# AI Copyright Neg

### Notes

This is one of the files that you should use if you are negative against the copyright affirmative! There are 4 big sections of this file: the science research disadvantage, the opt-out counterplan, answers to the aff’s journalism advantage, and answers to the aff’s model collapse advantage.

If you are negative against the copyright affirmative, you might want to have other files open too! For instance, you might want to read the ai industry disadvantage, the inflation disadvantage, the court clog disadvantage, or the capitalism k, all of which are in separate files.

#### Journalism advantage answers

These are direct responses to the case that the affirmative presented in the 1ac. Included in these responses are arguments that challenge the aff’s ability to solve for the decline of journalism (arguments 1-3: licensing is impossible, the absolution turn, and tech consolidation) as well as arguments minimizing the need for solving for the decline of journalism (arguments 4 and 5: journalism is resilient and democracy doesn’t solve war).

In the 1nc, you should read from the header “1nc journalism adv frontline.” In the negative block, each of those 5 arguments has extension evidence. You can choose to extend any or all of the arguments from the 1nc and read the new extension evidence. You could choose to extend all 5 arguments, or you could choose to extend a smaller number of them.

#### Model collapse advantage answers

These are direct responses to the case that the affirmative presented in the 1ac. Included in these responses are arguments that challenge the aff’s ability to solve for model collapse (arguments 2 and 3: aff can’t solve in time and aff causes worse ai models) as well as arguments minimizing the need for solving for model collapse (arguments 1 and 4: no model collapse and data poisoning doesn’t cause war).

In the 1nc, you should read from the header “1nc journalism adv frontline.” In the negative block, each of those 4 arguments has extension evidence. You can choose to extend any or all of the arguments from the 1nc and read the new extension evidence. You could choose to extend all 4 arguments, or you could choose to extend a smaller number of them.

#### Science research disadvantage

This is a new offcase position that you can introduce in the 1nc. It argues that, by increasing copyright protection, the aff also causes an unintended consequence of preventing works from being used for scientific research. In particular, it would limit TDM (text and data mining) research, which is relevant to a bunch of different scientific fields. Because the aff makes clear that AI trainers need to pay for copyrighted works, scientists would also have to pay for those same works since they are doing the same thing (scraping the internet for different texts and using them to study new uses). By undermining scientific research, the aff hurts our ability to use science to adapt to future global challenges.

In the 1nc, you just need to read the header called “1nc science research da.” In the negative block, you can read any of the 2nc/1nr cards that respond to what the 2ac said.

#### Opt-out counterplan

This is a different way of trying to solve for the worst abuses of AI competing with journalism. However, it differs from the affirmative in a key way. The affirmative makes it so that AI companies have to start from the presumption that they **are not allowed to use works to train ai.** If they want to use different works, they can **only once they have received permission.**

The counterplan approaches this with the opposite presumption: **AI companies are allowed to use works to train ai**. If creators do not want their work used for training AI, they have to tell AI companies that and prove that their own work is being harmed. They can prove that in two ways: 1. They can prove the AI would make an exact replica of their work or 2. They can prove that the AI makes something that is “substantially similar”.

Once the creator provides notice, the AI company will have to take some sort of action to respond (this is where the “notice and action” part comes from). The actions the AI company can take are to: 1. remove the author’s work from the datasets used by its AI system; 2. embed filtering technology into the AI system to prevent generation of similar content that would infringe the author’s copyright again; and/or 3. initiate a “machine unlearning” process where the AI forgets the things it learned and is trained differently

In the 1nc, you just need to read the header called “1nc opt-out cp.” In the negative block, you can read any of the 2nc/1nr cards that respond to what the 2ac said.

If you read this counterplan, you need to read a “net benefit” with it. For this counterplan, that net benefit should either be the inflation da or the court clog da, both of which are in separate files.

## answers to journalism advantage

### 1nc journalism advantage frontline

#### On the journalism advantage:

#### 1. Licensing is impossible for gen ai training data—the scale is too large

Damle 23 (Sy Damle, Partner at Latham & Watkins LLP; former General Counsel of the U.S. Copyright Office. “Artificial intelligence and intellectual property: Part 1 – Interoperability of AI and copyright law” Witness testimony before the U.S. House of Representatives Committee on the Judiciary, Subcommittee on Courts, Intellectual Property, and the Internet. 5/17/2023. Accessed 5/14/2024. Transcript downloaded from <https://judiciary.house.gov/committee-activity/hearings/artificial-intelligence-and-intellectual-property-part-i>) wtk

Even setting those policy objections aside, any effort to implement a statutory or collective licensing regime would need to grapple with a series of (perhaps insurmountable) practical problems.

The first and most obvious problem with a statutory licensing regime for AI training is scale. AI models can train on any media including, for example, long-form text (e.g., blog posts), short-form text (e.g., Tweets or forum comments), or images (full size or thumbnail). Often, AI developers train their models by pulling this content from the internet.52 Because virtually all of this content is eligible for copyright protection—regardless of its perceived aesthetic or expressive value53— any statutory licensing scheme would need to include virtually the entire internet within its scope. As a result, any licensing regime that triggered a payment obligation to any rightsholder whose work is used to train an AI model would require the administration of that license (and the payment of royalties) for billions upon billions of works—many of which are published online with no ownership information.

As precedent, some have pointed to the blanket licensing regime for musical works established by 2018 Music Modernization Act (“MMA”).54 At a high level, that regime created a new mechanism by which “digital music providers” could obtain a “blanket mechanical license” through a designated entity known as the “mechanical licensing collective” or “MLC.”55 The MLC administers the blanket license on behalf of musical work copyright owners by, for example, collecting royalties from digital music providers, seeking to identify the owners of the musical works in each song for which royalties are paid (a process known as “matching”), and distributing collected royalties to the appropriate owners.56 Congress required the MLC to hold any “unmatched” royalties for a certain period of time, and then distribute those funds to “copyright owners identified in the records of the collective” based on their “relative market share.”57

A statutory licensing regime for AI training would be massively more complicated than the one created by the MMA. Most obviously, the scope of a statutory license covering AI training activities would be orders of magnitude larger than any statutory license ever implemented. The MMA’s blanket license, for instance, covers a universe of musical works numbering in the tens of millions.58 By contrast, because AI models use a vast array of media for training, a statutory licensing scheme for AI training would need to cover, at least, every publicly accessible work on the internet—including every forum comment, online review, social media post, and business website. That universe of works likely numbers in the billions or tens of billions.

In addition, the works covered by this statutory license differ substantially from the works covered by the MMA’s blanket license. While it is not easy to identify the musical work owner for each sound recording, the endeavor is relatively straightforward if the MLC has access to the relevant data sources, which are in large part available to the MLC’s matching team.59 Acquiring sufficient data to identify the copyright owner of the billions of images or pieces of text on the internet would be much harder, if not impossible. Many photographs, social media posts, business websites, blog posts, or online reviews on the internet are published anonymously or pseudonymously, or simply with no information to identify the content’s author. I cannot estimate the breadth of the issue, but it seems likely that, for nearly all of the content that AI models could use for training purposes, identifying the rightful copyright author or owner will be an unattainable goal.

#### 2. Absolution Turn—licensing lets tech companies off the hook, devaluing human journalism

Lessin 24 (Jessica Lessin is the founder of The Information, a subscription publication covering the technology and finance industries. “Media Companies Are Making a Huge Mistake With AI” 5/24/2024. Accessed 5/25/2024. https://www.theatlantic.com/technology/archive/2024/05/fatal-flaw-publishers-making-openai-deals/678477) wtk

Publishers are deep in negotiations with tech firms such as OpenAI to sell their journalism as training for the companies’ models. It turns out that accurate, well-written news is one of the most valuable sources for these models, which have been hoovering up humans’ intellectual output without permission. These AI platforms need timely news and facts to get consumers to trust them. And now, facing the threat of lawsuits, they are pursuing business deals to absolve them of the theft. These deals amount to settling without litigation. The publishers willing to roll over this way aren’t just failing to defend their own intellectual property—they are also trading their own hard-earned credibility for a little cash from the companies that are simultaneously undervaluing them and building products quite clearly intended to replace them.

Late last year Axel Springer, the European publisher that owns Politico and Business Insider, sealed a deal with OpenAI reportedly worth tens of millions of dollars over several years. OpenAI has been offering other publishers $1 million to $5 million a year to license their content. News Corp’s new five-year deal with OpenAI is reportedly valued at as much as $250 million in cash and OpenAI credits. Conversations are heating up. As its negotiations with OpenAI failed, The New York Times sued the firm—as did Alden Global Capital, which owns the New York Daily News and the Chicago Tribune. They were brave moves, although I worry that they are likely to end in deals too.

That media companies would rush to do these deals after being so burned by their tech deals of the past is extraordinarily distressing. And these AI partnerships are far worse for publishers. Ten years ago, it was at least plausible to believe that tech companies would become serious about distributing news to consumers. They were building actual products such as Google News. Today’s AI chatbots are so early and make mistakes often. Just this week, Google’s AI suggested you should glue cheese to pizza crust to keep it from slipping off.

OpenAI and others say they are interested in building new models for distributing and crediting news, and many news executives I respect believe them. But it’s hard to see how any AI product built by a tech company would create meaningful new distribution and revenue for news. These companies are using AI to disrupt internet search—to help users find a single answer faster than browsing a few links. So why would anyone want to read a bunch of news articles when an AI could give them the answer, maybe with a tiny footnote crediting the publisher that no user will ever click on?

#### 3. The aff can’t solve journalism decline because of tech company consolidation

Peinert 23 (Erik Peinert is the Research Manager and Editor for the American Economic Liberties Project. “Saving the News from Big Tech: The Journalism Competition and Preservation Act” 6/9/2023. Accessed 5/29/2024. https://www.economicliberties.us/our-work/saving-the-news-from-big-tech-the-journalism-competition-and-preservation-act/) wtk

The rise and consolidation of digital ad markets in the 2000s changed this environment, both in terms of the distribution of news and the advertising structure financing it. Today, news has moved online and away from print. As a result, news consumers are often not going directly to journalism outlets but rather through various news “aggregators,” like Google Search, Google News, Facebook Timeline, and others. More than half of Americans report getting their news from social media, while 65% do so using a search engine like Google.[4] This is particularly true for smaller and more local outlets: whereas large publishers like the New York Times can attract readers directly to their websites, local and regional outlets obtain much more of their traffic through internet search and social media.

At the same time, Google and Facebook (now Meta) came to dominate online advertising markets through both acquisitions and specific business practices involving compiling dossiers on citizens and serving them ads as they move around the web. Both companies also serve as key news aggregators. Today, most money spent on digital advertising globally goes to Google or Facebook.[5] Facebook has a dominant position in social media, brought on by its own platform as well as its acquisitions of Instagram and WhatsApp.[6] For users reaching news primarily through social media, Facebook serves as an essential gatekeeper.

Google, though acquisitions and certain business practices, has built a complex and dominant online ad ecosystem, acting as a representative for both buyers and sellers, and as the market regulator. Its software is the plumbing for online advertising. When a publisher, such as a newspaper, wants to sell digital advertising, they use Google Ad Manager to sell ads, which are then sold on Google’s online ad exchange AdX, where they are sold and traded in rapid transactions akin to a stock market. Advertisers on the other side of this market buy the ad space through Google’s DV360 buying tools.[7]

As a result of this consolidation of the digital economy, most publishers are dependent on Facebook and Google for both distribution and ad financing. When a publisher goes to sell ads or set the terms for the commercial use of its journalism, it finds itself standing alone against one of two multinational technology giants. A news provider does not have meaningful outside options to take their business elsewhere, and therefore also does not have the position to push for better terms if it were to negotiate with the platforms. If a single news organization were to refuse to list links on or buy ads through Facebook or Google, neither platform’s revenue would be impacted in a significant way, even if journalism from publishers is, collectively, a critical component of the platforms’ traffic.

Key is that this is a difference solely of bargaining power. By virtue of being predominantly advertising businesses themselves — in 2021, 81% of Google’s revenue and 98% Facebook’s came from online advertising[8] — Facebook and Google benefit greatly from the journalism content that populates their platforms. Nonetheless, this disparity in bargaining power allows the tech firms to force news organizations to accept unfair and exploitative terms with limited compensation, knowing they have no alternatives.

#### 4. Or journalism is resilient to technological changes

Edge 24 (Marc Edge is a journalism researcher and author who lives in Ladysmith, BC. “Is journalism really dying? No, it’s just evolving” 1/28/2024. Accessed 6/15/2024. https://canadiandimension.com/articles/view/is-journalism-really-dying-no-its-just-evolving) wtk

Well, I guess that just leaves little old me. First let me say, as Shafer and Jarvis do, that journalism is not dying. It will always be with us, and anyone who doubts that should read Mitchell Stephens’ brilliant book A History of News, which chronicles the undeniable human impulse to both spread and consume news. Second, as I have long been saying, layoffs don’t necessarily mean that news media are dying. They could be a sign that they are adapting and, better yet, are able to pivot and thus survive.

The biggest question now is what kind of shape news provision will be in when newspapers finally crash and burn, not because anybody reads newspapers anymore, but because newspapers still provide most of the news most of us read for free online. I still firmly believe, as outlined in my 2014 book Greatly Exaggerated: The Myth of the Death of Newspapers, that print will survive in some diminished capacity. Newspapers that have not fallen into the clutches of private equity players and US hedge funds, after all, are making a successful transition to hybrid print and digital publications that rely more on reader revenues than on advertising. By investing in quality content and a successful online subscription scheme, the Globe and Mail has led the way in this country.

The problem, here as in the US, is that most of the newspaper industry is in the clutches of private equity players and US hedge funds, which are hindering their transition to the digital world by cutting their costs and stripping their assets in a short-term “harvesting” strategy. It’s a long story, which I tell in my recent book The Postmedia Effect, and American University professor Margot Susca tells in her new book Hedged.

A century ago, newspapers were the only fish in the news pond, but the industry began contracting in the 1920s when radio came along. This new medium didn’t kill off newspapers despite predictions that it would because, after all, who would want to read a newspaper when you could listen to someone read the news to you on the radio? Television likewise didn’t kill off either newspapers or radio, as instead each found its niche as smaller fish in an increasingly crowded pond. Will the Internet prove the meteor that finally kills off the dinosaurs? It hasn’t yet, and it’s been 30 years.

For the record, however, I have now adjusted my stance on this, changing the headline on my website from “Newspapers Are Not Dying” to “Newspapers Are Evolving.”

