# AI Training Copyright Aff

### Notes

#### Background information

This affirmative says that we should strengthen intellectual property rights in copyright. In order to train large language models like Chat GPT, generative AI companies require billions of words of text that they can use as training data. To obtain that data, they are scraping the internet for any type of text- books, newspaper articles, blog posts, pictures, etc. The problem is that those works are copyrighted!

In the US, copyright automatically covers “original works of authorship” as soon as an author “fixes the work in a tangible form of expression.” That means anytime somebody writes something, or records a video of them performing some original work, or creates a work of art, it becomes their intellectual property.

So why are generative ai models allowed to train on these copyrighted materials in the status quo? Because of the **fair use exception** to copyright law. Copyright does not mean that nobody can ever use a work without permission under any circumstances. For example, if your English teacher copies a paragraph from a book and gives it to the class, that is probably not violating copyright (but if they copied the whole book, they would be). The reason for that is because US copyright law allows “fair uses” of a work without getting permission. What “fair use” means is very vague and changes on a case-by-case basis. Generative AI companies are claiming that training AI models is a fair use of copyrighted materials, so they don’t need to get permission or pay for the works.

This affirmative says that training generative AI is no longer a “fair use” and that if AI companies want to use copyrighted works to train their models, they have to get a license to use that work first. Licenses achieve two goals: permission and compensation. Gen AI companies and publishers would negotiate the terms under which copyrighted works can be used, which means they have to agree on a price. After the affirmative plan, publishers and creators would be compensated for the works that are used in generative AI training.

There are two advantages to this plan.

#### Journalism advantage

The first advantage to this affirmative is that it would help prevent the decline of journalism. This advantage says that AI is not only infringing copyrights of journalists, but it is then using their information to create a new substitute good. For example, when you go to google now, before you even click on a link, it gives you an “AI overview” that summarizes the content from articles. If that overview gives you the information you were looking for, you now no longer have to click on the actual article that the information came from. But if you don’t click on the article, then that newspaper loses out on important revenue that they need to keep doing their job!

Journalism fills an important role in democracy- it is a watchdog against corruption and a check against misinformation. Having high-quality, human-created, investigative journalism is important to ensure that the information floating around the internet is factual. Without high-quality information, our democracy could crumble.

#### Model collapse advantage

The second advantage to this affirmative is that it will prevent AI models from becoming completely useless. The way that AI is trained, it takes a large amount of text, identifies patterns in those texts, and tries to replicate those patterns. Inevitably, AI will either pick up on some things that were not intended by humans or it will *not* pick up on things that *were* intended by humans. This is relatively normal and expected.

But what happens if we use text created by Gen AI to train future generations of Gen AI? What happens is a game of **ai telephone.** If the first generation is trained on all human-written text, it will do mostly a good job, but have a few errors. If the second generation is trained on the first generation, it will pick up on those errors and magnify them. The third generation will then pick up on the magnified errors and magnify them further. That process, called **model collapse**, can continue until AI models are spouting gibberish. This increases the risk of **hallucination** (ai either making up facts or misattributing quotes/ideas to the wrong sources). It also increases the risk of **data poisoning** (when data in an ai algorithm is incorrect, leading to the AI coming to the wrong conclusion).

#### Tips, tricks, and navigating the file

For the 1ac, there are a few cards that are asterisked. If you are struggling to get through the whole 1ac in 8 minutes, try taking out some of those asterisked cards. Once you practice enough and can get through them all, you can add those asterisked cards back to the 1ac.

For the 2ac on the case, there are cards to respond to every negative case argument. **You do not have to read a card to respond to every negative case attack.** In fact, the most helpful thing you could do is **explain the 1ac evidence** and tell the judge why the 1ac evidence answers the negative’s attack. Then, if you have enough time, you could read an additional card.

There are answers to 2 offcase arguments in this file: the science research disadvantage and the opt-out counterplan. If you are negative, you can find those two arguments in the copyright case negative file.

There are answers to every other offcase argument in the novice packet for this affirmative, but they are **not in this file.** They will be at the bottom of the negative file. So for example, if the negative reads the inflation da, you can find affirmative answers at the bottom of the inflation da file.

## 1ac

### 1ac—plan

#### The United States federal government should create licensing requirements for the use of copyrighted material for training and output of commercial generative artificial intelligence.

### 1ac—journalism advantage

#### Advantage 1 is journalism:

#### Journalism is declining because of a perfect storm created by the unlicensed use of copyrighted material by gen AI companies—that undermines the lifeblood of US democracy

Blumenthal 24 (Richard Blumenthal, Chair of the Senate Judiciary Subcommittee on Privacy, Technology, and the Law, “Oversight of A.I.: The Future of Journalism” hearing of the Senate Judiciary Committee, Subcommittee on Privacy, Technology, and the Law. Committee on the Judiciary. 1/10/2024. Accessed 5/20/2024 on ProQuest Congressional via University of Michigan online library) wtk

[\*]RICHARD BLUMENTHAL: Welcome, everyone. I'm pleased to convene the subcommittee and welcome our witnesses, welcome everyone who's come to here. And of course, my colleagues from both sides of the aisle for hearing that is critical to our democracy. It's critical to the future of journalism in the United States because local reporting is the lifeblood of our democracy.

And local reporting by newspapers and broadcast stations are in existential crisis. It is, in fact, a perfect storm, the result of increasing costs, declining revenue and exploding disinformation. And a lot of the cause of this perfect storm is, in fact, technologies like artificial intelligence, not new.

Not original for me to observe it, but it is literally eating away at the lifeblood of our democracy, which as we all know, is essential to local jobs, local accountability, local awareness, and knowledge. Everything from obituaries to the planning and zoning commissions. And you can't get it anywhere else.

National news, you can buy by the yard or by the word, but local reporting is truly the result of sweat and tears and sometimes even blood of local reporters. What we're seeing and it's a stark fact about American journalism in existential crisis is the decline and potential death of local reporting. As a result of that perfect storm local papers are closing at a staggering rate.

A third of our newspapers have been lost in the last two decades. I don't need to tell anyone here that some of the oldest newspapers in the country like the Hartford Courant have closed their newsrooms, and it is a national tragedy, a painful and traumatic time for reporters, editors and their industry.

And a deep danger for our democracy. Hedge funds are buying those papers, not for the news they can communicate, but often for the real estate that they own. They are publishing weekly instead of daily. And they are buying out reporting staffs, sometimes firing them. Millions of Americans now live in a news desert where there's no local paper.

And that's especially true for rural populations and communities of color, so there's an equity aspect to this challenge as well. For any of us on this panel, we know the importance of local news, but so should people who live in those communities. Because just as local police and fire services are the first responders, local reporters are often the first informers.

And that information is no less important to them than fire and police service in the long term. The rise of Big Tech has been directly responsible for the decline in local news. And it is largely the cause of that perfect storm, accelerating and expanding the destruction of local reporting. First, Meta, Google and OpenAI are using the hard work of newspapers and authors to train their AI models without compensation or credit.

Adding insult to injury. Those models are then used to compete with newspapers and broadcast, cannibalizing readership and revenue from the journalistic institutions that generate the content in the first place. As The New York Times recent lawsuit against OpenAI and Microsoft shows, those AI models will even essentially plagiarize articles and directly permit readers to evade paywalls to access protected content free of charge.

#### Unlicensed use of copyrighted content enables mass misinformation campaigns and undermines trust in all news content

Lynch 24 (Roger Lynch, CEO of Condé Nast, “Oversight of A.I.: The Future of Journalism” Testimony before the U.S. Senate Committee on the Judiciary, Subcommittee on Privacy, Technology, and the Law. 1/10/2024. Accessed 5/17/2024 https://www.judiciary.senate.gov/download/2024-01-10-testimony-lynch) wtk

Unfortunately, current Gen Al tools have been built with stolen goods. Gen Al companies copy and display our content without permission or compensation in order to build massive commercial businesses that directly compete with us. Such use violates copyright law and threatens the continued production of high-quality media content. These companies argue that their machines are just "learning" from reading our content just as humans learn, and that no licenses are required for that. But Gen Al models do not learn like humans do. There are many examples where the Chatbots display content plainly derived from the works they ingest. In effect, they are mashing up copies at enormous scale and speed. Just as copyright law does not permit a human to replicate and regurgitate copyrighted material for purposes of commercial gain without a license, it does not allow Gen Al companies to do so either.

Moreover, Gen Al technology enables misinformation/disinformation on an unprecedented scale. In the wrong hands, Gen Al can generate outputs that are customized for individuals, making misinformation in all forms - fake photographs, audio, video and documents -look real. Widely available Gen Al tools hallucinate and generate misstatements that are sometimes attributed to real publications like ours, damaging our brands. When confronted with unprecedented amounts of misinformation, Americans can't possibly spend all of their time determining what's true and what's false. They are far more likely to stop trusting any source of information, which would have devastating consequences for our already polarized society.

