incomes in multiple ways.[5] Incentivizing piracy will only make the problem worse, and vindicate patently illegal uses of copyrighted works

****

Together, the harms caused by AI companies' uncontrolled and uncompensated use of our books will add up to enormous losses, devaluing human writing; and will directly

---

[4] *See* Written Testimony of Dr. Michael D. Smith, Carnegie Mellon University at 4 (July 16, 2025) ("Indeed, this the unlicensed use of pirated content could create a new illicit licensing business model for pirate networks: adding new stolen content to their collections, knowing that AI developers will want access to them."), https://www.judiciary.senate.gov/imo/media/doc/64bc45b6-9e04-22e4-34c1-12d0efad69ef/2025-07-16%20-%20Testimony%20-%20Smith.pdf.
[5] *See id.*

impact the incentives to do the hard work it takes to become a good writer and to be able to keep writing.

These harms will also have ripple effects throughout other sectors of the writing profession. The vast majority of authors today rely on other types of writing—freelance journalism, advertising, website content, etc. to make up for the losses in book income. Fewer book publishing deals will mean that more authors will need even more supplemental income while working on their personal writing projects. But AI is already replacing many of those jobs as well, resulting in writers being attacked on multiple flanks. And that means that even our most talented and highly trained writers will have to leave the writing profession altogether.

This is clearly not what the country's founders intended. The AI community is pirating our work, breaking the laws of copyright and now asks to be excused from these abuses, only after their crimes were discovered, by arguing their technology is so transformational that the very laws intended to stop such pillaging should not apply.

One of the first lessons I learned in law school was that the slippery slope is indeed slippery. Thus, if the AI community succeeds in their argument, there is no more copyright protection for anyone. The exception will have swallowed the rule, as everyone in the future will argue that their theft of our work is also transformational. And while Professor Lee at the July 16 hearing argued that we should let the court cases play out, the reality is that the vast majority of authors, or even groups of authors, do not have the time or financial wherewithal to take on the largest corporations in the world and their armies of lawyers in a protracted court battle. Being a member of a class action lawsuit currently, I can attest to how disruptive, costly and time consuming such litigation is. That is why legislation is desperately needed, to stop this juggernaut of copyright thievery before the damage becomes irreversible. Unfortunately, we are already perilously close to that point of no return.