#### 5. Democracy doesn’t prevent war

**Doorenspleet 19** Renske Doorenspleet, Politics Professor at the University of Warwick. [Rethinking the Value of Democracy: A Comparative Perspective, Palgrave Macmillan, p. 239-243]

**The value of democracy has been taken for granted** until recently, but this assumption seems to be under threat now more than ever before. As was explained in Chapter 1, democracy’s claim to be valuable does not rest on just one particular merit, and scholars tend to distinguish three different types of values (Sen 1999). This book focused on the instrumental value of democracy (and hence not on the intrinsic and constructive value), and investigated the value of democracy for peace (Chapters 3 and 4), control of corruption (Chapter 5) and economic development (Chapter 6). **This study was based on a search of an enormous academic database** for certain keywords,6 **then pruned the thousands of articles** down to a few hundred articles (see Appendix) **which statistically analysed the connection between the democracy and the four expected outcomes**. The frst fiding is that **a reverse wave away from democracy has not happened** (see Chapter 2). Not yet, at least. **Democracy is not doing worse than before**, at least not **in comparative perspective**. While it is true that there is a dramatic decline in democracy in some countries,7 **a general trend downwards cannot yet be detected**. It would be better to talk about ‘stagnation’, as not many dictatorships have democratized recently, while democracies have not yet collapsed. Another fnding is that the instrumental value of democracy is very questionable. The feld has been deeply polarized between researchers who endorse a link between democracy and positive outcomes, and those who reject this optimistic idea and instead emphasize the negative effects of democracy. **There has been ‘no consensus’ in the quantitative literature on whether democracy has** instrumental value which leads some **beneficial general outcomes. Some scholars claim there is a consensus, but they only do so by ignoring a huge amount of literature which rejects their own point of view**. After undertaking a large-scale analysis of carefully selected articles published on the topic (see Appendix), **this book can conclude that the connections between democracy and expected benefts are not** as **strong** as they seem. Hence, we should not overstate the links between the phenomena. **The overall evidence is weak**. Take the expected impact of democracy on peace for example. As Chapter 3 showed, the study of democracy and interstate war has been a fourishing theme in political science, particularly since the 1970s. However, there are four reasons why democracy does not cause peace between countries, and why the empirical support for the popular idea of democratic peace is quite weak. **Most statistical studies have not found a strong correlation between democracy and interstate war at the dyadic level**. They show that **there are other—more powerful—explanations for war** and peace, and even that **the impact of democracy is a spurious one** (caveat 1). Moreover, **the theoretical foundation of the democratic peace hypothesis is weak**, and **the causal mechanisms are unclear** (caveat 2). In addition, **democracies are not** necessarily **more peaceful in general**, and **the evidence** for the democratic peace hypothesis **at the monadic level is inconclusive** (caveat 3). Finally, the process of **democratization is dangerous**. Living in a democratizing country means living in a less peaceful country (caveat 4). With regard to peace between countries, we cannot defend the idea that democracy has instrumental value. **Can the** (instrumental) **value** of democracy **be found in the prevention of civil war?** Or is the evidence for the opposite idea more convincing, and does democracy have a ‘dark side’ which makes civil war more likely? The findings are confusing, which is exacerbated by the fact that different aspects of civil war (prevalence, onset, duration and severity) are mixed up in some civil war studies. Moreover, defining civil war is a delicate, politically sensitive issue. Determining whether there is a civil war in a particular country is incredibly diffcult, while measurements suffer from many weaknesses (caveat 1). Moreover, **there is no linear link: civil wars are just as unlikely in democracies as in dictatorships** (caveat 2). Civil war is most likely in times of political change. **Democratization is a very unpredictable, dangerous process, increasing the chance of civil war significantly**. Hybrid systems are at risk as well: the chance of civil war is much higher compared to other political systems (caveat 3). More specifcally, both the strength and type of political institutions matter when explaining civil war. However, **the type of political system (e.g. democracy or dictatorship) is not the decisive factor at all** (caveat 4). Finally, democracy has only limited explanatory power (caveat 5). **Economic factors are far more significant than political factors** (such as having a democratic system) when explaining the onset, duration and severity of civil war. To prevent civil war, it would make more sense to make poorer countries richer, instead of promoting democracy. **Helping countries** to **democratize would even be a very dangerous idea**, as countries with changing levels of democracy are most vulnerable, making civil wars most likely. It is true that there is evidence that the chance of civil war decreases when the extent of democracy increases considerably. The problem however is that most countries do not go through big political changes but through small changes instead; those small steps—away or towards more democracy—are dangerous. Not only is the onset of civil war likely under such circumstances, but civil wars also tend to be longer, and the confict is more cruel leading to more victims, destruction and killings (see Chapter 4). A more encouraging story can be told around the value for democracy to control corruption in a country (see Chapter 5). Fighting corruption has been high on the agenda of international organizations such as the World Bank and the IMF. Moreover, the theme of corruption has been studied thoroughly in many different academic disciplines—mainly in economics, but also in sociology, political science and law. Democracy has often been suggested as one of the remedies when fghting against high levels of continuous corruption. So far, the statistical evidence has strongly supported this idea. As Chapter 5 showed, dozens of studies with broad quantitative, cross-national and comparative research have found statistically signifcant associations between (less) democracy and (more) corruption. However, there are vast problems around conceptualization (caveat 1) and measurement (caveat 2) of ‘corruption’. Another caveat is that **democratizing countries are the poorest performers with** regard to **controlling corruption** (caveat 3). Moreover, it is not democracy in general, but particular political institutions which have an impact on the control of corruption; and a free press also helps a lot in order to limit corruptive practices in a country (caveat 4). In addition, democracies seem to be less affected by corruption than dictatorships, but at the same time, there is clear evidence that economic factors have more explanatory power (caveat 5). In conclusion, more democracy means less corruption, but we need to be modest (as other factors matter more) and cautious (as there are many caveats). The perceived impact of democracy on development has been highly contested as well (see Chapter 6). Some scholars argue that democratic systems have a positive impact, while others argue that high levels of democracy actually reduce the levels of economic growth and development. Particularly since the 1990s, statistical studies have focused on this debate, and **the empirical evidence is clear: there is no direct impact of democracy on development**. Hence, both approaches cannot be supported (see caveat 1). The indirect impact via other factors is also questionable (caveat 2). Moreover, **there is too much variation in** levels of economic **growth** and development **among the dictatorial systems**, and there are huge regional differences (caveat 3). Adopting a one-size-ftsall approach would not be wise at all. In addition, in order to increase development, it would be better to focus on alternative factors such as improving institutional quality and good governance (caveat 4). There is not suffcient evidence to state that democracy has instrumental value, at least not with regard to economic growth. However, future research needs to include broader concepts and measurements of development in their models, as so far studies have mainly focused on explaining cross-national differences in growth of GDP (caveat 5). Overall, **the instrumental value of democracy is**—at best—tentative, or—if being less mild—**simply non-existent. Democracy is not necessarily better than any alternative form of government. With** regard to **many** of the expected **benefts—such as less war, less corruption and more economic development—democracy does deliver, but so do nondemocratic systems**. High or low levels of democracy do not make a distinctive difference. Mid-range democracy levels do matter though. Hybrid systems can be associated with many negative outcomes, while this is also the case for democratizing countries. Moreover, other explanations—typically certain favourable economic factors in a country—are much more powerful to explain the expected benefts, at least compared to the single fact that a country is a democracy or not. The impact of democracy fades away in the powerful shadows of the economic factors.8

### 2nc/1nr licensing fails

#### Licensing fails-- we have no ability to administer the licenses

Hansen and Brooke 23 (David Hansen, Executive Director of the Authors Alliance. Rachel Brooke, Senior Staff Attorney of the Authors Alliance. “RE: Policy Study on Artificial Intelligence, Docket Number 2023-6” 10/30/2023. Accessed 5/24/2024. [https://www.regulations.gov/comment/COLC-2023-0006-8976) wtk](https://www.regulations.gov/comment/COLC-2023-0006-8976)wtk) \*ECL= Extended Collective Licensing; CMO = Collective Management Organization

Finally, no party exists to administer such licenses, nor is it likely that an appropriate party or parties will emerge. As the Office observed in 2011 and again in 2015, ECL systems have typically been applied only to “limited types of works and uses, such as the use of published works for educational and scientific purposes, or the reproduction of works within an organization solely for internal use. . . . applying ECL to a mass digitization project that provides access to a wide range of works . . . would be a dramatic extension of the concept.”58 While a handful of EU countries have explored wider application (for example, to out of commerce works), those efforts are still new and are being done within the CMOs that have a history and experience working with ECLs.59

A credible system of adequate scale and scope would require an incredible investment of resources, and there are no clear candidates to make this investment. More than ten years ago, in the narrow field of book publishing, Google estimated a cost of $35 million to establish the Books Rights Registry (“BRR”) as part of its negotiated settlement in the Google Books litigation.60 The BRR which would have fulfilled many of the functions of a collective management organization administering an ECL-like regime under the proposed settlement, representing the interests of authors. While Google was willing to pay costs associated with such a system at that time, subsequent efforts (e.g., the Office’s mass digitization ECL pilot) revealed little appetite among others for such an investment.61 The costs of an ECL system broad enough to address the scope and scale of materials used for AI training datasets would likely dwarf the costs of any similar system we have seen to date.

More significantly, there are no CMOs nor combinations of existing CMOs that could adequately represent the range of rightsholders who would be implicated in training datasets, as the U.S. lacks the kinds of CMOs present in other jurisdictions that would be needed in order to represent the full range of rightsholders’ interests. Even in book publishing—an industry far more mature than industries representing many of the content types necessary to be licensed for AI training— the U.S. does not have an adequate CMO.62 The development of a whole cadre of new CMOs— representing groups as diverse as illustrators, graphic artists, authors, book publishers, academics, journalists, newspapers, photographers, software developers, bloggers, social media creators, and more—with sufficient infrastructure and wherewithal to represent their various constituencies would require years of effort to develop.

#### Licensing would result in very low payment to journalists

Damle 23 (Sy Damle, Partner at Latham & Watkins LLP; former General Counsel of the U.S. Copyright Office. “Artificial intelligence and intellectual property: Part 1 – Interoperability of AI and copyright law” Witness testimony before the U.S. House of Representatives Committee on the Judiciary, Subcommittee on Courts, Intellectual Property, and the Internet. 5/17/2023. Accessed 5/14/2024. Transcript downloaded from <https://judiciary.house.gov/committee-activity/hearings/artificial-intelligence-and-intellectual-property-part-i>) wtk

But it is also well understood that mandating bilateral negotiations between AI developers and individual rightsowners will be counterproductive. It will be impossible for legitimate AI developers to negotiate with each and every rightsholder who owns a copyright interest in one of the billions of individual pieces of data the developers’ models require. And attempts to build AI models using smaller sets of licensed or public domain material will lead to models that are less effective and, ironically, more likely to (inadvertently) create outputs that simply regurgitate their training data. Moreover, the economic dynamics of any bilateral licensing negotiation will be impossibly skewed: the supply of potential training data is effectively unlimited, meaning that no individual rightsholder will be able to demand more than nominal compensation for the use of its works.

#### Orphan works mean we can’t even identify copyright holders for most works

Hansen and Brooke 23 (David Hansen, Executive Director of the Authors Alliance. Rachel Brooke, Senior Staff Attorney of the Authors Alliance. “RE: Policy Study on Artificial Intelligence, Docket Number 2023-6” 10/30/2023. Accessed 5/24/2024. [https://www.regulations.gov/comment/COLC-2023-0006-8976) wtk](https://www.regulations.gov/comment/COLC-2023-0006-8976)wtk) \*ECL= Extended Collective Licensing

Second, at this scale, training datasets are likely to include massive numbers of orphan works for which licensing would be inappropriate. To effectively license those works under either an ECL or compulsory licensing scheme, adequate information about copyright ownership and licensing arrangements would either need to exist already or be readily obtainable in order to remit license payment to the rightsholders. While some rightsholders choose to make such information easily findable (for example, through registration with the U.S. Copyright Office), those registrations represent a tiny subset of all copyrighted works. For many works, however, no copyright information is available and—particularly for online works such as blog posts, online comments, Wikipedia contributions, and more—identifying and locating rightsholders would be virtually impossible.

As the Office’s previous studies on orphan works and mass digitization have highlighted, searches for rightsholders in far more narrowly scoped projects have incurred significant costs, with mostly had poor results.55 If a broad-based compulsory licensing or ECL regime were meant to be anything more than just a tax on generative AI systems, there would need be a way to pay out funds collected to the appropriate rightsholders, and that includes resolving the likely massive orphan works problem embedded within these training datasets. Experience from abroad in jurisdictions with systems that encourage searches for orphan works owners is not encouraging.56

#### Datasets are too large and too complex for licensing to work

Hansen and Brooke 23 (David Hansen, Executive Director of the Authors Alliance. Rachel Brooke, Senior Staff Attorney of the Authors Alliance. “RE: Policy Study on Artificial Intelligence, Docket Number 2023-6” 10/30/2023. Accessed 5/24/2024. https://www.regulations.gov/comment/COLC-2023-0006-8976) wtk

First, both compulsory licensing and extended collective licensing schemes are logistically infeasible because of the scale and complexity of the training datasets needed to train AI models. One of the key advantages of some of the largest and most successful AI models is that they are built upon an extremely large and diverse array of training materials. While creators of many of the largest models such as GPT-3 and GPT-4 have not fully disclosed the sources of training datasets, it is known that others have relied on sources such as Common Crawl (millions of works from across the internet), Books3, Arxiv, Stackexchange, and many millions others incorporating materials from every corner of the internet and across content types, including books, newspapers, websites, code, social media posts, legal materials, emails, YouTube subtitles, and more.53

These training datasets likely include millions of works owned by copyright holders from around the world. In a legal environment where permission is required to use copyrighted materials as training materials, licenses would need to cover a broad and diverse array of content. Doing otherwise (for example, creating a compulsory licensing or ECL system that applies to only a subset of copyrighted content) would both unfairly privilege certain rightsholders over others, while also exacerbating issues of bias in the models themselves (e.g., by encouraging use of only public domain content or easily licensable corporate-owned assets).54

### 2nc/1nr absolution turn

#### The turn outweighs the case—the tech industry is more powerful and will negotiate on their terms—only protecting the value of human journalism solves

Lessin 24 (Jessica Lessin is the founder of The Information, a subscription publication covering the technology and finance industries. “Media Companies Are Making a Huge Mistake With AI” 5/24/2024. Accessed 5/25/2024. https://www.theatlantic.com/technology/archive/2024/05/fatal-flaw-publishers-making-openai-deals/678477) wtk

The news industry finds itself in this dangerous spot, yet again, in part because it lacks a long-term focus and strategic patience. Once-family-owned outlets, such as The Washington Post and the Los Angeles Times, have been sold to interested billionaires. Others, like The Wall Street Journal, are beholden to the public markets and face coming generational change among their owners. Television journalism is at the whims of the largest media conglomerates, which are now looking to slice, dice, and sell off their empires at peak market value. Many large media companies are run by executives who want to live to see another quarter, not set up their companies for the next 50 years. At the same time, the industry’s lobbying power is eroding. A recent congressional hearing on the topic of AI and news was overshadowed by OpenAI CEO Sam Altman’s meeting with House Speaker Mike Johnson. Tech companies clearly have far more clout than media companies.