In comparison, our company and companies like ours depend on customer trust. We have extensive processes to check our facts and make sure our content is accurate and fair. To ameliorate these issues, Gen Al providers should be required to disclose instances when output is generated without human review and should ensure that they do not attribute content to our brands that have not been created by us.

Luckily, there is a path forward that's good policy, good business and already the law: licensed and compensated use of publisher content for both training and output. This will ensure a sustainable and competitive ecosystem in which high quality content continues to be produced and trustworthy brands can endure, giving society and democracy the information it needs. This will also ensure that the current multiplicity of viewpoints continues to be represented in journalism without the big tech companies becoming the gatekeepers for news. Without licensed use, however, the ecosystem will collapse, and Gen Al powered search will have nothing to search for beyond poor quality content and misinformation.

More than 20 lawsuits are pending against Gen Al companies that are refusing to license content for training and output. Big tech companies are depending on the expense, delay, and uncertainty of litigation to avoid coming to the table. In fact, they have told us directly that they believe our content should be available to them for free. Congress should eliminate any doubt that licenses are required.

#### Without licensing, AI erodes trust in information, causes newspaper closure, and leads to mass job loss of journalists—that kills democracy

Coffey and Smith 23 (Danielle Coffey, President and CEO of News/Media Alliance. Regan Smith, Senior Vice President and General Counsel for News/Media Alliance. “Artificial Intelligence and Copyright” Comments of the News/Media Alliance before the U.S. Copyright Office. Docket No. 2023–6. 10/30/2023. Accessed 5/21/2024. <http://www.newsmediaalliance.org/wp-content/uploads/2023/10/Final-NMA-USCO-AI-NOI-Response-Submitted-10.30.23.pdf>) wtk

To be sustainable, however, generative AI development and use must be responsible, regulated, and accountable, with appropriate permission and compensation paid to publishers for the copying and use of their protected works throughout the product cycle. Without effective enforcement, regulation, and standards—including a requirement for AI developers to seek permission from rightsholders for uses of their protected content to train competitive products—AI can lead to considerable harms. These harms may include the undermining of the foundation of our democracy through the further weakening or outright closure of newspapers, magazines, and digital outlets—especially local ones—the spread of mis- and disinformation, and reduced access to reporting that can fundamentally only be created by humans—based on extensive fact-gathering, interviews, and judgment. An engaged and informed citizenry depends on the existence and availability of reliable and accurate reporting and analysis by outlets the public trusts. Unlike generative AI systems that may make up facts and disclaim liability for doing so,19 publishers accept responsibility for the content they publish, ensuring that the information presented to the public is of high quality. In a world flooded by easily accessible, synthetic information of unknown quality, real information becomes harder to identify and trust in our democratic system harder to upkeep.

In addition to these significant societal harms, the negative effects of unsustainable AI development practices on publishers small and large can lead to substantial job losses and a devaluing of journalistic content that will undermine these creative industries. In short, while AI presents many potential benefits to both publishers and the public at large, unregulated generative AI risks driving existing publishers out of business and disincentivizing continued investments in new, original content. This result would undermine the goal and purpose of the Copyright Clause of the Constitution, and diminish the essential role of the Press envisioned by the Founders. (And potentially also harming the further development of generative AI models through model collapse, as discussed further below.)

#### AI-driven erosion of trust in democracy in the US causes global rise in autocracy and war

Etchenique 23 (Nicolas Cimarra Etchenique was the Project on Europe's Rafael del Pino-Spanish Ministry of Foreign Affairs Fellow. He graduated in Law from Universidad Complutense of Madrid. “Democracy and the Liberal World Order Amid the Rise of Authoritarianism” 8/14/2023. Accessed 5/27/2023 https://www.belfercenter.org/publication/democracy-and-liberal-world-order-amid-rise-authoritarianism-0) wtk

Democracies must strengthen their home front to succeed in the competition with Russia and China that has resulted in a new Cold War and a new battle between democracy and autocracy. Democracies are experiencing high polarization, which is undermining democratic governance and trust in democracy itself. Democratic backsliding in the West has major geopolitical implications. It compromises the West’s international position and degrades U.S. and European soft power. Furthermore, distrust and internal divisions are exploited by autocracies. Autocracies will continue undermining the liberal world order as long as the crisis of liberal democracy is not reverted. To uphold democracy and the liberal world order, the U.S. and the EU must lead an effort to regenerate democracies’ public spheres, which are seriously damaged.

Media, and particularly social media, is playing a central role on the rise of distrust and polarization in democracies. Large for-profit digital platforms, such as Twitter (now “X”), TikTok or Facebook, are contributing to this corrosive process. Furthermore, social media has been weaponized by Russia and China to manipulate political divisions and interfere in electoral processes in democracies.

The quickly accelerating development of AI makes the situation more pressing. Autocracies are using digital technologies and AI to enhance their surveillance and repressive capacities. Within democracies, unregulated, or poorly regulated AI, may work as a Trojan horse of autocracy: it may deliver productivity gains and geopolitical advantages to the U.S., but in a way that could backfire against democracy, as it could dramatically erode social trust. Today, democracies’ public spheres can be hacked by autocratic actors using digital platforms and increasingly sophisticated AI tools. Tomorrow, if human responsibility is massively outsourced to inadequately regulated AI, trust levels in democracy could further plummet.

In order to address the crisis of democracy and preserve the liberal world order, the U.S. and the EU should cooperate to revive trust in democracy at the domestic and the international levels. Such cooperation must synchronically confront three sets of challenges:

Great power competition and the authoritarian offensive against democracy,

The rise of political polarization and illiberalism, and

The adverse impact of social media and AI on democracy.

These sets of challenges feed on distrust and amplify it. As a result, they have evolved into a series of cycles of distrust within democracies and into a great geopolitical cycle of distrust.

The entanglement and feedback loops among the domestic and the geopolitical cycles of distrust have resulted in a cohesive threat to democracy: a downward political spiral that is pulling societies towards enmity. This spiral feeds on and generates destructive human emotions at massive scale, such as outrage and hatred, that lead to violence, war, and autocracy, so it can be better understood as a dangerous global maelstrom of distrust, which could sink democracy worldwide. As showcased by historical evidence, domestic and international forces do not act in isolation from each other. Democratic backsliding, the rise of authoritarianism and totalitarianism, and the politics of aggression generated feedback loops in the 1930s, that resulted in WWII. Similar forces are again working in the 2020s.

#### Global democracy collapse causes extinction

Haydn Belfield, 2023 – Research Associate and Academic Project Manager at the University of Cambridge's Centre for the Study of Existential Risk “Collapse, Recovery, and Existential Risk,” in *How Worlds Collapse: What History, Systems, and Complexity Can Teach Us About Our Modern World and Fragile Future*, p. 74-76. Accessed online via University of Michigan //DH

A world dominated by totalitarian states would be more incompetent, more war-prone, less cooperative, and more inhibitive of progress than one dominated by democratic states. Our current world is not particularly competent, peaceful, cooperative, or progressive—a totalitarian-dominated world would be worse. It would increase the risk of another collapse and extinction and could shape the future toward less desirable trajectories (Beckstead, 2013).

Totalitarian states are incompetent. They are bad at forecasting and dealing with disasters (Caplan, 2008).16 This can be seen most clearly in the great famines of Communist China and the USSR, in which millions died (Applebaum, 2017; Becker, 1996; Dikötter, 2010; Snyder, 2010). In comparison, functioning multiparty democracies rarely, if ever, experience famines (Sen, 2010). “Established autocracies” (or “personal”/“sultanist”) are particularly bad, as there are few checks or restraints on arbitrary rule and the whims and ideology of the single individual, even from other elites (Svolik, 2012). From the inside, the “inner circle” around Mao, Stalin, and Hitler seems incredibly chaotic, with elites strongly incentivized to conceal information and encouraged by the autocrat to squabble and feud—so they are divided (Conquest, 1992; Kershaw, 2008; Zhang & Halliday, 2006). If totalitarian states are worse at addressing social, environmental, and technological problems, then a world dominated by them would likely be worse at responding to risks of collapse and extinction.

A world dominated by totalitarian states is more likely to have major wars. States with near-universal adult suffrage rarely (if ever) go to war with one another (Barnhart et al., 2020), so a world dominated by democracies has fewer wars. Miscalculation might be a particular problem for totalitarian states due to personalization and disincentives for accurate information, leading to well-known strategic disasters such as Hitler and Stalin’s blunders in World War II (Bialer, 1970; Noakes & Pridham, 2001), or at a smaller level, Saddam Hussein’s rejection of diplomacy (Atkinson, 1993). War makes collapse and extinction more likely, by raising the chance of weapons of mass destruction being used.

Linked to this, totalitarian states are less cooperative than democratic states. While cooperation is possible (Ginsburg, 2020), their internal norms are characterized by paranoia and treachery, and their lack of transparency limits their ability to credibly commit to agreements. This is bad for all risks that require cooperation such as pandemics or climate change (Tomasik, 2015).

Finally, continued social and scientific progress is likely to reduce risks of collapse and extinction. Social progress could reduce global inequality and other risk factors. Scientific progress could help address natural risks and climate change (Sandberg, 2018), differentially increase defensive rather than offensive power (Garfinkel & Dafoe, 2019), and solve safety challenges in AI or biotechnology (Russell, 2019). However, as we will now discuss totalitarian states would likely inhibit social progress.

A central question from a longtermist perspective is: Which values should shape the future? I would argue that we should prefer it to be shaped by liberal democratic values. This is not to say that the current democracy-dominated world is perfect—far from it. The fate of billions of factory-farmed animals or hundreds of millions of people in extreme poverty makes that abundantly clear. However, democracies have two advantages. First, democracies have space for cosmopolitan values such as human rights, plurality, freedom, and equality. These are better than those that characterize life under totalitarianism: Fear, terror, subjection, and secrecy. Second, they have within themselves the mechanism to allow progress. In the last 100 (or even 50) years, the lives of women, LGBT people, religious minorities, and non-white people have dramatically improved. Our “moral circle” has expanded, and could continue to expand (Singer, 1981). The arc of the moral universe is long, but given the right conditions, it might just bend toward justice (King, 1968). A global society dominated by these values, and with the possibility of improving more, has a better longterm potential. A totalitarian-dominated world, on the other hand, would reduce the space for resistance and progress—distorting the human trajectory.

We should be particularly concerned about “bottlenecks” at which values are particularly important—where there is a risk of “locking-in” some particular set of (possibly far from optimal) values. While they are currently faroff, future technologies such as artificial general intelligence, space settlement, life extension (of autocrats), or much better surveillance could enable lock-in (Caplan, 2008).17

Conditional on them avoiding new catastrophes, world orders dominated by totalitarians could be quite long-lasting (Caplan, 2008). Democracies can undermine authoritarian and totalitarian regimes through the following ways: Control, including conquest; contagion through proximity; and consent, promoting receptivity toward democratization (Whitehead, 2001). Democracies can actively undermine these regimes through war, sanctions, hosting rebellious exiles, or sponsoring internal movements. Passively, through contagion, they offer a demonstration that a better, more prosperous life is possible. For example, in the final years of the USSR, ordinary Soviet citizens were able to see that the West had a higher standard of living—more innovation, more choice, and more consumer goods. The elites were able to read books from the outside, and travel—Gorbachev’s contacts and friendships with European politicians may have made him more favorable to social democracy (Brown, 1996). Democracies can undermine the will and capacity of the coercive apparatus (Bellin, 2004). However, in a world not dominated by democracies, all these pressures would be far less.

A world in which, say, totalitarian regimes emerged as dominant after World War II (for example if the USA was defeated) could be self-reinforcing and long-lasting, like the self-reinforcing relationship of Oceania, Eurasia, and Eastasia (Orwell, 1949). Orwell’s fictional world is characterized by constant low-grade warfare to justify emergency powers and secure elites, and with shifting alliances of convenience as states bandwagon and balance, thereby preventing any resolution. A totalitarian-dominated world order could be rather robust, perhaps for decades or even centuries.

A long-lasting totalitarian-dominated world would extend the period of time humanity would spend with a heightened risk of collapse or extinction, as well as increased potential for distortion of the human trajectory and the possibility that a “lock-in” event may occur. This example illustrates the possibility of a “negative recovery,” resulting in a trajectory with less or no scientific and social progress and a less favorable geopolitical situation, which would threaten the destruction of humanity’s longterm potential.

#### \*Copyright protections are key—it’s the only way to ensure that revenue goes to creators of content

NMA 23 (News/Media Alliance staff, trade association representing approximately 2,000 newspapers in the United States and Canada. “White Paper: How the Pervasive Copying of Expressive Works to Train and Fuel Generative Artificial Intelligence Systems Is Copyright Infringement And Not a Fair Use” 10/21/2023. Accessed 5/21/2024. http://www.newsmediaalliance.org/wp-content/uploads/2023/10/AI-White-Paper-with-Technical-Analysis.pdf) wtk

The modes of distribution and consumption of publisher content are rapidly changing in the digital age, and the systematic copying and use of publisher content to fuel GAI systems and applications and to disseminate competing content poses what could be an existential threat to far too many publishers and is not a fair use. By diverting readers and the digital advertising dollars that follow them away from original sources, and by interfering with a potential source of licensing revenue for granting permissions, GAI models disincentivize investment in creation of those sources in the first place.

The continued unlicensed use of reporting also disserves the public interest: an online world that is dominated by GAI-generated, substitutional content is poised to leave the public with watered-down, less reliable outputs and fewer news outlets with the resources necessary to provide critical original reporting. As district court judge Denise Cote’s decision in Associated Press v. Meltwater U.S. Holdings, Inc. explained with respect to direct scraping of news content that is economically indistinguishable from that now being laundered through GAI systems, copyright law should not allow for democracy to be imperiled in this manner: [T]he world is indebted to the press for triumphs which have been gained by reason and humanity over error and oppression … Permitting Meltwater to take the fruit of AP’s labor for its own profit, without compensating AP, injures AP’s ability to perform [its] essential function of democracy.10

GAI is now further threatening the ability of journalists and publishers to perform that “essential function of democracy.” At a time when governments and experts around the world warn of the risk AI poses to democratic functioning,11 it is critical that the copyright laws continue to protect publisher content to help safeguard the indispensable role of a flourishing and free press.

#### \*The brink is now—generative AI exacerbates existing market imbalances in news media

Coffey 24 (Danielle Coffey, President & CEO of News/Media Alliance. “Oversight of A.I.: The Future of Journalism” testimony before the U.S. Senate Committee on the Judiciary’s Subcommittee on Privacy, Technology, and the Law. 1/10/2024. Accessed 5/20/2024. <https://www.judiciary.senate.gov/download/2024-01-10-testimony-coffey>) wtk \*GAI = Generative AI

However, as dire as the current dynamic is, this marketplace imbalance will only be increased by GAI. GAI developers crawl websites and reach behind paywalls to train their models. An analysis that the N/MA commissioned, as well as complaints pending before the courts, demonstrate the significant ingestion of quality news content in AI training models. Adding insult to injury, GAI “output” results to user inquiries often contain summaries, excerpts, and even full verbatim copies of articles written and fact-checked by human journalists. These outputs compete in the same market, with the same audience, serving the same purpose as the original articles that feed the algorithms in the first place. GAI is an exacerbation of an existing problem where revenue cannot be generated by, but in fact is diverted from, those who create the original work.

In the long term, protecting creators and rights holders from the unauthorized use of their works for training GAI will help both technology companies and news publishers provide better products and services to consumers. GAI models and products will not be sustainable if they eviscerate the quality content that they feed upon. Copyright-protected, expressive works have been taken without authorization and without consent, and used repeatedly in model training, processing, and display. Because these uses go far beyond the guardrails set by courts, now openly substituting for copyrighted content and usurping licensing markets, this should not be considered fair use under current copyright law.

#### Licensing is the only way to save journalism from AI

Radsch 24 (Courtney C. Radsch, PhD in international relations from American University, Nonresident Fellow in Governance Studies at the Center for Technology Innovation and Director of the Center for Journalism and Liberty at the Open Markets Institute for Brookings. “Can journalism survive AI?” 3/25/2024. Accessed 5/22/2024. https://www.brookings.edu/articles/can-journalism-survive-ai/) wtk

Requiring tech companies to license the use of news publishers’ content through this type of legislation would help ensure that smaller, local, niche, and non-English language news publishers would also be able to negotiate for the use of their content and data. This type of journalism could be particularly useful for localizing generative search, summarization, content creation, and other applications that make use of journalism to provide more accurate, timely, and relevant results, particularly in languages other than English.

Journalism can also be an important source of data for improving the quality of foundation models, which suffer from bias, misinformation, and spam that make access to diverse sources of quality, factual information, especially in low-resourced digital languages, even more valuable. Furthermore, as the quality of data becomes as important as the quantity of data, journalism provides a constant source of new, timely, human-generated data.

We are in a moment when the news industry needs to unite. As giant media conglomerates and major publications strike deals with the tech giants, they need to demand a framework that will benefit journalism in the public interest, not just line the pockets of their corporate owners. That is why the only way journalism will survive AI is to double down on its journalists. As important as it will be for journalism to adapt to and integrate AI, newsrooms that replace journalists will hasten its demise, with profound ramifications for democracy in the U.S. and around the world.

News outlets must consider how to optimize revenue streams and assert their pricing autonomy throughout the AI value chain. They will need to figure out how to unlock the value of journalism by adopting sophisticated and dynamic compensation frameworks and pricing strategies for news content in various parts of AI systems and AI applications. They will need access to information about the way their content is used in AI systems, including data sets and foundational model weights. And they will need government regulations that enable them to do so.