Things are about to get worse. Legacy and upstart media alike are bleeding money and talent by the week. More outlets are likely to shut down, while others will end up in the hands of powerful individuals using them for their own agendas (see the former GOP presidential candidate Vivek Ramaswamy’s activist play for BuzzFeed).

The long-term solutions are far from clear. But the answer to this moment is painfully obvious. Publishers should be patient and refrain from licensing away their content for relative pennies. They should protect the value of their work, and their archives. They should have the integrity to say no. It’s simply too early to get into bed with the companies that trained their models on professional content without permission and have no compelling case for how they will help build the news business.

Instead of keeping their business-development departments busy, newsrooms should focus on what they do best: making great journalism and serving it up to their readers. Technology companies aren’t in the business of news. And they shouldn’t be. Publishers have to stop looking to them to rescue the news business. We must start saving ourselves.

### 2nc/1nr can’t solve—consolidation

#### Ad revenue consolidation outweighs—ai is just one of many factors for decline

Karanicolas 23 (Michael Karanicolas is the executive director of the Institute for Technology, Law, and Policy, a new research collaboration between the UCLA School of Law and the UCLA Samueli School of Engineering. “Disrupting Journalism: How platforms have upended the news” 2/13/2023. Accessed 5/29/2024. https://www.cjr.org/special\_report/disrupting-journalism-how-platforms-have-upended-the-news-intro.php) wtk

After decades of shrinking revenues, and an increasing expectation among consumers that journalism should be free, the global media industry has reached a crisis point. As legacy news outlets shut down or lay off staff, misinformation and conspiracy theories run rampant, blurring the line between fantasy and reality. Trust in our institutions of governance continues to decline, fueling an alarming rise in extremism and political violence across previously stable democracies. In the Global South, the impact of journalism’s decline has been even more striking, with the rise of a new generation of autocrats skilled in manipulating the online conversation to suit their consolidation of power.

While there is no single cause for the decline in traditional news media, much of the blame has been focused on online platforms, whose profits traditionally have grown even as funding for journalism dries up. There is a connection between these two trends, since the platforms’ control over how we access information means that news media organizations are dependent on them to distribute their products. Facebook and Google have used this to leverage a dominant position over the online advertising market, keeping the lion’s share of profits for themselves and disrupting a key revenue source for supporting journalism.

#### Bargaining power is the fundamental problem

Peinert 23 (Erik Peinert is the Research Manager and Editor for the American Economic Liberties Project. “Saving the News from Big Tech: The Journalism Competition and Preservation Act” 6/9/2023. Accessed 5/29/2024. https://www.economicliberties.us/our-work/saving-the-news-from-big-tech-the-journalism-competition-and-preservation-act/) wtk

A key reason for the decline in journalism is that Facebook and Google, as dominant platforms in online ad markets, have redirected the flow of ad money from publishers to themselves through a variety of exploitative terms imposed on publishers and news organization. These include:

Denying news organizations access to their own readers’ data;

Forcing publishers to provide free articles (for example, Google’s “First Click Free” policy);

Requiring publication formats that direct readers to the platform’s content, such as Google’s AMP format or Facebook’s “Instant Articles”;

Failing to reward or protect original content against publications copying content; and

Insufficient compensation for content.[2]

The fundamental problem is a bargaining imbalance. Google may need the news, but it doesn’t need any one newspaper, whereas newspapers all individually need Google to survive. And whereas antitrust enforcement has condoned the platforms’ recent acquisition of their dominant position, it prohibits newspapers from banding together to bargain over pricing, data, ad revenue, or other terms.

### 2nc/1nr journalism resilient

#### Journalism is stronger than ever despite technological challenges

Nassau 23 (Julieta Nassau, 2023 WPI Fellow, Journalist, La Nación in Buenos Aires, Argentina. “Why investigative journalism is now stronger than ever” 4/21/2023. Accessed 6/15/2024. [https://worldpressinstitute.org/why-investigative-journalism-is-now-stronger-than-ever/) wtk](https://worldpressinstitute.org/why-investigative-journalism-is-now-stronger-than-ever/)wtk)

New technologies have been very transformative for investigative journalism, especially to help journalists’ processes, like indexation of large amounts of documents. There are many processes that can be automated that 20 years ago or less could only be done manually, and that took a great deal of time.

Another great contribution is the possibility of undertaking what I call “impossible investigations,” which could not have been done without the use of this artificial intelligence. For example, we promoted an investigation to map illegal landing tracks in the Amazon. The only way in such a complex geography and so difficult to penetrate was to use “eyes on the sky.” We taught the computer to recognize how the land looks when these irregular tracks are built. Then the computer searches through a number of satellite images for matching images that may indicate that illegal tracks are there.

What do you think are the risks of the use of artificial intelligence for journalism?

Technology is always at the service of journalism and not the other way around. The processes have to be guided by editorial criteria. Technology is a tool, and many times it is imperfect. These new technologies complement the traditional work of investigative journalism, which requires meticulous checking, interviews and field reporting. But they never replace the reporter.

### 2nc/1nr democracy doesn’t solve war

#### Democracy doesn’t solve war

**Ghatak et al. 17**—Sam Ghatak is a Lecturer in Political Science at the University of Tennessee Knoxville; Aaron Gold is a PhD Student in Political Science at UT Knoxville; Brandon C. Prins is a Professor and Director of Graduate Studies of Political Science at UT Knoxville [“External threat and the limits of democratic pacifism,” *Conflict Management and Peace Science*, Vol. 34, No. 2, p. 141-159, Emory Libraries]

**Conclusion**

**It has become a stylized fact that dyadic democracy lowers the hazard of armed conflict**. While **the Democratic Peace has faced many challenges**, we believe **the most significant challenge has come from the argument that the pacifying effect of democracy is epiphenomenal to territorial issues, specifically the external threats that they pose**. **This argument sees the lower hazards of armed conflict among democracies not as a product of shared norms or institutional structures, but as a result of settled borders**. Territory, though, remains only one geo-political context generating threat, insecurity, and a higher likelihood of armed conflict. Strategic rivalry also serves as an environment associated with fear, a lack of trust, and an expectation of future conflict. **Efforts to assess democratic pacifism have largely ignored rivalry as a context conditioning the behavior of democratic leaders**. To be sure, research demonstrates rivals to have higher probabilities of armed conflict and democracies rarely to be rivals. But **fundamental to the Democratic Peace is the notion that even in the face of difficult security challenges and salient issues, dyadic democracy will associate with a lower likelihood of militarized aggression. But the presence of an external threat, be that threat disputed territory or strategic rivalry, may be the key mechanism by which democratic leaders, owing to audience costs, resolve and electoral pressures, fail to resolve problems nonviolently**.

**This study has sought a ‘‘hard test’’ of the Democratic Peace by testing the conditional effects of joint democracy on armed conflict when external threat is present**. We test three measures of threat: territorial contention, strategic rivalry, and a threat index that sums the first two measures. **For robustness checks, we use two additional measures of our dependent variable: fatal MID onset, and event data from the Armed Conflict Database**, which can be found in our Online Appendix. As most studies report, democratic dyads are associated with less armed conflict than mixed-regime and autocratic dyads. In every one of our models, when we control for each measure of external threat, joint democracy is strongly negative and significant and each measure of threat is strongly positive and significant. Here, liberal institutions maintain their pacific ability and **external threats clearly increase conflict propensities**. However, **when we test the interactive relationship between democracy and our measures of external threat, the pacifying effect of democracy is less visible**. Park and James (2015) find some evidence that when faced with an external threat in the form of territorial contention, the pacifying effect of joint democracy holds up. This study does not fully support the claims of Park and James (2015). **Using a longer timeframe, we find more consistent evidence that when faced with an external threat, be it territorial contention, strategic rivalry, or a combination, democratic pacifism does not survive**. What are the implications of our study? First, while it is clear that we do not observe a large amount of armed conflict among democratic states, if we organize interstate relationships along a continuum from highly hostile to highly friendly, we are probably observing what Goertz et al. (2016) and Owsiak et al. (2016) refer to as ‘‘lesser rivalries’’ in which ‘‘both the frequency and severity of violent interaction decline. Yet, the sentiments of threat, enmity, and competition that remain—along with the persistence of unresolved issues—mean that lesser rivalries still experience isolated violent episodes (e.g., militarized interstate disputes), diplomatic hostility, and non-violent crises’’ (Owsiak et al., 2016). Second, our findings show that **the pacific benefits of liberal institutions or externalized norms are not always able to lower the likelihood of armed conflict when faced with external threats, whether those hazards are disputed territory, strategic rivalry, or a combination of the two**. The structural environment clearly influences democratic leaders in their foreign policy actions more than has heretofore been appreciated. **Audience costs, resolve, and electoral pressures, produced from external threats, are powerful forces that are present even in jointly democratic relationships**. **These forces make it difficult for leaders to trust one another, which inhibits conflict resolution and facilitates persistent hostility**. **It does appear**, then, **that there is a limit to the Democratic Peace**.

## answers to model collapse advantage

### 1nc model collapse advantage frontline

#### On the model collapse advantage:

#### 1. No model collapse—the original training set will anchor new models even if they are trained with synthetic data

Claburn 24 (Thomas Claburn is a senior reporter at The Register. “Big brains divided over training AI with more AI: Is model collapse inevitable?” 5/9/2024. Accessed 6/10/2024. https://www.theregister.com/2024/05/09/ai\_model\_collapse/?td=keepreading) wtk

AI model collapse – the degradation of quality expected from machine learning models that recursively train on their own output – is not inevitable, at least according to 14 academics.

The risk that ongoing generative AI output, known as synthetic data, will dilute human-created organic data and impair the performance of models trained on this increasingly fabricated corpus was highlighted by a separate group last year, in a paper titled: "The Curse of Recursion: Training on Generated Data Makes Models Forget."

Ilia Shumailov, lead author of that paper, spoke to The Register earlier this year about this phenomenon, which has been documented in other studies.

Now another set of boffins – Matthias Gerstgrasser, Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Henry Sleight, John Hughes, Tomasz Korbak, Rajashree Agrawal, Dhruv Pai, Andrey Gromov, Daniel Roberts, Diyi Yang, David Donoho, and Sanmi Koyejo – contend that the problem of training AI on AI-made data isn't significant, given the way that model training is actually done.

This latest baker's dozen plus one – from Stanford, AI safety group Constellation, the University of Maryland at College Park, MIT, and Sequoia Capital – make the case for not worrying in a paper titled: "Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data."

It's worth noting that some of these boffins acknowledge support through grants from commercial entities including OpenAI and Google, although the authors insist their research results do not necessarily reflect the positions or policies of their funders.

Gerstgrasser, a postdoctoral research associate at Harvard SEAS and visiting postdoctoral scholar at Stanford, outlined on social media the argument he and his colleagues want to make.

"As AI-generated content becomes more prevalent on the internet, there's a growing concern that future AI models will be trained on this 'tainted' data," he asserted. "It's like a virus that could infect the entire AI ecosystem!

"Many experts have warned that this could lead to a doomsday scenario for AI. If models keep getting worse and worse with each generation, we could face an 'AI apocalypse'! But don't panic just yet …"

Gerstgrasser argued that while previous studies have warned about this "doomsday scenario," all that research relies on the assumption that each succeeding generation of AI would train exclusively on the synthetic data produced by the previous generation model.

He argues that legacy data won't just be discarded. Instead of being replaced every generation, it's more likely to accumulate – the synthetic data will just get mixed with the organic data, and the resulting model will continue to perform.

"Our findings extend these prior works to show that if data accumulates and models train on a mixture of 'real' and synthetic data, model collapse no longer occurs," Gerstgrasser et al declare in their "Is Model Collapse Inevitable?" paper.

"[T]hese results strongly suggest that the 'curse of recursion' may not be as dire as had been portrayed – provided we accumulate synthetic data alongside real data, rather than replacing real data by synthetic data only."

#### 2. Or the aff can’t solve in time because we will run out of high quality data by 2026.

Zuckerman 23 (Ethan Zuckerman is associate professor of public policy, information and communication at the University of Massachusetts Amherst. “What happens when AI trains itself?” October 2023. Accessed 6/10/2024. https://www.prospectmagazine.co.uk/ideas/technology/62810/ai-artificial-intelligence-trains-itself-zuckerman) wtk

Even if AI entrepreneurs can find their way through these legal thickets, another more existential barrier may cap the growth of large language models: the limits of human creativity. A recent paper from Epoch, a team of researchers focused on the future of AI, predicts that AI companies will run out of “high-quality” language data like “books, news articles, scientific papers, Wikipedia, and filtered web content” as soon as 2026. While the number of books and scientific papers authored per year is massive, it’s also finite, and the appetites of large language models have grown exponentially. Companies like OpenAI may be able to train models on lower-quality data, such as comments on social media, but as regards filtered, edited content, “The high-quality language stock will almost surely be exhausted before 2027 if current trends continue.”

#### 3. The plan causes worse AI models

Mantegna 24 (Micaela Mantegna, Affiliate at the Berkman Klein Center at Harvard University (United States) and the Center for Technology and Society at San Andres University (Argentina). “ARTificial: Why Copyright Is Not the Right Policy Tool to Deal with Generative AI” *Yale Law Journal* Volume 133, 4/22/2024. Accessed 5/27/2024. https://www.yalelawjournal.org/forum/artificial-why-copyright-is-not-the-right-policy-tool-to-deal-with-generative-ai) wtk

6. Licenses Incentivize Data Substitution, Impacting Core AI Ethics Principles

Another consideration is that a decision made in a GAI case will have an impact well beyond it, affecting the whole AI-development ecosystem, as it might create incentives for data substitution. AI models based on ML learn from the training data. If the most adequate data to train a model requires paying a licensing fee, it would incentivize developers to cut costs and substitute that quality data for other sources that might not be as appropriate, but less costly. Doing so would be akin to changing the ingredients of a recipe with inferior substitutes. Inferior data could result in a concept commonly referred to as “garbage in, garbage out,”173 where the quality of an AI model’s output is directly affected by the quality of the input data.

Web-scraping, data collection, and databases are a vital component of the AI ecosystem. Legally, there are different approaches to regulating databases. In the EU there is a “sui generis” right to protect the content of the database, and the structure under copyright, if certain conditions are met.174 In the United States, as explained above, the content of the database will only be protectable if there is enough originality in the criteria, selection, arrangement, or composition of the data.

This difference in regulatory approaches has an enormous geopolitical impact. It influences where AI companies will locate themselves, and which databases they will use to train the systems, as companies will try to reduce the cost of having to license the data.

Strong copyright provisions and a lack of carve-outs for data mining for training purposes could potentially incentivize the use of data substitutions as a way to avoid paying for licensed datasets. This would translate directly to subpar performance from AI models for the reasons explained above, which would likely be more biased and less fair—and consequently, human beings and society will be harmed in the process.

#### 4. Data poisoning doesn’t cause war—AI weeds out misinformation

Jessica Cox & Heather Williams, 21 – \*Director, Nuclear Policy Directorate. NATO International Staff, \*\*Director, Project on Nuclear Issues @csisponi Senior Fellow, Intl Security Program, “The Unavoidable Technology: How Artificial Intelligence Can Strengthen Nuclear Stability” The Washington Quarterly Volume 44, 2021 - Issue 1, Taylor & Francis database)

AI can similarly enhance the ability to discriminate between real and false information, which is critical to preventing miscalculation or mistake in a crisis. A number of “close calls” have occurred throughout the years, in which nuclear weapons use was only avoided through human intervention based on uncertain information.22 In 1983, NATO undertook Command Post Exercise Able Archer to “practice command and staff procedures, with particular emphasis on the transition from conventional to non-conventional operations, including the use of nuclear weapons.”23 At the time, Soviet paranoia about a US nuclear first strike, with preparations under cover of a war game, was at an all-time high. When NATO began its Able Archer exercise, Soviet officials thought the exercise was real and put mobile intercontinental ballistic missiles (ICBMs) on a three-minute alert.24 The same year, the Soviet early warning computer system signaled five incoming US Minuteman ICBMs. The watch officer on duty, Soviet Lt. Col. Stanislav Petrov, did not report the incident, concluding that it must be a false alarm. If he had reported that US nuclear missiles were inbound, the Soviets would have followed their nuclear doctrine and retaliated with no time to double-check or negotiate with the United States.25 Petrov ultimately made the right call—as the Soviet early warning system had mistaken sun shining off clouds for incoming missiles.

Even after the end of the Cold War, in 1995, a Norwegian civilian research rocket launch was detected by the radar crews from the Russian Missile Attack Warning System (MAWS) and mistaken for a US Trident II submarine-launched ballistic missile. Command and control procedures were enacted, including notification of President Boris Yeltsin and activation of the Russian “cheget” system—essentially its nuclear football. Russian authorities eventually determined that this was not a nuclear-armed ballistic missile launch after reviewing satellite and other intelligence information.26

While these and other incidents were ultimately resolved successfully through the intervention of military experts and political leaders,27 they illustrate the dangers of imperfect systems that rely on human analysts with potentially inaccurate or inadequate information. As defense researcher Jaganath Sankaran argues, “future applications of AI to nuclear command and control should aspire to create an algorithm that could argue in the face of overwhelming fear of an impending attack that a nuclear launch isn’t happening.”28 Such an approach could both reduce the fog of war and reassure decision-makers that their course of action is correct in the face of uncertain information, either using nuclear weapons or refraining from it.

Although most nuclear-armed nations now use multiple systems to reduce the chances of a false positive warning signal, advancements in AI could significantly improve confidence in and functionality of these systems when integrated with human decision-making. AI tools could play a significant role in helping to identify patterns of life and reduce potential operator biases in conducting analysis. For instance, in order to be able to quickly discriminate between a country’s launch of a sounding/research rocket as opposed to an ICBM, which have very similar radar signatures, big data analytical tools could be used to collate and process massive amounts of electronic data—including signals, imagery, and open source collection—over time to identify patterns of behavior unique to each type of launch. Then, if there is an ambiguous or unexpected launch, these systems would be able to quickly determine whether or not the current circumstances more closely resemble one type of launch over the other—actually preventing future close calls or misinterpretations of data, as occurred in Able Archer in 1983. A human analyst would likely still need to make a final determination, but through the use of AI and data analytics, they would have a more accurate, timely, and complete picture on which to base any decisions.

### 2nc/1nr no model collapse

#### Reject aff studies—they assume AI generated text replaces data, not that it supplements it

Gerstgrasser et al. 24 (Matthias Gerstgrasser, postdoc jointly at Stanford Computer Science and at Harvard University’s School of Engineering and Applied Sciences. Co-authors: Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Henry Sleight, John Hughes, Tomasz Korbak, Rajashree Agrawal, Dhruv Pai, Andrey Gromov, Daniel A. Roberts, Diyi Yang, David L. Donoho, Sanmi Koyejo. “Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data” 4/29/2024. Accessed 6/15/2024 from [https://arxiv.org/abs/2404.01413) wtk](https://arxiv.org/abs/2404.01413)wtk)

The proliferation of generative models, combined with pretraining on webscale data, raises a timely question: what happens when these models are trained on their own generated outputs? Recent investigations into model data feedback loops proposed that such loops would lead to a phenomenon termed *model collapse*, under which performance progressively degrades with each model-data feedback iteration until fitted models become useless. However, those studies largely assumed that new data *replace* old data over time, where an arguably more realistic assumption is that data *accumulate* over time. In this paper, we ask: what effect does accumulating data have on model collapse? We empirically study this question by pretraining sequences of language models on text corpora. We confirm that replacing the original real data by each generation’s synthetic data does indeed tend towards model collapse, then demonstrate that accumulating the successive generations of synthetic data alongside the original real data avoids model collapse; these results hold across a range of model sizes, architectures, and hyperparameters. We obtain similar results for deep generative models on other types of real data: diffusion models for molecule conformation generation and variational autoencoders for image generation. To understand why accumulating data can avoid model collapse, we use an analytically tractable framework introduced by prior work in which a sequence of linear models are fit to the previous models’ outputs. Previous work used this framework to show that if data are replaced, the test error increases with the number of model-fitting iterations; we extend this argument to prove that if data instead accumulate, the test error has a finite upper bound independent of the number of iterations, meaning model collapse no longer occurs. Our work provides consistent empirical and theoretical evidence that data accumulation avoids model collapse.

#### Both experiments and regression analysis prove no model collapse

Gerstgrasser et al. 24 (Matthias Gerstgrasser, postdoc jointly at Stanford Computer Science and at Harvard University’s School of Engineering and Applied Sciences. Co-authors: Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Henry Sleight, John Hughes, Tomasz Korbak, Rajashree Agrawal, Dhruv Pai, Andrey Gromov, Daniel A. Roberts, Diyi Yang, David L. Donoho, Sanmi Koyejo. “Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data” 4/29/2024. Accessed 6/15/2024 from [https://arxiv.org/abs/2404.01413) wtk](https://arxiv.org/abs/2404.01413)wtk)

Our findings extend these prior works to show that if data accumulates and models train on a mixture of “real” and synthetic data, model collapse no longer occurs. We show this both experimentally on causal transformers for language modeling, diffusion models for molecule generation, and variational auto-encoders on image data as well as theoretically for linear regression. Together, these results strongly suggest that the “curse of recursion” may not be as dire as had been portrayed – provided we accumulate synthetic data alongside real data, rather than replacing real data by synthetic data only.

### 2nc/1nr can’t solve in time

#### AI will run out of human-made text—AI consumes information too fast

O’Brien 24 (Matt O’Brien, Associated Press, “AI ‘gold rush’ for chatbot training data could run out of human-written text as early as 2026” 6/6/2024. Accessed 6/15/2024. https://www.pbs.org/newshour/economy/ai-gold-rush-for-chatbot-training-data-could-run-out-of-human-written-text-as-early-as-2026) wtk

Artificial intelligence systems like ChatGPT could soon run out of what keeps making them smarter — the tens of trillions of words people have written and shared online.

A new study released Thursday by research group Epoch AI projects that tech companies will exhaust the supply of publicly available training data for AI language models by roughly the turn of the decade — sometime between 2026 and 2032.

Comparing it to a “literal gold rush” that depletes finite natural resources, Tamay Besiroglu, an author of the study, said the AI field might face challenges in maintaining its current pace of progress once it drains the reserves of human-generated writing.

In the short term, tech companies like ChatGPT-maker OpenAI and Google are racing to secure and sometimes pay for high-quality data sources to train their AI large language models – for instance, by signing deals to tap into the steady flow of sentences coming out of Reddit forums and news media outlets.

In the longer term, there won’t be enough new blogs, news articles and social media commentary to sustain the current trajectory of AI development, putting pressure on companies to tap into sensitive data now considered private — such as emails or text messages — or relying on less-reliable “synthetic data” spit out by the chatbots themselves.

“There is a serious bottleneck here,” Besiroglu said. “If you start hitting those constraints about how much data you have, then you can’t really scale up your models efficiently anymore. And scaling up models has been probably the most important way of expanding their capabilities and improving the quality of their output.”

#### High quality text will be completely tapped by 2026

Rao 24 (Devika Rao is a staff writer at The Week. “All-powerful, ever-pervasive AI is running out of internet” 6/5/2024. Accessed 6/15/2024. [https://theweek.com/tech/ai-running-out-of-data) wtk](https://theweek.com/tech/ai-running-out-of-data)wtk)

Artificial intelligence (AI) has relied on high-quality language data to train its models, but supply is running low. That depletion is forcing companies to look elsewhere for data sourcing as well as to change their algorithms to use data more efficiently.

What is the scope of AI's data problem?

Artificial intelligence needs to be trained, and data and information is used to accomplish that. Trouble is, the data is running out. A paper by Epoch, an AI research organization, found that AI could exhaust all the current high-quality language data available on the internet as soon as 2026. This could pose a problem as AI continues to grow. "The issue stems from the fact that, as researchers build more powerful models with greater capabilities, they have to find ever more texts to train them on," said the MIT Technology Review. The quality of the data used in training AI is important. "The [data shortage] issue stems partly from the fact that language AI researchers filter the data they use to train models into two categories: high-quality and low-quality," said the Review. "The line between the two categories can be fuzzy," but "text from [high-quality data] is viewed as better-written and is often produced by professional writers."

AI models require vast amounts of data to be functional. For example, "the algorithm powering ChatGPT was originally trained on 570 gigabytes of text data, or about 300 billion words," said Singularity Hub. In addition, "low-quality data such as social media posts or blurry photographs are easy to source but aren't sufficient to train high-performing AI models," and could even be "biased or prejudiced or may include disinformation or illegal content which could be replicated by the model." Much of the data on the internet is considered useless for AI modeling. Instead, "AI companies are hunting for untapped information sources and rethinking how they train these systems," said The Wall Street Journal. "Companies also are experimenting with using AI-generated, or synthetic, data as training material — an approach many researchers say could actually cause crippling malfunctions."

### 2nc/1nr plan causes worse ai

#### The plan creates barriers to quality training data and undermines broader AI safety research

Thamkul 23 (Janel Thamkul, Deputy General Counsel for Anthropic PBC, “Notification of Inquiry Regarding Artificial Intelligence and Copyright: Public Comments of Anthropic PBC” Before the United States Copyright Office Washington, D.C. 10/30/2023. Accessed 5/23/2024. Downloaded from https://www.regulations.gov/comment/COLC-2023-0006-9021) wtk

Because training LLMs is a fair use, we do not believe that licensing is necessary per se. To be sure, for a variety of reasons, developers may choose to procure special access to or use of particular datasets as part of commercial transactions. However, a regime that always requires licensing for use of material in training would be inappropriate; it would, at a minimum, effectively lock up access to the vast majority of works, since most works are not actively managed and licensed in any way. 32

Constraining use of existing works in this way would also impede efforts to address other concerns about AI, such as the potential for bias. 33 Having broad, diverse datasets is critical to combating the potential for bias, as well as other measures of model quality. Additionally, it will harm U.S. efforts to safely and effectively develop and deploy AI.

The likely result of preventing training on existing works absent permission would be not only less useful generative AI, undermining people’s ability to use them to create new works or perform other non-infringing tasks, 34 but also a more concentrated market. The developers of generative AI models would face higher barriers to entry, because they would not be able to rely on web crawling or other means of inexpensively analyzing content at scale. Only the most highly resourced entities would be able to engage in costly and burdensome data licensing processes. Efforts to research the safety and interpretability of these models would be particularly undermined, and likely result in only the most highly resourced entities being able to advance research in this space, as our empirical work shows that research on the largest and most capable systems is qualitatively different than for small models.

### 2nc/1nr no data poisoning impact

#### Data poisoning won’t cause war—their evidence ignores the positives of AI on nuclear decision-making

Jessica Cox & Heather Williams, 21 – \*Director, Nuclear Policy Directorate. NATO International Staff, \*\*Director, Project on Nuclear Issues @csisponi Senior Fellow, Intl Security Program, “The Unavoidable Technology: How Artificial Intelligence Can Strengthen Nuclear Stability” The Washington Quarterly Volume 44, 2021 - Issue 1, Taylor & Francis database)

From a nuclear policy perspective, the majority of scholarship addressing the implications of AI, with a few important exceptions,8 argues that military applications of AI will have a destabilizing effect on strategic stability and global security. Areas regularly cited as problematic include the potential for deep fakes, human bias carried over into AI programming, cyber vulnerabilities, exacerbating great power rivalry, application to nuclear command and control, security dilemmas, and potential for loss of human control.9 This focus creates a rather defeatist paradigm, particularly for policymakers confronted with rapid advances in technology and limited thinking on how it might be applied in a stabilizing way. Whereas many of these studies treat AI as a monolith, we argue that the opportunities and risks associated with AI depend on its application, and considering their impact on nuclear weapons policies requires a greater degree of granularity. Can some military applications of AI strengthen deterrence, for example? Might it facilitate advances in arms control, at a time when the international community desperately needs new ideas and new verification techniques?