#### \*Licensing is feasible, even on a large scale

Kupferschmid 23 (Keith Kupferscmid, CEO of the Copyright Alliance, “Artificial Intelligence and Copyright: Comments of the Copyright Alliance” comment before the U.S. Copyright Office. 10/30/2023. Accessed 5/25/2024 from https://www.regulations.gov/comment/COLC-2023-0006-8935) wtk

Regarding the second part of this question, it is feasible to get consent on a mass scale. In fact, it happens frequently. Examples include licensing for music streaming services and voluntary collective licensing through organizations like the Copyright Clearance Center, both of which involve vast amounts of copyrighted works. Mass copyright infringement should not be rewarded by creating a special copyright exception for AI ingestion. Like others that have proceeded them, AI companies should be required to get authorization for any and all works they use (unless a defense applies under the law or if the works are in the public domain). The idea that just because it may be harder to get consent from copyright owners when large volumes of works are being used, it is therefore not infringement, would simply incentivize infringers to illegally copy more as a means for avoiding infringement—that cannot possibly be the law.

Getting a license, even when a large volume of works is being used is not difficult; there are flexible licensing models available and many different copyright management organizations to implement them. (See the response to question 10.2 for discussion of CMOs.) As noted in the Warhol decision, “licenses… are how [creators…] make a living. They provide an economic incentive to create original works, which is the goal of copyright.” 150 Thus, there is every incentive for copyright owners to make it easy for AI companies to license their works.

### 1ac—model collapse

#### Advantage 2 is Model Collapse

#### Status quo AI development is unsustainable—uncompensated use of human creations causes future AI to be trained on AI-generated works, which causes model collapse and widespread AI hallucination

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

Ironically, a policy of free appropriation of copyrighted work may even menace AI development itself. Simply put, it is not sustainable to expect training data to persist as a renewable resource when it is being mined, without compensation, in part to create substitutes for itself.67 Scholars in the field have identified a danger of LLMs “learning from data produced by other models,” a possibility that is more likely the less humans are compensated for their work.68 The researchers call this pathological outcome “model collapse,” “a degenerative process whereby, over time, models forget the true underlying data distribution, even in the absence of a shift in the distribution over time.” 69 Consider, for instance, a distribution of articles about a given topic existing at Time 1. Over time, early LLMs may generate material based on those articles. As later LLMs at Time 2 take in both the original human content, and the later LLM-generated content, their results can be skewed by the earlier LLMs’ random or otherwise unjustified selection and arrangement of key points from the human content, as well as the well-documented problems of hallucination and fabrication by LLMs.70

LLMs are *language* models, not *knowledge* models, and have no ability to independently reason about what is in the human-generated articles or images they process. Nor is text generated in response to requests for fiction or creative non-fiction reflective of a mind capable of apprehending the world, since LLMs are mere textpredictors. They do not interact with and sense the world as humans do.71 LLMs increasingly based on earlier LLM output may become, after sufficient iterations, like the faded analog copies of copies of copies that are familiar to those who recall widespread distribution of materials via copy machines—many of which became almost unrecognizably blurred and distorted over time.72

The bottom line here is grim. If uncompensated and uncontrolled expropriation of copyrighted works continues, many creatives are likely to be further demoralized and eventually defunded as AI unfairly outcompetes them, or effectively drowns them out. Low-cost automated content will strike many as a cornucopian gift—until it becomes clear that AI itself is dependent on ongoing input of human-generated works in order to improve and remain relevant in a changing world. At that point, it may be too late to reinvigorate creative industries left moribund by neglect. Much of an entire generation of writers, composers, journalists, actors, and other creatives may be missing, dissuaded from even trying to publish, disseminate, or profit from their expression, given how easily aspects of their expression can be mimicked via AI, and how rapidly their own contributions may be occluded or overwhelmed by AI expression. 73 Legislative interventions are critical to avoid such an unfair and ultimately selfdefeating outcome. Part III below describes a new opt-out mechanism that would give creatives more say over how their works are used.

#### Licensing solves—it prevents model collapse and hallucinations by creating a thriving market for original human content

Coffey and Smith 23 (Danielle Coffey, President and CEO of News/Media Alliance. Regan Smith, Senior Vice President and General Counsel for News/Media Alliance. “Artificial Intelligence and Copyright” Comments of the News/Media Alliance before the U.S. Copyright Office. Docket No. 2023–6. 10/30/2023. Accessed 5/21/2024. <http://www.newsmediaalliance.org/wp-content/uploads/2023/10/Final-NMA-USCO-AI-NOI-Response-Submitted-10.30.23.pdf>) wtk

Generative AI development is unlikely to succeed without a robust ecosystem that facilitates licensed use of valuable, authentic news media material. The failure to license publishing content may negatively impact the valuation of AI companies themselves, creating a cloud on the technology precisely because it is unlicensed.181 Companies that might otherwise want to license from and deploy generative AI products and services may hang back as long as the IP issues are unresolved.

And a market that facilitates licensed exchanges of human-created content is needed for continued innovation. Researchers have found “that use of model-generated content in training causes irreversible defects in the resulting models,” an effect they term “model collapse.”182 Even short of a complete model collapse under a deluge of synthetic content, there is an increased risk that generative AI chatbots could become an unattractive swamp of hallucinations without the ability to use human-created content that reflects thoughtful editorial judgment and creative expression.

The flourishing of AI technologies requires development that incorporates design principles that underscore public safety, security, and trust—as demonstrated by the recent voluntary commitments from leading AI companies to the Biden-Harris Administration and the Administration’s Executive Order.183 Companies that adequately account for intellectual property responsibilities in their business models at the outset will be better poised to enjoy the tremendous potential economic benefits promised by AI innovation.

#### Model collapse causes large-scale data poisoning—human-generated content is key to solve

Grad 23 (Peter Grad, writer for Tech Xplore, “AI models feeding on AI data may face death spiral” 6/19/2023. Accessed 6/10/2024. https://techxplore.com/news/2023-06-ai-death-spiral.html) wtk

Large language models are generating verbal pollution that threatens to undermine the very data such models are trained on.

That's the conclusion reached by a team of British and Canadian researchers exploring the impact of successive generations of ChatGPT generated text that will be culled for future models.

In a paper published on the arXiv preprint server and titled, "The Curse of Recursion: Training on Generated Data Makes Models Forget," the team predicted that the recursive nature of AI training will eventually lead to "model collapse."

"We discover that learning from data produced by other models causes model collapse—a degenerative process whereby, over time, models forget the true underlying data distribution," the team said.

Team member Ross Anderson, of University of Cambridge and University of Edinburgh, likened the effect to the diminishing quality of musical output.

"If you train a music model on Mozart," he said in a personal blog, "you can expect output that's a bit like Mozart but without the sparkle …and if [that version] trains the next generation, and so on, what will the fifth or sixth generation sound like?"

The authors note that model collapse is a threat similar to catastrophic forgetting and data poisoning.

In catastrophic forgetting, a model "forgets" previous data, sometimes abruptly, when learning new information. The impact is compounded over time.

In their new research, the team said, models don't forget previously learned data "but rather start misinterpreting what they believe to be real, by reinforcing their own beliefs."

Data poisoning is the malicious insertion of false information. Of course, this practice predated the use of large language models. But with the use of large-scale web crawls, the insertion of even a small amount of malicious data, the team said, can lead to widespread contamination.

"What is different with the arrival of large language models is the scale at which such poisoning can happen once it is automated," the team said.

Researcher Ilia Shumailov, of the University of Oxford, warned that "major degradation happens within just a few iterations, even when some of the original data is preserved."

"Errors from optimization imperfections, limited models and finite data," he continued, "ultimately cause synthetic data to be of low[er] quality. Over time mistakes compound and ultimately force models that learn from generated data to misperceive reality even further."

The researchers said that the nature of recursive learning is to dispense with low-probability events, referred to by statisticians as "tails of the distribution"

In his blog, Anderson warned, "using model-generated content in training causes irreversible defects. The tails of the original content distribution disappear. Within a few generations, text becomes garbage."

"Low-probability events are … vital to understand complex systems," the report noted.

The first large language models were trained on human-generated text. But with the rapid adoption of ChatGPT by industry and general users, enormous amounts of data are populating online sites.

The researchers urged that steps be taken to distinguish AI content from human-generated content and that efforts be made to preserve original content for future training purposes.

"Large language models are like fire," team member Anderson said, "a useful tool, but one that pollutes the environment. How will we cope with it?"

#### Specifically, model collapse makes intentional data poisoning more effective

Shumailov et al. 23 (Ilia Shumailov, University of Oxford, PhD in Computer Science from Cambridge. Zakhar Shumaylov, University of Cambridge. Yiren Zhao, Imperial College London. Yarin Gal, University of Oxford. Nicolas Papernot, University of Toronto & Vector Institute. Ross Anderson, University of Cambridge & University of Edinburgh. “THE CURSE OF RECURSION: TRAINING ON GENERATED DATA MAKES MODELS FORGET” last revised 4/14/2024. Arxiv. Accessed 6/10/2024 from [https://arxiv.org/abs/2305.17493) wtk](https://arxiv.org/abs/2305.17493)wtk)

Poisoning attacks are crafted and inserted during training in order to degrade the model’s performance when deployed [Biggio et al., 2012]. Malicious data can be inserted into training data to induce unintended behaviors that can be activated by special triggers [Gu et al., 2017]. The early literature on data poisoning focused mainly on supervised learning, where classifiers are trained with labeled samples. But with the emergence of contrastive learning [Radford et al., 2021] and LLMs [Brown et al., 2020], more recent models are trained with large-scale web crawls, making data poisoning attacks more feasible on these untrustworthy web sources. Recent studies have demonstrated that web-scale datasets can be poisoned by introducing malicious data into a small percentage of samples [Carlini and Terzis, 2021, Carlini et al., 2023].