Despite the potential risks inherent with adopting AI in nuclear systems, many of the potential benefits to both deterrence and arms control have been overlooked or ignored. Additionally, many in the nuclear policy community rely on the most extreme negative examples to make the case that AI is too dangerous to adopt into nuclear policies and practices. These views are short-sighted and ignore the reality that AI will increasingly play a role in all aspects of our civil and military policies—and that potential great power competitors such as China and Russia are investing heavily in AI-related military technology.10 The broader nuclear policy community should therefore start thinking now about how to harness the power of AI as a tool for greater stability, transparency, and security as well as steps that can be put in place to mitigate the most significant risk factors.

We argue that considering AI’s impact on nuclear weapons—both in the field of nuclear deterrence and for strategic arms control—requires a more nuanced approach, whereby it offers both opportunities as well as challenges for strategic stability. Because the risks of AI are increasingly familiar and discussed in increasing scope and detail by experts and scholars, we choose here to focus our exploration on the potential benefits of AI on nuclear policy, which have received far less attention. We do not mean to diminish the potential risks associated with increasing reliance on AI, but rather to offer a more optimistic picture of how AI might also contribute to nuclear deterrence and arms control in unique and helpful ways in order to answer this question: how can decision-makers pursue potentially stabilizing applications of AI while managing the risks associated with AI and nuclear systems?

## science research disadvantage

### 1nc science research da

#### The next offcase position is the science research da

#### Broad “fair use” exceptions in the status quo are key to scientific research—the plan reverses that

MacKie-Mason and Li 23 (Jeffrey K. MacKie-Mason is the University Librarian and Chief Digital Scholarship Officer, University of California, Berkeley, and a professor in the UC Berkeley School of Information and the Department of Economics. Haipeng Li, University Librarian at University of California Merced. “Re: Notice of inquiry (“NOI”) and request for comments, Artificial Intelligence and Copyright, Docket No. 2023-6” 10/30/2023. Accessed 5/22/2024. [https://drive.google.com/file/d/1nNvynoekNREUhLA\_h5AqE5HzVuTxymQm/view) wtk](https://drive.google.com/file/d/1nNvynoekNREUhLA_h5AqE5HzVuTxymQm/view)wtk) \*TDM = text and data mining

And so, rightsholders, academic publishers, and content aggregators regularly “contract around” fair use by requiring libraries to negotiate for otherwise lawful activities (such as conducting TDM or training AI for research), and often to pay additional fees for the right to conduct these lawful activities on top of the cost of licensing the content, itself. When such costs are beyond institutional reach, the publisher or vendor may then offer similar contractual terms directly to research teams, who may feel obliged to agree in order to get access to the content they need. 40 Vendors may charge tens or even hundreds of thousands of dollars for this type of access. 41

This “pay-to-play” landscape of charging institutions for the opportunity to rely on existing statutory rights is particularly detrimental for TDM research methodologies. As global scholars have recently explained:

Licensing is not an affordable or viable option for many critical TDM projects. TDM research often requires use of massive datasets with works from many publishers, including copyright owners that cannot be identified or are unwilling to grant licenses. Forcing researchers to use only licensed or public domain content (i.e., content in which there is no enforceable copyright) can restrict topics of study, hamper reproducibility and validation, bias results, and dissuade researchers from undertaking projects. 42

If the Copyright Office were to enable rightsholders to opt-out of training AI for research and teaching fair uses, then academic institutions and scholars would face even greater hurdles in licensing content for research purposes. 43 It would be operationally difficult for academic publishers and content aggregators to amass and license the “leftover” body of copyrighted works that remain eligible for AI training. Costs associated with publishers’ efforts in compiling “AI-training-eligible” content would be passed along as additional fees charged to academic libraries, who are already financially constrained to preserve TDM and other fair uses for scholars. In addition, rightsholders might opt out of allowing their work to be used for AI training fair uses, and then turn around and charge AI usage fees to scholars (or libraries)—essentially licensing back fair uses for research. These scenarios would impede scholarship by or for research teams who lack grant or institutional funds to cover these additional expenses; penalize research in or about underfunded disciplines or geographical regions; and result in bias as to the topics and regions studied.

Copyright exceptions like Section 107 matter for research: they result in a higher production of new works of scholarship, and drive TDM mining as a means of extracting information and advancing knowledge. 44, 45 Scholars need to be able to utilize existing knowledge resources to create new knowledge goods. 46 Indeed, the availability of openly accessible scholarship (without legal or technical barriers) during the COVID-19 pandemic directly benefited public health policy- and decision-making. 47 Congress and the Copyright Office clearly understand the importance of facilitating access and usage rights, having implemented Section 107 without any statutory or regulatory exclusions or opt-outs. This status quo should be preserved: to promote the progress of science and useful arts, copyright holders should not be permitted to opt out of having their works used for AI training when such AI training or practices would be a fair use—and particularly in the nonprofit educational or research contexts.

#### Research is key to sustainability and human survival—copyright exceptions are key

Geiger and Jütte 22 (Christophe Geiger is Professor of law at the Luiss Guido Carli University in Rome (Italy) and President of the International Association for the Advancement of Teaching and Research in Intellectual Property (ATRIP). Bernd Justin Jütte is Assistant Professor in Intellectual Property Law, University College Dublin, Sutherland School of Law (Ireland) and Senior Researcher Vytautas Magnus University, Faculty of Law (Kaunas, Lithuania). “Conceptualizing a ‘Right to Research’ and its Implications for Copyright Law: An International and European Perspective” Joint PIJIP/TLS Research Paper Series, July 2022. Accessed 5/28/2024. https://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=1079&context=research) wtk

Research is a precondition to create a sustainable future. It is indispensable to realize substantive and programmatic human and fundamental rights, rights which inform global, European and national policies and which take shape in concrete actions envisaged to create a more sustainable global community. Copyright plays an important role in enabling access to information to create in order to meet the ambitious goals set out, for example, in EU policies and in the Sustainable Development Goals at international level. A right to research, expressed as such, can help to realize these goals by providing convincing arguments for copyright reform, but has not been included expressly in international or European fundamental rights instruments. Its pieces are, however, present in the canon of European and international fundamental rights and which are an elementary part of the aims and objectives of the Union.

As a result, a right to research as we propose is influenced by existing concepts, interpretations and understandings present in the fundamental rights of the ECHR and the CFREU, as well as international human rights instruments, including the UCDHR, the ICCPR, and the ICESCR. A right to research as we construct is also rooted in the political mission of the European Union and the United Nations Sustainable Development Goals, which establish political goals for which continued and persistent research efforts are a precondition. A right to research is, therefore, conceived as a constitutional imperative. Giving shape and express recognition to this right under its own label will give better weight to research as a necessary policy goal in political discourses and negotiations on the future shape of copyright law. This mission is admittedly much broader than designing a sustainable copyright system, but copyright plays an important part in building a sustainable global future.

For copyright law, a right to research will serve two main functions: first, it will inform legislatures when debating and (re-) drafting copyright law, when creating, repealing and shaping exclusive rights and when designing exceptions and limitations that permit uses necessary and indispensable for research; second, it will inform the judiciary when applying copyright law in the light of fundamental rights, a technique that is of paramount importance in the Member States of the European Union.

The impact of a right to research will, of course, be much broader than a simple corrective of existing copyright norms. It can, and we have argued that it should, create a normative shift in copyright law towards a more paradigmatic understanding of copyright as a system of rules that is intended to promote and enable creativity and innovation for the benefit of society at large. Designing or excavating a new (fundamental) right based on specific policy considerations must inevitably shift the balance in normative systems that rely on fundamental rights as guiding norms.

We have demonstrated that this shift would change copyright, interpreted and understood in light of an ‘upgraded’ fundamental rights regime, into a right that permits access to information if used for purposes that promote copyright’s mission. Putting copyright in the service of our societal mission and our normative goals will be the logical result of a right-to-research-infused copyright regime.

In the coming decades, as the last years have demonstrated, the importance of research for human development, and to a clear extent for human survival, will depend on continuous and intensive research to face and master the challenges we will face as a global society in a physically and digitally interconnected world. A right to research can help to avoid that a legal institution designed to enhance progress and access to science and culture – copyright – does not stand in the way of the best possible sustainable development of our global society.

#### Scientific research provides resilience to a range of existential risks

Matthews 18 (Dylan Matthews, Senior Correspondent and Lead Writer for vox. “How to help people millions of years from now” 10/26/2018. Accessed 5/28/2024. https://www.vox.com/future-perfect/2018/10/26/18023366/far-future-effective-altruism-existential-risk-doing-good) wtk

But in a set of slides he made in 2013, Beckstead makes a compelling case that while that’s certainly part of what caring about the far future entails, approaches that address specific threats to humanity (which he calls “targeted” approaches to the far future) have to complement “broad” approaches, where instead of trying to predict what’s going to kill us all, you just generally try to keep civilization running as best it can, so that it is, as a whole, well-equipped to deal with potential extinction events in the future, not just in 2030 or 2040 but in 3500 or 95000 or even 37 million.

In other words, caring about the far future doesn’t mean just paying attention to low-probability risks of total annihilation; it also means acting on pressing needs now.

For example: We’re going to be better prepared to prevent extinction from AI or a supervirus or global warming if society as a whole makes a lot of scientific progress. And a significant bottleneck there is that the vast majority of humanity doesn’t get high-enough-quality education to engage in scientific research, if they want to, which reduces the odds that we have enough trained scientists to come up with the breakthroughs we need as a civilization to survive and thrive.

### 2nc/1nr uq— they say “science research declining”

#### Scientific research is increasing across the board

Ofer et al. 24 (Dan Ofer, Hadasah Kaufman, and Michal Linial, all from the Department of Biological Chemistry, Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel. “What's next? Forecasting scientific research trends” *Heliyon* Volume 10, Issue 1. 1/15/2024. Accessed 5/29/2024. https://www.sciencedirect.com/science/article/pii/S2405844023109893) wtk

The number of scientific papers published has been accelerating for at least four decades, and citation and annotation behaviors have changed with it [1]. Changes might be attributed to the continuous increase in research institutes and researchers, the increasing impact of publications for funding agencies, and academic careers [2,3]. In addition, acceleration in publication may reflect better automated data indexing (e.g., Science Citation Index (SCI) [4]), and the establishment of a keyword annotation scheme (e.g., the medical subject headings system, MeSH [5]). Other changes within the last four decades include the expansion of open access policies, increased research originating in industry, and the ongoing increase in the total number of researchers and expected research productivity output in many fields. While presenting the current state can be based on a historical view of analytical methods, predicting future trends is far more challenging [6,7].

#### Science is getting more disruptive—their studies rely on flawed methodologies

Eckhardt and Church 23 (Juergen Eckhardt is a senior vice president and head of Leaps by Bayer, the impact investment unit of Bayer AG. George Church is a professor of genetics and of health sciences and technology at Harvard Medical School and the Massachusetts Institute of Technology. “Disruptive science is leaping forward, not limping along” 2/1/2023. Accessed 5/29/2024. https://www.statnews.com/2023/02/01/disruptive-innovation-science-leaping-forward/) wtk

The negative overall trend may partially stem from an increasing denominator of “non-disruptive papers,” given that more than 3,000 papers are now published per day — a staggering increase since the 1940s — driving down the “disruptiveness index.” Yet total innovation and impact may still be increasing. Furthermore, the paper’s analysis also stops at 2010, missing the last crucial decade of rapid progress.

Here’s the big question: Does this analysis really represent a decline in radical innovation that improves fundamental understandings and technologies?

The example of the “most disruptive” innovation in the last half century the authors cite is a 1983 patent on a better way to get DNA into cells. But moving DNA into cells was not at all surprising in 1983 with famous precedents in 1944 and 1978. It was soon largely replaced by electroporation, cationic lipids, viral capsids, and microinjection. Meanwhile, the “least disruptive” innovation in the Nature analysis is David Baltimore’s 1970 paper on reverse transcriptase, an enzyme that catalyzes the formation of DNA from an RNA template. At the time, going from RNA to DNA was conceptually very surprising and the paper, as a bonus, provided a way to study and commercialize RNA that has not been replaced. That low-disruption paper earned its author a Nobel Prize, while the supposedly bolder example did not.

Consider the incremental advances of today, like the mRNA-based Covid-19 vaccines. The authors of the Nature paper view these vaccines not as a genuine breakthrough but as the result of decades of stepwise research, from the discovery of mRNA itself to viral genome mapping and to the development of the lipid nanoparticles that stabilize the fragile molecule for delivery into the human body. But given that these vaccines saved an estimated 20 million lives in a year, we believe it’s fair to call their arrival in record time truly groundbreaking. These vaccines literally represent a break from the vaccines of the past, a new paradigm upon which vaccine developers and scientists will continue to build.

In the near future, what if new categories of cancer therapy (building on the development of CAR-T immunotherapy), break-even nuclear fusion reactions, exponential improvements in reading and editing genomes, machine learning solving the six-decade old protein design challenges, the colonization of planets, carbon sequestration, nitrogen fixation, and aging reversal therapies emerge? One could imagine that none would nudge the disruptiveness index upward because long-standing, profound problems are defined by multiple mature disciplines. That said, these achievements would rank among the most amazing achievements of humanity!

Gunther Stent declared “the end of progress” in 1969, decades before the life science revolution got underway. Similar sentiments about the end of physics breakthroughs immediately preceded quantum and relativity revolutions and many decades of breathtaking applications. Perhaps it’s time to embrace the spirit of anti-complacency and pro-innovation without hallucinating a depressing decline in the face of remarkable progress.

### 2nc/1nr link extensions

#### The aff undermines the freedom of inquiry and TDM research—that undermines scientific progress

Samberg et al 24 (Rachael G. Samberg, Timothy Vollmer, and Samantha Teremi, professionals within the Office of Scholarly Communication Services at UC Berkeley Library. “Licensing research content via agreements that authorize uses of artificial intelligence” 1/10/2024. Accessed 5/22/2024. [https://www.authorsalliance.org/2024/01/10/licensing-research-content-via-agreements-that-authorize-uses-of-artificial-intelligence/) wtk](https://www.authorsalliance.org/2024/01/10/licensing-research-content-via-agreements-that-authorize-uses-of-artificial-intelligence/)wtk) \*TDM = text and data mining

Previous court cases like Authors Guild v. HathiTrust, Authors Guild v. Google, and A.V. ex rel. Vanderhye v. iParadigms have addressed fair use in the context of TDM and confirmed that the reproduction of copyrighted works to create and conduct text and data mining on a collection of copyright-protected works is a fair use. These cases further hold that making derived data, results, abstractions, metadata, or analysis from the copyright-protected corpus available to the public is also fair use, as long as the research methodologies or data distribution processes do not re-express the underlying works to the public in a way that could supplant the market for the originals.