3 What is Model Collapse?

Definition 3.1 (Model Collapse). Model Collapse is a degenerative process affecting generations of learned generative models, where generated data end up polluting the training set of the next generation of models; being trained on polluted data, they then mis-perceive reality. We separate two special cases: early model collapse and late model collapse. In early model collapse the model begins losing information about the tails of the distribution; in the late model collapse model entangles different modes of the original distributions and converges to a distribution that carries little resemblance to the original one, often with very small variance.

#### \*And it causes widespread AI hallucinations

Ghose 24 (Shomit Ghose, lecturer at UC Berkeley's College of Engineering, “Why Hallucinations Matter: Misinformation, Brand Safety and Cybersecurity in the Age of Generative AI” 5/2/2024. Accessed 6/10/2024. https://scet.berkeley.edu/why-hallucinations-matter-misinformation-brand-safety-and-cybersecurity-in-the-age-ofgenerative-ai/)

Why Does Generative AI Hallucinate?

We’ve been writing computer software for 80 years and we still produce bugs in our source code, leading to execution errors. It should come as no surprise to us that, as we find ourselves engulfed by data-driven technologies such as AI, we can find “bugs” within data’s complexity and volume, leading to AI hallucinations. The etiology of AI hallucination includes biased training data, the computational complexity inherent in deep neural networks, lack of contextual / domain understanding, adversarial attack, training on synthetic data (“model collapse”), and a failure to generalize the training data (“overfitting”).

#### \*That causes escalation of military and diplomatic crises

Ghose 24 (Shomit Ghose, lecturer at UC Berkeley's College of Engineering, “Why Hallucinations Matter: Misinformation, Brand Safety and Cybersecurity in the Age of Generative AI” 5/2/2024. Accessed 6/10/2024. https://scet.berkeley.edu/why-hallucinations-matter-misinformation-brand-safety-and-cybersecurity-in-the-age-ofgenerative-ai/)

Just as LLMs have shown themselves responsive to automated prompt engineering to yield desired results, so too have they shown themselves susceptible to adversarial attack via prompt engineering to yield malicious results (Yao et al. 2023, Deng et al. 2023, Jiang et al. 2024, Anil et al. 2024, Wei et al. 2023, Rao et al. 2024). Exemplary work in using an automated framework to jail-break text-to-image gen-AI, resulting in the production of not-suitable-for-work (NSFW) images, was done in the SneakyPrompt project (Yang et al. 2023). “Given a prompt that is blocked by a safety filter, SneakyPrompt repeatedly queries the text-to-image generative model and strategically perturbs tokens in the prompt based on [reinforcement learning and] the query results to bypass the safety filter.”

Beyond chatbots, the risks of hallucination extend into many other application areas of AI. LLMs have slipped into the technologies utilized to build autonomous robots (Zeng et al. 2023, Wang et al. 2024) and vehicles (Wen et al. 2024). Hallucinations have also shown themselves to be a significant issue in the field of healthcare, both in LLM-driven applications (Busch et al. 2024, Ahmad et al. 2023, Bruno et al. 2023) and in medical imaging (Bhadra 2021). Data as attack vector has been shown in self-driving vehicle technology, through “poltergeist” (Ji et al. 2021) and “phantom” (Nassi et al. 2020) attacks, and has also been demonstrated in inaudible voice command “dolphin” attacks (Zhang et al. 2017).

More unsettling still is LLM-based agents being “integrated into high-stakes military and diplomatic decision making”, as highlighted by Stanford’s center for Human-Centered Artificial Intelligence (Rivera et al. 2024). Here, the researchers found “that LLMs exhibit difficult-to-predict, escalatory behavior, which underscores the importance of understanding when, how, and why LLMs may fail in these high-stakes contexts”. In such settings, the risk hallucination may bring in escalating human conflict is clearly an unacceptable one.

#### AI hallucinations cause inadvertent escalation and flash wars

Klare 23 (Michael T. Klare, Five Colleges professor of Peace and World Security Studies, whose department is located at Hampshire College, “AI vs. AI: Flash Wars and Human Extinction” 7/13/2024. Accessed 6/10/2024. https://www.counterpunch.org/2023/07/13/ai-vs-ai-flash-wars-and-human-extinction/) wtk

Such a prospect should be ample cause for concern. To start with, consider the risk of errors and miscalculations by the algorithms at the heart of such systems. As top computer scientists have warned us, those algorithms are capable of remarkably inexplicable mistakes and, to use the AI term of the moment, “hallucinations” — that is, seemingly reasonable results that are entirely illusionary. Under the circumstances, it’s not hard to imagine such computers “hallucinating” an imminent enemy attack and launching a war that might otherwise have been avoided.

And that’s not the worst of the dangers to consider. After all, there’s the obvious likelihood that America’s adversaries will similarly equip their forces with robot generals. In other words, future wars are likely to be fought by one set of AI systems against another, both linked to nuclear weaponry, with entirely unpredictable — but potentially catastrophic — results.

Not much is known (from public sources at least) about Russian and Chinese efforts to automate their military command-and-control systems, but both countries are thought to be developing networks comparable to the Pentagon’s JADC2. As early as 2014, in fact, Russia inaugurated a National Defense Control Center (NDCC) in Moscow, a centralized command post for assessing global threats and initiating whatever military action is deemed necessary, whether of a non-nuclear or nuclear nature. Like JADC2, the NDCC is designed to collect information on enemy moves from multiple sources and provide senior officers with guidance on possible responses.

China is said to be pursuing an even more elaborate, if similar, enterprise under the rubric of “Multi-Domain Precision Warfare” (MDPW). According to the Pentagon’s 2022 report on Chinese military developments, its military, the People’s Liberation Army, is being trained and equipped to use AI-enabled sensors and computer networks to “rapidly identify key vulnerabilities in the U.S. operational system and then combine joint forces across domains to launch precision strikes against those vulnerabilities.”

Picture, then, a future war between the U.S. and Russia or China (or both) in which the JADC2 commands all U.S. forces, while Russia’s NDCC and China’s MDPW command those countries’ forces. Consider, as well, that all three systems are likely to experience errors and hallucinations. How safe will humans be when robot generals decide that it’s time to “win” the war by nuking their enemies?

If this strikes you as an outlandish scenario, think again, at least according to the leadership of the National Security Commission on Artificial Intelligence, a congressionally mandated enterprise that was chaired by Eric Schmidt, former head of Google, and Robert Work, former deputy secretary of defense. “While the Commission believes that properly designed, tested, and utilized AI-enabled and autonomous weapon systems will bring substantial military and even humanitarian benefit, the unchecked global use of such systems potentially risks unintended conflict escalation and crisis instability,” it affirmed in its Final Report. Such dangers could arise, it stated, “because of challenging and untested complexities of interaction between AI-enabled and autonomous weapon systems on the battlefield” — when, that is, AI fights AI.

## journalism advantage extensions

### 2ac they say: “licensing fails”

#### Empirics prove licensing is feasible and creates new revenue streams for journalism

Lynch 24 (Roger Lynch, CEO of Condé Nast, “Oversight of A.I.: The Future of Journalism” Testimony before the U.S. Senate Committee on the Judiciary, Subcommittee on Privacy, Technology, and the Law. 1/10/2024. Accessed 5/17/2024 https://www.judiciary.senate.gov/download/2024-01-10-testimony-lynch) wtk

Big tech companies claim that getting permission for the use of copyrighted content isn't practical, but it is. There are a great many situations where multitudes of rights owners license multitudes of users in efficient ways. In music publishing, ASCAP, BMI, SESAC, GMR and others fulfill this role. The Copyright Clearance Center, global mechanical rights organizations, and companies like Shutterstock and Getty Images - all of them are private market solutions to aggregating rights for license. I am confident that the free market can generate efficient licensing solutions once the Gen Al companies acknowledge the need to license. The constitutional foundation of copyright is as important now as it was long ago. Our nation's founders believed that by creating a property right for authors and inventors, market forces, rather than politics, would provide incentives for the "progress of science and useful arts."

And there's good reason to believe that licensing deals will directly result in future investment in content. Following adoption of a law requiring that big tech platforms pay for news content in Australia, $140M in incremental annual revenue for journalism is resulting in hiring of a significant number of journalists and other investments.

### 1ar they say: “licensing fails—scale”

#### The scale is manageable—don’t conflate ai companies wanting to ingest every work with them needing to

Kupferschmid 23 (Keith Kupferscmid, CEO of the Copyright Alliance, “Artificial Intelligence and Copyright: Comments of the Copyright Alliance” comment before the U.S. Copyright Office. 10/30/2023. Accessed 5/25/2024 from [https://www.regulations.gov/comment/COLC-2023-0006-8935) wtk](https://www.regulations.gov/comment/COLC-2023-0006-8935)wtk) \*italics in original

As discussed in our response to question 8.1, some have argued that licensing copyright protected material for AI ingestion is not practical because, in order to be successful, AI systems *must* use *every* piece of available content. It might be desirable to train an AI tool on as much content as practically can be scraped from the internet, but—as evidenced by the many successful AI developers that do not scrape the internet for ingestion purposes—that does not make it necessary.

Just because AI developers *want* to use everything to train their systems doesn’t mean that it is necessary or justified. There is evidence that generative AI can succeed when ingesting only copyrighted materials that are properly licensed or in the public domain. We understand that the large general purpose LLMs do require a lot of ingestion material, but there is no reason that they cannot rely on public domain, open access and licensed texts. In the image space, Adobe’s Firefly suite of generative AI tools—which are trained on proprietary stock images, licensed images, and public domain images whose copyrights have expired—has seen broad consumer adoption.158 Additionally, leading AI developer Nvidia has partnered with Getty Images to build new generative AI technologies that ingest only fully licensed works, and IBM recently announced a collaboration with Adobe to assist customers in implementing generative AI models based on Adobe’s Firefly technology.159 These are just some examples that demonstrate that the foundation of an AI model can be built on licensed works.

### 1ar they say: “licensing fails—can’t administer”

#### Licenses are easy to administer—there are tons of organizations that can do it

Kupferschmid 23 (Keith Kupferscmid, CEO of the Copyright Alliance, “Artificial Intelligence and Copyright: Comments of the Copyright Alliance” comment before the U.S. Copyright Office. 10/30/2023. Accessed 5/25/2024 from https://www.regulations.gov/comment/COLC-2023-0006-8935) wtk

There are numerous voluntary collective management organizations (CMOs) that administer copyright owners’ rights. Perhaps the most well-known and established CMOs are in the music industry, namely, ASCAP, BMI, SESAC, and GMR. 160 Other CMOs in the music space include Vydia, 161 AWAL, 162 and Merlin.163 Additionally, there have been agreements reached between music publishers and platforms using copyrighted works that establish licensing systems through which royalties are distributed.164 And there are many others.

Outside of music, there are many other CMOs for other types of works. These include, but are not limited to:

• American Society for Visual Arts Licensing (ASCRL), which collects foreign payments for works of visual art that are mandated by foreign law and distributes those payments to its members.165

• Artists Rights Society (ARS), which collects foreign payments for works of fine art that are mandated by foreign law and distributes those payments to its members. 166

• Copyright Clearance Center (CCC), which collects and distributes license royalties for literary works.167

• Motion Picture Licensing Corporation (MPLC) and SWANK, which license on a nonexclusive basis the public performance of copyrighted motion pictures, television programs and other audiovisual works that were originally intended for personal use only.168

There are many other CMOs and more CMOs are in the process of being developed as result of AI.169

### 1ar they say: “licensing fails—complexity”

#### Licensing is not too complex—it’s been done thousands of times

Hirono and Legeyt 24 (Mazie Keiko Hirono is a Japanese- American lawyer and politician serving as the junior United States senator from Hawaii. Curtis LeGeyt is the President and Chief Executive Officer National Association of Broadcasters. “Oversight of A.I.: The Future of Journalism” hearing of the Senate Judiciary Committee, Subcommittee on Privacy, Technology, and the Law. Committee on the Judiciary. 1/10/2024. Accessed 5/20/2024 on ProQuest Congressional via University of Michigan online library)

MAZIE K. HIRONO: Thank you very much. As we wrestle with what kind of regulation or parameters we should put on the training and AI, it just seems to make sense that the creators of the content should get some kind of compensation for the -- the use of their material in training the AI platforms, so. This is for Mr. LeGeyt, don't we already have a -- do you not already have a lot of expertise, experience in figuring out what would be an appropriate way to license for copyrighted material and other -- the use of these -- this kind of content?

And how would that be applied in the case of AI?

CURTIS LEGEYT: Senator, thank you for the question. Over the last three decades, local television broadcasters have literally done thousands of deals with cable and satellite systems across the country for the distribution of their programing. The notion that the tech industry is saying that it is too complicated to license from such a diverse array of content owners just doesn't stand up. We negotiate with some of the largest cable systems in the country.

We negotiate with small mom and pop cable systems. And the result of that is that instead of just doing national deals with our networks, our broadcast networks or news organizations in New York and LA, every local community in the country is served by a locally focused broadcast station. And that station's programing is carried on the local cable and satellite system.

So, we have a lot of experience here. Similarly on the radio side, there are licensing organizations, Mr. Lynch referred to this in his testimony, um that allow for collective licensing of songwriter rights, performing rights. All of this, our industry has been in the center of both as a licensor and licensee.

And has been tremendously beneficial.

### 2ac they say: “absolution turn”

#### Licensing agreements are necessary for maintaining journalism and keeping an informed citizenry

Calvert 24 (Clay Calvert, Nonresident Senior Fellow in Technology Policy Studies at the American Enterprise Institute JD from McGeorge School of Law, University of the Pacific and PhD in communication from Stanford University. “Content Creators vs. Generative Artificial Intelligence: Paying a Fair Share to Support a Reliable Information Ecosystem” AEI. 1/3/2024. Date Accessed: 5/13/2024. <https://www.aei.org/technology-and-innovation/content-creators-vs-generative-artificial-intelligence-paying-a-fair-share-to-support-a-reliable-information-ecosystem/>) wtk

National newspapers like the Times and Wall Street Journal will play an increasingly vital role in informing citizens as reliable local journalism dwindles. Northwestern University’s November 2023 report on the state of local news predicts that given “the current trajectory, by the end of [2024], the country will have lost a third of its newspapers since 2005. Discouragingly, the growth in alternative local news sources—digital and ethnic news outlets, as well as public broadcasting—has not kept pace with what’s being lost.” The report adds that “residents in more than half of U.S. counties have no, or very limited, access to a reliable local news source—either print, digital or broadcast.”

Large daily newspapers also are suffering, with fewer journalists employed to cover the news. The Northwestern report asserts that “many of the large dailies owned by chains employ less than a fifth of the journalists on staff in 2005.” Gannett, the country’s largest newspaper chain, has severely cut its number of employees in the past three years.

One solution involves generative AI companies entering into licensing agreements with content creators, paying them for using their copyrighted works. The venerable Associated Press news service did just that in July 2023 with OpenAI. Additionally, Axel Springer, owner of Politico and Business Insider, entered into an agreement with OpenAI last month that allows content usage from “Springer media brands for advancing the training of OpenAI’s sophisticated” LLMs.

OpenAI virtuously proclaims it wants to “redistribute profits from our work to maximize the social and economic benefits of AI technology.” Forget––temporarily––post-hoc redistribution: Paying the Times and news organizations upfront for content that economically benefits OpenAI’s generative beast is essential for a well-informed citizenry.

#### Licensing is key to build trust in information and journalism

Coffey and Smith 23 (Danielle Coffey, President and CEO of News/Media Alliance. Regan Smith, Senior Vice President and General Counsel for News/Media Alliance. “Artificial Intelligence and Copyright” Comments of the News/Media Alliance before the U.S. Copyright Office. Docket No. 2023–6. 10/30/2023. Accessed 5/21/2024. <http://www.newsmediaalliance.org/wp-content/uploads/2023/10/Final-NMA-USCO-AI-NOI-Response-Submitted-10.30.23.pdf>) wtk

The continued unlicensed use of reporting—including entire corpora of unique publisher content, amounting up to millions of stories—portends injury, not just to the news industry, but to the public interest that it serves: an online world that is dominated by AI-generated, inferior yet substitutional content will leave the public with watered-down, less reliable outputs and fewer news outlets with the resources necessary to provide critical original reporting. As district court judge Denise Cote’s decision in Associated Press v. Meltwater U.S. Holdings, Inc. explained with respect to direct scraping of news content, copyright law does not allow for democracy to be imperiled in this manner: [T]he world is indebted to the press for triumphs which have been gained by reason and humanity over error and oppression … Permitting [Meltwater] to take the fruit of [AP’s] labor for its own profit, without compensating [AP], injures [AP’s] ability to perform [its] essential function of democracy.23 In addition to decreasing readership, the unauthorized use of publisher content to produce outputs that include inaccuracies also devalues publisher brands and creative content by muddling the source of the original content and misattributing information or misinformation to unrelated publishers or journalists.24 This is especially damaging as many of N/MA’s members have spent years or decades—sometimes even centuries—building their reputation as reliable and trustworthy content producers, providers, and curators. This reputation is vital for their success, with readers associating their brands with content that has been researched, vetted, proofread, and carefully considered by consummate professionals they know and trust. Indeed, trusted journalism can be an antidote to the mis- and disinformation problem.25

It is therefore particularly concerning when a generative AI system attributes material that is blatantly false to a publisher who has never published such information. As one example, take the case of Jonathan Turley, a law professor who ChatGPT falsely accused of sexually harassing a student, attributing the information to a non-existent news article by The Washington Post.26 In the same research experiment, conducted by Professor Eugene Volokh, ChatGPT made other similarly false allegations, citing articles that did not exist from publishers such as the Miami Herald and the Los Angeles Times. These “hallucinations,'' or massive errors, are a recognized propensity of many generative AI models that can spread misinformation and cause real harm to publisher brands. Other examples of the dangers of “hallucinations'' and other harms include summaries of articles by reputable publishers combining information from unreputable sources and the proliferation of deepfake photographs in politics.27 Publishers recognize these pitfalls and while some may use AI as a tool in newsgathering and content production processes, they accept legal responsibility for the content they publish and understand that the outputs are often not reliable and require human editing and supervision before publication—something that generative AI systems typically do not have.