For the same reasons that the TDM processes constitute fair use of copyrighted works in these contexts, the training of AI tools to do that text and data mining is also fair use. This is in large part because of the same transformativeness of the purpose (under Fair Use Factor 1) and because, just like “regular” TDM that doesn’t involve AI, AI training does not reproduce or communicate the underlying copyrighted works to the public (which is essential to the determination of market supplantation for Fair Use Factor 4).

But, while AI training is no different from other TDM methodologies in terms of fair use, there is an important distinction to make between the inputs for AI training and generative AI’s outputs. The overall fair use of generative AI outputs cannot always be predicted in advance: The mechanics of generative AI models’ operations suggest that there are limited instances in which generative AI outputs could indeed be substantially similar to (and potentially infringing of) the underlying works used for training; this substantial similarity is possible typically only when a training corpus is rife with numerous copies of the same work. And a recent case filed by the New York Times addresses this potential similarity problem with generative AI outputs.

Yet, training inputs should not be conflated with outputs: The training of AI models by using copyright-protected inputs falls squarely within what courts have already determined in TDM cases to be a transformative fair use. This is especially true when that AI training is conducted for non-profit educational or research purposes, as this bolsters its status under Fair Use Factor 1, which considers both transformativeness and whether the act is undertaken for non-profit educational purposes.

Were a court to suddenly determine that training AI was not fair use, and AI training was subsequently permitted only on “safe” materials (like public domain works or works for which training permission has been granted via license), this would curtail freedom of inquiry, exacerbate bias in the nature of research questions able to be studied and the methodologies available to study them, and amplify the views of an unrepresentative set of creators given the limited types of materials available with which to conduct the studies.

#### AI is uniquely important for advancing scientific progress

Hansen and Brooke 23 (David Hansen, Executive Director of the Authors Alliance. Rachel Brooke, Senior Staff Attorney of the Authors Alliance. “RE: Policy Study on Artificial Intelligence, Docket Number 2023-6” 10/30/2023. Accessed 5/24/2024. https://www.regulations.gov/comment/COLC-2023-0006-8976) wtk

Generative AI is also a powerful tool for academic nonfiction authors. AI models are increasingly being used as part of the research and writing process for nonfiction works, such as scientific research articles and books. Like fiction authors, researchers are turning to generative AI systems like ChatGPT to ideate and to find more concise and accurate ways of writing academic nonfiction works.15 And tools like elicit.org are increasingly used to assist with literature reviews, summarizing conclusions and highlighting areas of uncertainty or disagreement within the literature.16

AI in this context allows researchers to refocus attention on the core of their work—generating new scientific discoveries, innovative new technologies, and new applications in a wide range of fields—rather than on the process of writing and communicating their research. For many of these authors, as explained by Michael Eisen, a computational biologist at the University of California, Berkeley (also former editor-in-chief of the journal eLife and a member of the Authors Alliance Advisory Board), “[i]t’s never really the goal of anybody to write papers—it’s to do science.”17

Perhaps most significantly, AI has the potential to lower barriers to participation in the scientific community by allowing more researchers to contribute. As explained by a recent review article, there are dozens of efforts underway to apply AI systems in ways that will lower language barriers in scientific communication: “[M]ost scientific papers in the world literature are written in English by non-native English speakers. Non-native English-speaking scientists face many difficulties in writing clearly, succinctly, and without grammatical errors.”18

Generative AI can also support the progress of science as part of the research process itself. It can achieve this by speeding up data processing, automating data acquisition, making it easier and faster to write code, generating new research hypotheses, and more.19 Other commenters’ submissions, particularly the submission from the UC Berkeley Library, explain in more detail how generative AI (and other AI technology) is already being used to aid researchers by allowing them to more efficiently and effectively explore large corpora of data in fields ranging from computer science to medicine.

### 2nc/1nr link—they say: “commercial doesn’t spillover”

#### Commercial exclusions undermine research as a whole

Fiil-Flynn et al. 22 (Sean M. Fiil-Flynn, Program on Information Justice and Intellectual Property, Project on the Right to Research and International Copyright, American University Washington College of Law, Washington, DC, USA. Et. al. “Legal reform to enhance global text and data mining research” *Science* Volume 278, No. 6623. 12/1/2022. Accessed 5/22/2024. https://www.science.org/doi/10.1126/science.add6124) wtk

The extension of TDM exceptions to commercial uses may be controversial. On one hand, many commercial users might be capable of paying licensing fees and other transaction costs, and copyright exceptions that simply transfer wealth from copyright owners to commercial TDM users, might seem arbitrary and unjustified. On the other hand, many socially beneficial uses of TDM—including the BlueDot program that originally tracked COVID-19 (1) or internet search engines that copy and mine the entire internet (3)—would likely not exist if commercial uses were excluded from copyright exceptions. Some countries see commercial TDM as a way to invest in domestic innovation and technology transfer. The EU recently adopted a rule that, although not fully tested, permits copyright holders to opt out of commercial (but not “scientific” or “cultural”) TDM uses.

### 2nc/1nr link—they say: “licensing solves”

#### Copyright protections create a cost barrier to scientific research

Palmedo 19 (Michael Palmedo, MA Economics, MA International Affairs, Ph.D Candidate, American University, Department of Economics. “The Impact of Copyright Exceptions for Researchers on Scholarly Output” 6/1/2019. Accessed 5/23/2024. [https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3090022) wtk](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3090022)wtk)

Copyright law incentivizes the creation of new works, but it does so by limiting access to existing works to those who cannot afford it. In the modern world of academic publishing, high book and journal prices present barriers to researchers who wish to build on previous knowledge. One possible solution to this problem is for countries to broaden copyright exceptions that allow scholars to access works for the purpose of conducting further research.

This paper has demonstrated that scientists residing in countries which have implemented more robust copyright exceptions for research published approximately 17- 22% more papers and books in subsequent years. It has found some evidence that the effect is greater in countries where copyright protection of existing works is stronger, and that the effect is greater in poorer countries, though the evidence regarding these interactions is less robust.

#### Broad fair-use exception to copyright protections is key—empirics prove

Fiil-Flynn et al. 22 (Sean M. Fiil-Flynn, Program on Information Justice and Intellectual Property, Project on the Right to Research and International Copyright, American University Washington College of Law, Washington, DC, USA. Et. al. “Legal reform to enhance global text and data mining research” *Science* Volume 278, No. 6623. 12/1/2022. Accessed 5/22/2024. https://www.science.org/doi/10.1126/science.add6124) wtk

The rights granted by copyright are not absolute. All international copyright treaties permit, and all countries have, exceptions from copyright protection for various purposes, some of which may authorize TDM research. In the US, for example, a flexible exception exists for “fair use” for purposes such as education and research and has been interpreted by courts to permit at least some TDM uses. Copyright laws in many other countries contain exceptions for research (or “scientific”) uses that can be interpreted to apply to TDM uses (4). But only about a fifth of these research exceptions are broad enough to permit the full range of TDM research, which requires the ability to copy, share, and analyze whole works in collaboration with others (8) (see the figure and table). For example, some countries have research exceptions that permit uses only of excerpts of a work (e.g., Argentina), do not apply to uses of books or other kinds of works (e.g., most post-Soviet countries), or require membership in a specific research institute (e.g., Sweden).

Empirical studies show that copyright exceptions for research matter—with correlations between more permissive research exceptions and higher production of citable works of scholarship (9) and increased academic use of TDM methodologies (10). But until legal enabling environments for TDM research can be harmonized, the full benefits from this new research frontier will remain inadequately explored.

#### TDM exceptions are uniquely important—research proves

Geiger and Jütte 22 (Christophe Geiger is Professor of law at the Luiss Guido Carli University in Rome (Italy) and President of the International Association for the Advancement of Teaching and Research in Intellectual Property (ATRIP). Bernd Justin Jütte is Assistant Professor in Intellectual Property Law, University College Dublin, Sutherland School of Law (Ireland) and Senior Researcher Vytautas Magnus University, Faculty of Law (Kaunas, Lithuania). “Conceptualizing a ‘Right to Research’ and its Implications for Copyright Law: An International and European Perspective” Joint PIJIP/TLS Research Paper Series, July 2022. Accessed 5/28/2024. https://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=1079&context=research) wtk

A broader research exception must be technologically neutral to enable research with different types of data and with different methodologies. It must extend to a variety of uses of protected works and other subject matter to allow for the progressive of science and technology to achieve the goals set out in Article 3 TEU and in particular those that relate to the realization of the internal market. A broad research exception should incentivize risk-taking, but these risks should be confined to potential outcomes of research endeavours and not in the form of potential legal liability. In other words, a broadly expressed research exception must enable researchers to work with data and information to contribute to scientific and technological advancement to work toward a more sustainable and socially just internal market in the EU, an eventually for global societies. And, of course, the idea of a broad research exception should eventually not be confined to the EU’s internal market but should work to the benefits of researchers that cooperate with counterparts in the EU, or individually or in groups around the globe.

4. Specific TDM Exception

The recent introduction of a text and data-mining (TDM) exception in Articles 3 and 4 of the CDSM Directive remedies some of the shortcomings of the general research exception under Article 5(3)(a) InfoSoc Directive. It was indeed, introduced to “benefit the research community and, in so doing, support innovation.”251 However, the new exception itself is not without flaws and underlines some of the existing systemic deficiencies of copyright law252. In general, a specific exception for TDM should be welcomed as it clarifies the lawfulness of TDM in principle. However, the definition of the scope raises further problems, but also shows the way forward.

### 2nc/1nr i/l—they say “research not key to science”

#### Publication is key to truth testing and shared knowledge—that ensure broad scientific progress

Foster 23 (Jamie Foster, PhD in Physics from the Massachusetts Institute of Technology, “The Critical Importance Of Publishing Results In Science” 11/5/2023. Accessed 5/29/2024. https://www.jamiefosterscience.com/why-is-publishing-results-important-in-science/) wtk

Publishing research results is a fundamental part of the scientific process. Without the dissemination of findings through peer-reviewed journals, conferences, and other outlets, science would fail to progress and build on itself.

In this comprehensive article, we’ll examine why sharing results publicly is so vital to the enterprise of science.

If you’re short on time, here’s a quick answer: Publishing results enables scrutiny from the scientific community, sparks new discoveries, prevents duplication of effort, rewards researchers for their work, and allows findings to benefit society.

Enables Scrutiny and Verification

Publishing results in science is of critical importance as it allows for thorough scrutiny and verification of research findings. The peer review process plays a crucial role in this regard, as it involves the evaluation of a study by experts in the field before it is accepted for publication.

This rigorous evaluation helps to ensure that the study meets certain standards of quality and methodology.

Peer review process

During the peer review process, reviewers carefully examine the research methods, data analysis, and interpretation of results. They assess the study’s validity, reliability, and overall contribution to the scientific community.

This step helps to identify any potential errors or limitations in the research, ensuring that only reliable and accurate findings are published.

Moreover, peer review also provides valuable feedback to the researchers, allowing them to improve their work and address any potential weaknesses. This collaborative process helps to maintain the integrity and credibility of scientific research.

Spotting errors and limitations

By publishing their results, scientists open up their findings to a wider audience, including fellow researchers, scholars, and experts in the field. This broader scrutiny increases the chances of errors or limitations being identified.

With more eyes on the research, potential flaws in methodology, data analysis, or interpretation can be spotted and addressed.

Additionally, publishing results allows for constructive criticism and discussions, which can lead to further improvements and advancements in the field. It encourages researchers to be thorough, transparent, and accountable for their work, ultimately benefiting the scientific community as a whole.

Confirming reproducibility

Another crucial aspect of publishing results is the ability to confirm reproducibility. Reproducibility is the cornerstone of scientific research, as it ensures that findings can be independently verified and validated by other researchers.

When results are published, it allows other scientists to conduct their own experiments and attempt to reproduce the findings.

This process not only adds to the body of knowledge but also helps to identify any anomalies or discrepancies in the original study. Reproducibility strengthens the reliability and validity of scientific research, fostering confidence in the scientific community and among the general public.

Allows Theories and Knowledge to Accumulate

Publishing results in science is of critical importance as it allows theories and knowledge to accumulate over time. By sharing their findings with the scientific community and the wider public, researchers contribute to the collective body of knowledge and pave the way for future advancements.

Building on previous findings

When scientists publish their results, they provide a foundation for others to build upon. By sharing their methodologies, data, and conclusions, researchers allow others to replicate their experiments and validate their findings.

This replication is crucial in science as it helps to confirm the reliability and reproducibility of results. Additionally, by publishing their results, scientists enable others to expand upon their work and explore new avenues of research.

Sparking new hypotheses

Published results often spark new hypotheses and avenues of exploration. When researchers share their findings, they provide valuable information that can inspire other scientists to ask new questions and develop new theories.

This collaborative nature of science allows for the continuous growth and development of knowledge.

Progress depends on shared knowledge

Scientific progress relies heavily on the sharing of knowledge. By publishing their results, scientists contribute to the collective understanding of the scientific community and enable others to build upon their work.

This sharing of knowledge is essential for the development of new technologies, treatments, and solutions to global challenges.

#### Increasing publishable research is key to scientific progress

Robens 22 (Jeffrey Robens, PhD in Pharmacology from the University of Pennsylvania and Senior Editorial Development Manager, Nature. “The Importance of Academic Publishing and the Open Access Evolution” 2/23/2022. Accessed 5/29/2024 https://www.aje.com/arc/the-importance-of-academic-publishing-and-the-open-access-evolution/) wtk

It does not matter how important a researcher's results are, if no one knows about them, they'll have no impact. So, it is essential for researchers to communicate their results with others to advance their field.

What do you gain from publishing in a journal?

People often ask me if it is important where researchers publish their results. Definitely. A scientist needs to publish their findings on a platform that is discoverable by others in their field worldwide. The greater the recognition of the platform, the higher the likelihood it will be for people to find and read that research. And that is how scientists improve their impact.

However, with the ease of discoverability on the internet, is it necessary to publish research in academic journals? Can a scientist simply post their results online? That certainly seems much more efficient and easier for the author. And it is. But it is not convenient for readers.