To mitigate these risks, it is essential that generative AI training datasets, systems, and applications be based on reliable, trustworthy, and high-quality content with adequate safeguards to deter misinterpretations and the creation of false information based on that content. To do so sustainably and lawfully—in a manner that protects the public interest, including professional journalism—generative AI developers should license content from publishers for training and grounding purposes based on fair and transparent negotiations, as discussed in more detail below and in the attached White Paper. Only then can publishers recoup their investments in high-quality journalism while assuring developers that generative AI systems are built on authentic—not synthetic—content that is essential for reliable and trustworthy AI.

In the worst-case scenario, without an enforceable licensing market, high-quality publishers will slowly fail, forcing generative AI systems to rely on each other for training, leading to the gradual degradation in the availability of reliable and trustworthy reporting to our communities and system of democratic governance.28 In fact, without human-generated quality content to train AI, researchers have found “that use of model-generated content in training causes irreversible defects in the resulting models,” an effect they term “model collapse”29 or “Model Autophagy Disorder (MAD),”30 an analogy to mad cow disease: For instance, start with a language model trained on human-produced data. Use the model to generate some AI output. Then use that output to train a new instance of the model and use the resulting output to train a third version, and so forth. With each iteration, errors build atop one another. The 10th model, prompted to write about historical English architecture, spews out gibberish about jackrabbits.31 It is therefore in all of our collective interest that generative AI companies adhere with the letter and spirit of intellectual property law.

### 2ac they say: “can’t solve—consolidation”

#### AI is the brink—it further entrenches the power of big tech companies

Radsch 24 (Courtney C. Radsch, PhD in international relations from American University, Nonresident Fellow in Governance Studies at the Center for Technology Innovation and Director of the Center for Journalism and Liberty at the Open Markets Institute for Brookings. “Can journalism survive AI?” 3/25/2024. Accessed 5/22/2024. https://www.brookings.edu/articles/can-journalism-survive-ai/) wtk

The rapid advances in artificial intelligence are becoming yet another way for a handful of powerful tech corporations to extend and entrench their already dominant market positions. This will make it difficult, if not impossible, for sectors like journalism or the creative industries to remain independent, much less to maintain a public interest orientation as should be the case for the news industry.

The AI revolution underway extends the “platformization” of journalism and the power that a handful of tech firms maintain over our information channels and our public discourse. This, in turn, will exacerbate the ways in which these corporations are already threatening and cheapening real journalism while exploiting the labor of millions of journalists and others to build their models and develop applications that alter our economies and societies.

### 2ac they say: “journalism resilient”

#### Journalism isn’t resilient—AI is eroding search traffic

Lynch 24 (Roger Lynch, CEO of Condé Nast, “Oversight of A.I.: The Future of Journalism” Testimony before the U.S. Senate Committee on the Judiciary, Subcommittee on Privacy, Technology, and the Law. 1/10/2024. Accessed 5/17/2024 https://www.judiciary.senate.gov/download/2024-01-10-testimony-lynch) wtk

Gen Al companies are using our stolen intellectual property to build tools of replacement.

Today's Gen Al tools maintain complete copies of the works they train on, including our content, and output the substance — sometimes verbatim, sometimes paraphrased — while keeping 100% of the value for themselves. They are training consumers to come to them for information, not to us, and, unlike traditional search, they are keeping consumers within their experiences, depriving us of the opportunity to connect with our audiences directly, customize our content for them, and generate advertising and subscription revenue, sales leads and other valuable data. By misappropriating our content in this way, they are directly threatening the viability of the media ecosystem.

Some Gen AIS provide their own output, others use a combination of natural language outputs and search, which is called retrieval augmented generation (or "RAG"). Some Gen Al companies claim this is "just search" but that couldn't be further from the truth.

A traditional search engine response contains only a partial snippet and links to the content provider's website; think of it as a teaser that entices the reader to click through to read the full piece of content. Search engines are vehicles to discover content, not replace it. But in Gen Al responses, users are provided with a complete answer to their query and the opportunity to ask follow up questions. The user receives the information derived from our sites without further clicks and without being transferred to our sites.

This isn't just a traditional publisher versus tech dispute. In this context, many tech companies are also publishers — for example, Twitter/X and Meta are suing a scraper called Bright Data for misusing their scraped content and Reddit has taken a variety of steps to make clear that they are entitled to be compensated for training on their data. These tech companies also believe they are entitled to be compensated for the use of their systems and content.

Some Al companies will tell you that publishers can opt out from Al training. This is misleading. First, opting out of future training does not solve for the content they have already taken. Second, most publishers generate a substantial amount of traffic from search. In order to opt out of the new search engines powered by Al, such as Google's "Search Generative Experience", we would have to opt out of search, which would materially damage our businesses.

### 2ac they say: “democracy doesn’t solve war”

#### Democratic backsliding magnifies every existential threat.

Dr. Larry Diamond 19. Professor of Political Science and Sociology at Stanford University, Senior Fellow at the Hoover Institution, Senior Fellow at the Freeman Spogli Institute for International Studies, PhD in Sociology from Stanford University. “Ill Winds: Saving Democracy from Russian Rage, Chinese Ambition, and American Complacency.” p. 199-202.

The most obvious response to the ill winds blowing from the world’s autocracies is to help the winds of freedom blowing in the other direction. The democracies of the West cannot save themselves if they do not stand with democrats around the world. This is truer now than ever, for several reasons. We live in a globalized world, one in which models, trends, and ideas cascade across borders. Any wind of change may gather quickly and blow with gale force. People everywhere form ideas about how to govern—or simply about which forms of government and sources of power may be irresistible—based on what they see happening elsewhere. We are now immersed in a fierce global contest of ideas, information, and norms. In the digital age, that contest is moving at lightning speed, shaping how people think about their political systems and the way the world runs. As doubts about and threats to democracy are mounting in the West, this is not a contest that the democracies can afford to lose. Globalization, with its flows of trade and information, raises the stakes for us in another way. Authoritarian and badly governed regimes increasingly pose a direct threat to popular sovereignty and the rule of law in our own democracies. Covert flows of money and influence are subverting and corrupting our democratic processes and institutions. They will not stop just because Americans and others pretend that we have no stake in the future of freedom in the world. If we want to defend the core principles of self-government, transparency, and accountability in our own democracies, we have no choice but to promote them globally. It is not enough to say that dictatorship is bad and that democracy, however flawed, is still better. Popular enthusiasm for a lesser evil cannot be sustained indefinitely. People need the inspiration of a positive vision. Democracy must demonstrate that it is a just and fair political system that advances humane values and the common good. To make our republics more perfect, established democracies must not only adopt reforms to more fully include and empower their own citizens. They must also support people, groups, and institutions struggling to achieve democratic values elsewhere. The best way to counter Russian rage and Chinese ambition is to show that Moscow and Beijing are on the wrong side of history; that people everywhere yearn to be free; and that they can make freedom work to achieve a more just, sustainable, and prosperous society. In our networked age, both idealism and the harder imperatives of global power and security argue for more democracy, not less. For one thing, if we do not worry about the quality of governance in lower-income countries, we will face more and more troubled and failing states. Famine and genocide are the curse of authoritarian states, not democratic ones. Outright state collapse is the ultimate, bitter fruit of tyranny. When countries like Syria, Libya, and Afghanistan descend into civil war; when poor states in Africa cannot generate jobs and improve their citizens’ lives due to rule by corrupt and callous strongmen; when Central American societies are held hostage by brutal gangs and kleptocratic rulers, people flee—and wash up on the shores of the democracies. Europe and the United States cannot withstand the rising pressures of immigration unless they work to support better, more stable and accountable government in troubled countries. The world has simply grown too small, too flat, and too fast to wall off rotten states and pretend they are on some other planet. Hard security interests are at stake. As even the Trump administration’s 2017 National Security Strategy makes clear, the main threats to U.S. national security all stem from authoritarianism, whether in the form of tyrannies from Russia and China to Iran and North Korea or in the guise of antidemocratic terrorist movements such as ISIS.1 By supporting the development of democracy around the world, we can deny these authoritarian adversaries the geopolitical running room they seek. Just as Russia, China, and Iran are trying to undermine democracies to bend other countries to their will, so too can we contain these autocrats’ ambitions by helping other countries build effective, resilient democracies that can withstand the dictators’ malevolence. Of course, democratically elected governments with open societies will not support the American line on every issue. But no free society wants to mortgage its future to another country. The American national interest would best be secured by a pluralistic world of free countries—one in which autocrats can no longer use corruption and coercion to gobble up resources, alliances, and territory. If you look back over our history to see who has posed a threat to the United States and our allies, it has always been authoritarian regimes and empires. As political scientists have long noted, no two democracies have ever gone to war with each other—ever. It is not the democracies of the world that are supporting international terrorism, proliferating weapons of mass destruction, or threatening the territory of their neighbors. For all these reasons, we need a new global campaign for freedom. Everything I am proposing in this book plays a role in that campaign, but in this chapter, I am concerned more narrowly with the ways that we can directly advance democracy, human rights, and the rule of law in the twenty-first-century world. As with any policy area, many of the challenges can be somewhat technical, requiring smart design and the careful management of programs and institutions. Those operational debates I leave for another venue. Here, I make a more basic case for four imperatives. First, we must support the democrats of the world—the people and organizations struggling to create and improve free and accountable government. Second, we must support struggling and developing democracies, helping them to grow their economies and strengthen their institutions. Third, we must pressure authoritarian regimes to stop abusing the rights and stealing the resources of their citizens, including by imposing sanctions on dictators to make them think hard about their choices and separate them from both their supporters and the people at large. Finally, we need to reboot our public diplomacy—our global networks of information and ideas—for today’s fast-paced age of information and disinformation. For the sake of both our interests and our values, we need a foreign policy that puts a high priority on democracy, human rights, and the rule of law.