I like to see academic journals like curators at an art museum. Imagine you want to visit art and there are two choices of art museums.

The first has no restrictions; anyone can come in and hang up their artwork. This museum has complete freedom. So, when you walk in, you'll see some high-quality art and some very poor-quality art. Further, as there is no organization, everything is mixed together. So, it would likely take you a long time and a lot of effort to find artworks of interest to you.

The second museum, on the other hand, is curated. Meaning that art professionals inspect the various pieces of art to determine those that are of high quality. Additionally, they group this art by genre or age; e.g., one room will contain Japanese art from the 20th century, while another room will contain French art from the 19th century.

Which museum would you prefer to visit?

Most people would choose the second. Why? It is simply easier. You can quickly view the art that you are interested in, and you can feel confident that the art displayed are high quality works.

That is the role of journals in academic publishing, as curators of the scientific literature. Scientists first submit their work to specific journals (whose scope is related to that topic of research), and the journal editors and peer reviewers ensure that work has the necessary quality and relevance to be published.

In this way, readers can quickly find what they are looking for and feel confident that research will be relevant and useful.

This is the value of academic publishers for the scientific record — they filter and improve what is published. Journal editors are the first line ensuring that the research is suitable for the scope of their journal. This helps to group similar research together that support each other. It also helps readers find studies that are relevant for them.

For example, climate scientists can feel confident that the research published in Nature Climate Change will likely be relevant and interesting for their own research. This helps these researchers save time by not having to search through the internet looking for research.

Regarding improvement, journal editors consult experts in the field — peer reviewers — to evaluate the studies and give recommendations on how that research can be further improved in terms of robustness and transparency. Having these additional insights can make sure that the final published article is of the highest quality and also promotes reproducibility of that research.

Highly selective journals, like those from Nature, have very high standards regarding what they publish. The result of this strict quality control is that readers can feel confident that when they pick up an issue of Nature Cell Biology (or any other Nature title), they'll be reading the latest and most important breakthroughs in their field.

#### Copyright exceptions cause follow-on innovations

García et al. 20 (Kristelia García, University of Colorado Law School. James Hicks, University of California, Berkeley, School of Law. Justin McCrary, Columbia University Law School. “Copyright and Economic Viability: Evidence from the Music Industry” Journal of Empirical Legal Studies Volume 17, Issue 4, 696–721, December 2020 Accessed 5/23/2024 [https://scholar.law.colorado.edu/faculty-articles/1307) wtk](https://scholar.law.colorado.edu/faculty-articles/1307)wtk)

The extent of both the incentive effect and the costs of copyright are, in principle, empirically testable. On the question of cost, some attention has been paid to the effect of copyright on the availability and perceived quality of creative work. Exploiting the fact that works written before 1923 are in the public domain in the United States, Heald (2008, 2014) finds that copyright status is associated with a significant increase in the proportion of books that are out of print and functionally unavailable.8 Similarly, Reimers (2019) uses a regression discontinuity to identify the effect of copyrights on the price and availability of literature, and concludes that copyright results in higher prices (and fewer editions) for popular works, and lesser availability among low- and mediumquality works. Reimers draws a mixed picture of the welfare effects of this tradeoff, finding that, in this case, the decrease in consumer surplus from reduced access to lowerquality works was not offset by increased profits for owners of popular books.

Buccafusco and Heald (2013) see a similar result with respect to availability in the audiobook market. They also show experimentally that copyrighted audiobooks are no more likely to be perceived by listeners as “high quality” than audiobooks derived from public-domain works, somewhat contrary to theoretical predictions that copyright owners will have an incentive to be better stewards of older works. However, the picture is complicated: in the musical context, MacGarvie et al. (2018) show that the expiry of copyright leads to a doubling of rereleases of the underlying work but has no effect on the availability of the original track on Spotify’s streaming platform.

In the case of follow-on creation, new evidence from economic history suggests that copyright can limit the circulation of knowledge in the scientific community (Biasi & Moser forthcoming). Similarly, in the music industry, Didwania (2018) finds evidence that copyrighted musical works are sampled at less than half the rate of comparable public-domain tracks.

### 2nc/1nr impact—they say “science doesn’t prevent extinction”

#### Scientific research solves extinction through innovations and collective knowledge

Siegel 17 (Ethan Siegel, Ph.D. astrophysicist, author, and science communicator, “Humanity Needs Science To Survive And Thrive” 7/25/2017. Accessed 5/28/2024. [https://www.forbes.com/sites/startswithabang/2017/07/25/humanity-needs-science-to-survive-and-thrive/?sh=90df5d2e5a22) wtk](https://www.forbes.com/sites/startswithabang/2017/07/25/humanity-needs-science-to-survive-and-thrive/?sh=90df5d2e5a22)wtk)

Humanity is special because we have the capabilities to gather all the data related to these issues, synthesize them together into one overarching framework, address each problem both individually and as a part of our larger society, and plan for our continued thriving existence here on Earth. Over the thousands and thousands of years of recorded history, the one path that's been consistently successful has been to take a scientific approach.

That means a lot of hard work. It means we have to create a body of knowledge that stretches across an entire field, encompassing as many options and corner-cases as possible. It means that we need to observe and take note of what happens under all sorts of different circumstances, often when many variables are in play at the same time. It means we need to quantify and understand our uncertainties, errors, and biases in performing our studies. It means we need to focus on one small problem at a time in order to learn something meaningful. And perhaps most challenging of all: we need to always be open to the possibility that our best laws, conclusions, theories, and models may be wrong. As we learn more, we must always go where the full suite of data takes us, even if it requires us to revise what we previously thought we knew.

The beauty of it all is, therefore, what makes it so disconcerting to some. You can spend your entire life studying a problem, fascinated with the process of investigation and discovery. You can learn everything we know about a field and extend our knowledge within it. And you might have a beautiful idea, a preferred outcome or a conclusion you wish were true as respects your studies. You may even put together a new theory that has the potential to revolutionize everything we think we know. But no matter how elegant, beautiful, compelling or intuitive it is, your conclusions must be consistent with the entirety of the information we have. The most successful ideas in science aren't successful because we love them the most; they're successful because the evidence overwhelmingly validates and supports them.

Not all of science is about feeding the world, curing and preventing disease, preventing short-and-long-term threats, protecting our environment or extending our lifespans. Much of it is about looking at esoteric systems, at other creatures, at subatomic particles or at the Universe beyond our world. It's an enterprise that's driven by curiosity, and for the search for scientific understanding. It challenges us to find truths that go beyond politics, beyond ideology, beyond our desired outcomes. It dares us to know the Universe as it actually is, and the way we do that is by asking it questions about itself and then listening to the answers, no matter what they are.

Science is what's led our society to the present day, where food is plentiful, abundant and safe. Where diseases can be treated, cured or even prevented outright before you ever get sick. Where we can quantify the threat that dirty air, contaminated water or a hole in the ozone layer has on humanity. And where new advances lead to new technologies, enhancing our quality of life to levels that humans, even just a hundred years ago, couldn't possibly have foreseen. If we want this to continue, we absolutely need to listen to, accept, and value what science has to offer in all of these regards.

Society didn't get to where it is by focusing on dissenting scientific voices; it advanced by valuing and listening to what was validated and known. We didn't advance by delaying until we were 99.99999% sure we had it right; we took the best information we had and acted. And we didn't invest in the future by cutting funding for the enterprise that makes future advances possible; we chose a boldly ambitious project and saw it through, 100% of the way. It's up to humanity to write our own future, with science providing the ink in our pens. If we stop listening, stop investing, and stop valuing the lessons it has to teach us, it's only a matter of time before we suffer the same fate as an untended yeast colony. We can be better than that, but only if we decide to choose and value science.

#### Maximizing available research is key to deployment of important scientific principles

Vasilik 22 (Taylor, Popular Science, “Why Science Research Is Imperative For Human Survival” 12/6/2022. Accessed 5/28/2024. https://www.popularmechanics.com/science/math/a41409095/neil-degrasse-tyson-exponential-concepts/?scrlybrkr=04ba8ef9) wtk

Although nothing about our blip in time is truly unique, that doesn’t mean we should stop asking questions. In fact, building on past research by asking more nuanced questions—and searching for their answers—is key to long-term human survival.

So Much Damn Research

Tyson’s post-doc research began in 1995 at Princeton University. There, he studied the doubling time of research published from 1895 to 1995 in The Astrophysical Journal, a peer-reviewed scientific journal covering advances in astrophysics and astronomy. Tyson found that in any 15-year period, there was more research published in the journal than in the entire history of the journal, meaning the rate of publishing had exponentially grown (exponential growth means the rate of growth becomes increasingly rapid in proportion to the growing total). He found that this rate was about the same across all branches of sciences.

Humans are programmed to think linearly, but knowledge grows exponentially. So, how do we wrap our mind around exponential concepts?

One way is to consider the exponential growth of United States patents in the 21st century: between 2010 and 2020, the number of patents granted was greater than the number of patents granted between 1960 and 2000.

Tyson writes in his new book, Starry Messenger: Cosmic Perspectives on Civilization, that “with systems in place to disseminate thought, such as scientific conferences, peer-reviewed journals … every next generation can use discoveries of the previous generation as fresh starting points.” Einstein’s Theory of Relativity allowed for the discovery of GPS navigation, for instance. If Einstein hadn’t questioned the composition of our universe, we’d still be using paper maps to get from point A to point B instead of taking an Uber.

To Publish or To Perish

The long-standing tradition of published scientific research means something different as you progress in your academic career. For a graduate student, your research must be publishable, so that you stand out to employers. For established scientists publishing new research, your latest work must be as good as your last published work. In each of these situations, scientists are pressured to publish in order to succeed in their career, or perish without enough published work to stand out.

One of the most detrimental effects of publish-or-perish culture is what some may know as “salami slicing.” Researchers will purposely slice-and-dice their research into multiple publications, rather than as a whole experiment with all findings concluded. This causes an array of issues, such as a reader’s hesitance to trust your research, resulting in a diluted publication with less impact.

This has the potential to mess with the integrity of scientific research, including budget pressures and publication biases. On the flip side, there are more collaborations between researchers, creating a vast network of scientists that can work together to both answer and explore new questions.

How To Prevent Mass Extinction

To understand why science research is critical to human survival, let’s consider the threat of asteroids.

NASA launched the test mission known as DART (Double Asteroid Redirection Test) in 2021. DART intentionally smashed into a small asteroid named Dimorphous just a few months ago, providing important data on methods to prevent planetary disaster. With an already extensive history of astronomy and astrophysics, and the exponential growth of knowledge today, a technologically savvy generation of thinkers and scientists were able to find answers to questions about asteroids that we would otherwise never even know to ask.

## opt-out counterplan

### 1nc opt-out cp

#### The next offcase position is the opt-out counterplan

#### The United States federal government should implement streamlined notice and action procedures, requiring AI developers to promptly and in good faith take action to prevent copyright-infringing content from being generated by its system again.

#### The cp creates an opt-out process that protects creators from having their works used to create identical or substantially similar works by AI—that solves

Pasquale and Sun 24 (Frank A. Pasquale, Professor of Law at Cornell Tech and Cornell Law School. Haochen Sun, Associate Professor of Law and Director of the Law and Technology Center at the University of Hong Kong Faculty of Law. “Consent and Compensation: Resolving Generative AI’s Copyright Crisis” Cornell Legal Studies Research Paper Forthcoming, 5/1/2024. Accessed 6/10/2024. Available at SSRN: http://dx.doi.org/10.2139/ssrn.4826695) wtk

Regulators must reinstate creatives’ proprietary control of their works within an equitable knowledge ecosystem. A new opt-out mechanism for copyright owners would empower them to reclaim proprietary control of their works through streamlined “notice and action” procedures aimed at AI providers. This mechanism would allow authors to submit requests to such providers for the removal of their works from the datasets of relevant AI systems, and take additional actions, as described in Sections A and B below.

For ease of reference, we will assume for the rest of this article that the authors of given works discussed also own the copyright in those works. To be sure, copyrights are often bought by publishers, recording companies, and other firms, or accrue to institutions in workmade-for-hire scenarios. In such cases, our proposal is agnostic as to whether authors, copyright owners, or both should be able to deny consent to use their works in AI. This is a detail that would need to be worked out in a legislative process, perhaps with some reference to past debates on the proper extent and scope of moral rights, since many creatives will have ethical and cultural objections to works they created being used in AI training or reproduced by AI.

A. NOTICE AND ACTION PROCEDURES

Under the proposed mechanism, copyright owners can first request AI providers to take actions to effectively prevent their systems from generating outputs that appear identical or substantially similar to relevant copyrighted works. A copyright owner would be entitled to send a notice to an AI provider when he or she identifies that an output generated by the provider’s AI system contains either a verbatim or substantially similar copy of his or her work, or a derivative work. In the notice, the copyright owner would be obliged to document the unauthorized reproduction of the work and his or her copyright ownership, along with a digital copy or an online link to the work.

The notice would target AI-generated content that resembles or adapts the copyright owner’s work, potentially infringing upon the author’s right to reproduction74 or right to prepare derivative works. 75 For example, an exact replica of a copyrighted image or video generated by an AI system is highly likely to infringe on the right of reproduction. At the same time, certain adaptations of works by AI systems could also infringe on the right to prepare derivative works. 76 Such adaptations might include an image generator creating a painting based on a photograph, a chatbot condensing a novel into a novella, and a sound generator composing or performing a soundtrack derived from preexisting songs.

Upon receiving the notice, if the AI provider concerned accepts that the author has established a prima facie case of copyright infringement, it must promptly take actions to prevent such infringing content from being generated by its system again.77 These actions may include: (1) removing the author’s work from the datasets used by its AI system; (2) embedding filtering technology into the AI system to prevent generation of similar content that would infringe the author’s copyright again; 78 and (3) initiating a “machine unlearning” process to remove the influence of the author’s work from the AI system.79 Upon completion, the AI provider should inform the author of the actions taken and provide an appropriate explanation of the effects of such actions.

Under the proposed mechanism, AI providers are obligated to take relevant actions expeditiously in response to notices submitted by authors. 80 For more complex tasks, such as machine unlearning, providers should be granted additional time, as long as they take action in good faith. The proposed mechanism would initially impose monetary penalties on AI providers if they fail to promptly reply to legally valid notices submitted by copyright owners.

### 2nc/1nr solvency extensions

#### The cp solves sufficiently—it empowers authors to opt-out and forces disclosure of information by ai companies

Pasquale and Sun 24 (Frank A. Pasquale, Professor of Law at Cornell Tech and Cornell Law School. Haochen Sun, Associate Professor of Law and Director of the Law and Technology Center at the University of Hong Kong Faculty of Law. “Consent and Compensation: Resolving Generative AI’s Copyright Crisis” Cornell Legal Studies Research Paper Forthcoming, 5/1/2024. Accessed 6/10/2024. Available at SSRN: https://ssrn.com/abstract=4826695 or http://dx.doi.org/10.2139/ssrn.4826695) wtk

B. NORMATIVE RATIONALES FOR THE PROPOSED OPT-OUT MECHANISM

The proposed mechanism would effectively empower authors to opt out of AI systems that generate content infringing on their copyrights. As a result, it would enhance authors’ control over their works, enabling them to better protect their interests amidst the surge of copyright infringement facilitated by AI systems. Even if direct copying of copyrighted works by AI systems occurs in a small percentage of cases, it may still have great impact given AI systems’ vast output. One study of Stable Diffusion found that its models copied from its training data approximately 1.88% of the time.88 Given that AI is estimated to have generated over 150 billion images in a single year,89 even such a small percentage will generate myriad infringing images. The chance that the output of AI systems would infringe on at least some copyrighted content is high. 90 In response, the opt-out mechanism aims to minimize the impacts of infringing activities on authors’ interests in multiple ways.

First, the proposed mechanism draws on methods developed in a long-standing copyright regime governing the use of content online, while adapting them to the age of AI. The DMCA established a robust notice and takedown process online, enabling authors to swiftly remove copyright-infringing content from platforms. 91 Copyright owners aggrieved by AI providers’ infringing outputs have, at present, no recourse to such procedures, since AI creates works, rather than hosting and arranging them in the manner of DMCA-covered online service providers like YouTube or Facebook. 92 The DMCA offers protections to platforms that host infringing user-generated content, on the rationale that the platform cannot preemptively police users’ actions.93 AI content, by contrast, is being created by the AI firm itself, so it is clearly responsible for it under the proposed mechanism.

Second, the mechanism would serve an information-forcing function, empowering copyright owners to address infringing activities perpetrated by opaque sociotechnical systems utilizing AI. Copyright owners need a new mechanism allowing them to compel AI providers to disclose information about how their works are used in training models and generating content. Any AI provider wishing to contest a notice from the author would need to provide explanations about its methods, such as whether its datasets contain the copyrighted work in question, and the workings of technologies concerning content removal, filtering, and unlearning. Thus, this new mechanism would enable authors to regain proprietary control over how their works are used in in AI systems.

Third, the proposed mechanism would also address major problems with existing opt-out procedures offered by some AI providers. For example, OpenAI purports to provide creatives with an option to avoid incorporating their creations among the photos, paintings, and other visual items that its AI systems, such as DALL-E, utilize for training and subsequent image generation. However, many creatives claim that AI providers’ opt-out processes are burdensome and complex.94 Authors have lamented that such opt-out procedures are “a bad joke” and “a fake PR stunt” for AI providers. 95 Self-regulation will not be effective here.

#### Opt-out solves and is key to sustainable ai development

Sternburg et al. 23 (Ali Sternburg, Vice President, Information Policy; Josh Landau, Senior Counsel, Innovation Policy; Erin Sakalis, Law Clerk; all for the Computer & Communications Industry Association (CCIA). “Artificial Intelligence and Copyright: Comments of the Computer & Communications Industry Association (CCIA), 10/30/2023. Accessed 5/25/2023. Downloaded from https://www.regulations.gov/comment/COLC-2023-0006-8740) wtk

Because ingestion is a fair use, no affirmative consent is required by law. However, copyright owners who wish to may have effective means of opting out of allowing their works as training materials. For example, some may be able to put their content behind technological protection measures such as paywalls, which are legally protected by 17 U.S.C. § 1201. Copyright holders making content available on the web are able to use the widely-used robots.txt exclusion protocol to prevent the work posted to their websites from being crawled by specific AI bots.

Technical experts and standard setting organizations such as the World Wide Web Consortium (W3C) or the Internet Engineering Task Force (IETF) could work to develop an exclusion protocol with more granularity that would permit search engine bots but exclude other bots, or would permit a bot to ingest data from a site for some uses but not others. Some major AI developers are already beginning work on such a standard. Several companies have recently announced extensions that will allow website publishers to allow search bots but exclude AI training bots,10 though a more universal approach that does not rely on identifying and excluding specific bots would be helpful.

Further, given the enormous scale required for LLM creation, an opt-in regime is effectively going to block AI development. It will also likely have negative impacts on equity. While obtaining permission from, e.g., songwriters may be viable through existing collective licensing groups, training data created in less common languages or from various subcultures is far less likely to be organized and the appropriate entity to contact for permission may even be impossible to determine.

#### Opt-out policies effectively manages risk of ai development

Mendes and Stray 24 (Laura Schertel Mendes, Professor of Law at the Brazilian Institute for Development, Education and Research (IDP), and Rapporteur of the Jurists Commission advising the Brazilian Senate on AI regulation, Brazil. Jonathan Stray, Senior Scientist, UC Berkeley Center for HumanCompatible AI. Co-Charis of the Working Group on Artificial Intelligence and its Implications for the Information and Communication Space. “AI as a Public Good: Ensuring Democratic Control of AI in the Information Space” February 2024. Accessed 6/16/2024. [https://www.observacom.org/wp-content/uploads/2024/02/InformeIA.pdf) wtk](https://www.observacom.org/wp-content/uploads/2024/02/InformeIA.pdf)wtk)

AI companies and entities can proactively build more inclusive and trustworthy AI systems by:

* Setting up an inclusive and participatory process that at a minimum includes equitable, sustained and substantive participation by independent researchers, civil society, and affected communities, to determine the rules and criteria guiding dataset provenance and curation, human labeling for AI training, alignment, and red-teaming. These rules and criteria should be publicly available.
* Implementing risk mitigation and output moderation measures, including notice and action mechanisms for capturing user feedback, complaint mechanisms, collaboration with trusted flaggers and fact-checkers, and continuous red-teaming to tackle identified risks.
* Providing users with an easy and user-friendly opportunity to choose alternative recommender systems that do not optimize for engagement but build on ranking in support of positive individual and societal outcomes, such as reliable information, bridging content or diversity of information.
* Implementing a policy where both content and users must acquire a “right of recommendability” before getting promoted or seen in feeds. This right should be granted based on a valid cryptographic signature linked to trusted entities.
* Clearly communicating, in an easily accessible and visible way, as regards users’ inputs and interactions with an AI system, any data-processing activities, how data is stored, and whether the data is shared with third parties. Users should be able to opt out of their input data and interactions being stored and used, and opt-out should be the default setting.

### 2nc/1nr they say: “tech fails”

#### Tech exists to enable effective opt-out and machine unlearning

Pasquale and Sun 24 (Frank A. Pasquale, Professor of Law at Cornell Tech and Cornell Law School. Haochen Sun, Associate Professor of Law and Director of the Law and Technology Center at the University of Hong Kong Faculty of Law. “Consent and Compensation: Resolving Generative AI’s Copyright Crisis” Cornell Legal Studies Research Paper Forthcoming, 5/1/2024. Accessed 6/10/2024. Available at SSRN: https://ssrn.com/abstract=4826695 or http://dx.doi.org/10.2139/ssrn.4826695) wtk

AI providers may take several actions to safeguard valid interests of copyright owners as requested through the notices. Regarding the first major action that AI providers can undertake, numerous copyright owners have requested the removal of their works from AI systems’ datasets (or, when that is not possible, destroying the copy of the dataset including their works). 81 Going forward, technology can assist here: online platforms have already employed copyright filtering technology to detect infringing content and prevent it from being uploaded. Similarly, AI providers have developed and implemented filtering technologies, such as Microsoft’s Copilot and OpenAI’s Copyright Shield, to minimize instances of copyright infringement caused by their systems’ generated content.82 Some AI providers have devised innovative methods, enabling their models to selectively “unlearn” specific information. For example, by replacing particular content in the model’s dataset with generic data, Microsoft researchers have successfully made the model forget details related to Harry Potter.83

Hence, copyright owners may complement their removal requests with demands that AI providers take additional actions. They could ask AI providers to adjust the operation of the filtering technology to prevent the generation of copyright-infringing content. If AI providers are able to develop and apply machine unlearning technology, copyright owners may request them to utilize it to make their AI models “forget” authors’ works.

### 2nc/1nr they say: “no detection”

#### Detection is feasible through multiple different methods

Pasquale and Sun 24 (Frank A. Pasquale, Professor of Law at Cornell Tech and Cornell Law School. Haochen Sun, Associate Professor of Law and Director of the Law and Technology Center at the University of Hong Kong Faculty of Law. “Consent and Compensation: Resolving Generative AI’s Copyright Crisis” Cornell Legal Studies Research Paper Forthcoming, 5/1/2024. Accessed 6/10/2024. Available at SSRN: https://ssrn.com/abstract=4826695 or http://dx.doi.org/10.2139/ssrn.4826695) wtk

Though AI providers’ data and methods are often secret, copyright owners have several options for detecting AI-generated content that infringes their copyrights. A straightforward method is for copyright owners to test an AI system themselves. For example, to determine whether an AI system allows users to create exact replicas, authors can input prompts such as “make an exact copy of X.” 84 Copyright owners may also come across infringing content given marketing of AI system capabilities or programmed disclosures of provenance. 85 Furthermore, the application of watermarks to AIgenerated content can also facilitate copyright owners’ detection of infringing activities. Such watermarks may indicate the AI-generated nature of an output and the specific system that generated it.86

### 2nc/1nr they say: “permutation: do the counterplan”

#### The perm severs—the plan requires licensing, the counterplan does not—reject the permutation because it makes the aff a moving target

### 2nc/1nr they say: “permutation: do both”

#### The perm links to the net benefit—it still limits fair use, which links to the DA.

#### Perm fails—inclusion of opt-in undermines fair use and adds logistical complexity

Garcia and Rose 23 (Nicholas P. Garcia, Policy Counsel for Public Knowledge. Meredith Filak Rose, Senior Policy Counsel for Public Knowledge. “Artificial Intelligence and Copyright Request for Information” Comments of Public Knowledge before the United States Copyright Office. 10/30/2023. Accessed 6/16/2024. https://www.regulations.gov/comment/COLC-2023-0006-9172) wtk

AI companies should develop policies for honoring creators who want to “opt out” of their work being collected for AI training purposes. Any such system should provide artists and creators the ability to affirmatively express their preference to not have their work used in AI training in advance, and should also provide systems whereby they can later object to the use of their work in training, to prevent it from being further used for AI training.

An opt in, or affirmative consent, system would present an unacceptable barrier to fair use and creativity by creating a default of restriction. Given the vast amount of existing copyrighted work, and the quantity of work needed to train AI systems, any opt in system would be practically and logistically infeasible. Conversely, an opt out system requires creators to express a definitive preference to be more restrictive, creating a default of permissiveness that promotes an overall more open creative environment.

It is critical to note that any opt out system should cover only AI training. Opt outs should not limit other forms or purposes of web crawling, web scraping, or use of the work. Functions like historical web archiving and search indexing are vital for an open internet. Any system of opt outs must recognize and accommodate potential fair uses of copyrighted work.

### 2nc/1nr they say “links to net benefit”—inflation

#### The counterplan doesn’t link to the inflation da because it doesn’t change “fair use” exceptions

Pasquale and Sun 24 (Frank A. Pasquale, Professor of Law at Cornell Tech and Cornell Law School. Haochen Sun, Associate Professor of Law and Director of the Law and Technology Center at the University of Hong Kong Faculty of Law. “Consent and Compensation: Resolving Generative AI’s Copyright Crisis” Cornell Legal Studies Research Paper Forthcoming, 5/1/2024. Accessed 6/10/2024. Available at SSRN: https://ssrn.com/abstract=4826695 or http://dx.doi.org/10.2139/ssrn.4826695) wtk

It is important to emphasize that the proposed mechanism is not designed to enable creatives to undermine the fair use privileges that may be enjoyed by AI providers. The fair use doctrine generally does not authorize the creation of new works that infringe on another’s copyright, such as by making an exact copy without transformative use.99 Moreover, the proposed mechanism does not grant authors the right to prevent AI providers from using their works for data training processes without first documenting copyright infringement arising out of content generation. When an AI provider has a good faith belief that the output it generates is a fair use of the copyrighted work, it simply needs to indicate the basis of that belief in a reply to a complaining copyright owner, to avoid the civil penalties mentioned above.

### 2nc/1nr they say “links to net benefit”—court clog

#### The cp doesn’t link to court clog but the plan does

Pasquale and Sun 24 (Frank A. Pasquale, Professor of Law at Cornell Tech and Cornell Law School. Haochen Sun, Associate Professor of Law and Director of the Law and Technology Center at the University of Hong Kong Faculty of Law. “Consent and Compensation: Resolving Generative AI’s Copyright Crisis” Cornell Legal Studies Research Paper Forthcoming, 5/1/2024. Accessed 6/10/2024. Available at SSRN: https://ssrn.com/abstract=4826695 or http://dx.doi.org/10.2139/ssrn.4826695) wtk

Last but not least, the proposed mechanism would also provide authors with a more efficient and cost-effective alternative for dispute resolution than the judicial process. Litigation is often time-consuming and expensive. According to the American Intellectual Property Law Association, litigating a single copyright infringement case in a U.S. federal court from pre-trial to appeals costs an average of $278,000 and may take over a year in many instances.96 Such litigation costs might not pose a problem for large corporations and the wealthiest creatives and content owners. For example, when The New York Times credibly threatened to sue OpenAI for using its content for data training without consent, Common Crawl removed links to The New York Times’s content from its datasets.97 However, The New York Times still felt obliged to sue both OpenAI and Microsoft a few months later.98 Many authors lack the financial resources to litigate against AI vendors (many of which are massive firms, or are backed by such firms) over potentially lengthy periods.

In contrast, an opt-out mechanism offers authors a streamlined and cost-effective way to assert their rights. AI providers would be required to promptly review the request from an author, make a decision, and notify the complainant of their decision. Failure to do so in a good faith manner should subject the firm to civil penalties. Hence, this approach would ensure that authors have an accessible and efficient means of protecting their works before resorting to potentially lengthy and costly litigation.