## model collapse advantage extensions

### 2ac they say: “no model collapse”

#### AI model collapse is coming now because AI is being trained on AI-created works

Marr 24 (Bernard Marr, Contributor for Forbes, “Generative AI And The Risk Of Inbreeding” 3/28/2024. Accessed 6/10/2024. https://www.forbes.com/sites/bernardmarr/2024/03/28/generative-ai-and-the-risk-of-inbreeding/) wtk

What Is Inbreeding In Relation To Generative AI?

The term refers to the way in which generative AI systems are trained. The earliest large language models (LLMs), were trained on massive quantities of text, visual and audio content, typically scraped from the internet. We’re talking about books, articles, artworks, and other content available online – content that was, by and large, created by humans.

Now, however, we have a plethora of generative AI tools flooding the internet with AI-generated content – from blog posts and news articles, to AI artwork. This means that future AI tools will be trained on datasets that contain more and more AI-generated content. Content that isn’t created by humans, but simulates human output. And as new systems learn from this simulated content, and create their own content based on it, the risk is that content will become progressively worse. Like taking a photocopy of a photocopy of a photocopy.

It’s not dissimilar to human or livestock inbreeding, then. The “gene pool” – in this case, the content used to train generative AI systems – becomes less diverse. Less interesting. More distorted. Less representative of actual human content.

What Would This Mean For Generative AI Systems?

Inbreeding could pose a significant problem for future generative AI systems, rendering them less and less able to accurately simulate human language and creativity. One study has confirmed how inbreeding leads to generative AIs becoming less effective, finding that “without enough fresh real data in each generation … future generative models are doomed to have their quality (precision) or diversity (recall) progressively decrease.”

In other words, AIs need fresh (human-generated) data to get better and better over time. If the data they’re trained on is increasingly generated by other AIs, you end up with what’s called “model collapse.” Which is a fancy way of saying the AIs get dumber. This can happen with any sort of generative AI output – not just text but also images. This video shows what happens when two generative AI models bounce back and forth between each other, with one AI describing an image and then the other creating an image based on the description, and so on and so on in a loop. The starting point was the Mona Lisa, one of the world’s great masterpieces. The end result is just a freaky picture of squiggly lines.

### 2ac they say: “can’t solve in time”

#### Human generated content can prevent model collapse

Cooke 23 (Elizabeth Cooke is a trainee reporter, specialising in thematic intelligence. She graduated with a degree in French and Philosophy from Oxford University in 2022. “AI model collapse could spell disaster for AI development, say new studies” 7/3/2023. Accessed 6/6/2024. https://www.verdict.co.uk/ai-model-collapse-could-spell-disaster-for-ai-development-say-new-studies/?cf-view) wtk

Two new studies have found that when generated data begins to populate the training sets of future AI models, there is a significant degradation in the quality and diversity of the generated output, leading to “model collapse”.

“Model collapse” is a degenerative process whereby models, trained on data polluted by AI-generated data, forget the true underlying data distribution.

One of the studies, “The Curse of Recursion: Training on Generated Data Makes Models Forget”, says that Big Tech companies such as OpenAI and Google benefit from a “first mover advantage” when it comes to training large language models (LLM)s. This is because training of samples from another generative model can induce a “distribution shift”, which causes the model’s predictions to become less accurate over time.

The study, co-authored by researchers at the University of Oxford, the University of Cambridge, Imperial College London and the University of Toronto, emphasises the need to preserve access to the original data source and to continue creating new human-generated data sources.

### 2ac they say: “plan causes worse models”

#### Unlicensed scraping causes model collapse—plan solves

Miller 23 (Courtney Miller, Masters degree in Information Security, Programmer, “Artificial Intelligence and Copyright” Reply Comments of Courtney Miller before the U.S. Copyright Office. Docket No. 2023–6. 8/30/2024. Accessed 6/6/2024. [https://www.regulations.gov/comment/COLC-2023-0006-0725) wtk](https://www.regulations.gov/comment/COLC-2023-0006-0725)wtk) \*edited for ableist language

This problem will only become worse as time goes on, due to the concept of model collapse. Since AI companies are scraping the Internet without need to worry about copyright, they continue to train their models on data which contains an increasing amount of AI-generated material, itself. This eventually leads to complete model failure—a circumstance which could ~~cripple~~[hamper] AI companies, but only after they have first made the Internet almost entirely unusable. AI companies know that model collapse is likely, which is why they assign low-wage human workers to comb through material and except data which is obviously vulgar, dangerous, or fake… but even the average person is often taken in by AI-generated articles, which sound authoritative even while they spout complete nonsense. A low-wage worker has little chance of cleaning up this false data unless it is wildly obvious.

#### Limiting fair use exceptions solves model collapse—Australia proves

Bonyhady 23 (Nick Bonyhady, Technology writer for the Financial Review, “AI companies face ‘model collapse’. They should pay to fix it” 12/29/2023. Accessed 6/10/2024. [https://www.afr.com/technology/ai-companies-face-model-collapse-they-should-pay-to-fix-it-20231228-p5eu0r) wtk](https://www.afr.com/technology/ai-companies-face-model-collapse-they-should-pay-to-fix-it-20231228-p5eu0r)wtk)

It’s a symptom of a phenomenon known as “model collapse”.

Training models on fresh, original material is one antidote. It’s why sites like The New York Times are prominent in the training data used by companies including OpenAI. It’s also why OpenAI and its part owner, Microsoft, are now facing a lawsuit from the Times alleging mass copyright breaches.

But the landmark lawsuit is unlikely to resolve a central question for societies in the AI age: how to account for the works created by real people and posted online that individually are worth virtually nothing but collectively train technology that is worth tens of billions of dollars.

The news: The New York Times sued OpenAI and Microsoft on Thursday, AEDT, in the US District Court in Manhattan, alleging mass copyright breaches. “[The] defendants’ generative AI tools can generate output that recites Times content verbatim, closely summarises it, and mimics its expressive style, as demonstrated by scores of examples,” the lawsuit reads. “These tools also wrongly attribute false information to the Times.” The $US7.8 billion newspaper does not specify a remedy but asks the court to hold OpenAI and Microsoft “responsible for the billions of dollars in statutory and actual damages” it says they owe.

Between the lines: US copyright law permits “fair use” – a broad, principled exception to copyright protection – of published works without payment. If material has been “transformed”, then that is generally fair use. That could be as simple as recreating scenes from a movie for a documentary, putting legal briefs in a database or using lines of computer code for a fresh purpose.

So, OpenAI’s claim that it is engaging in “fair use” when it trains models on many millions of pieces of copyrighted material and remixes that mass into new text based on user prompts is a credible argument.

That’s why the Times’ legal filings highlight cases where OpenAI’s tools spit out text that is identical to real news articles. It points to paragraph after paragraph that ChatGPT and the Microsoft tool it powers, Copilot, reproduced from major Times stories on the New York taxi industry, a deadly avalanche and Israel’s failure to stop Hamas’ October attack. That practice of copying text or style looks much less transformative and could be vulnerable.

Down under: Australia’s copyright regime is different to the US. Here, there are just five “fair dealing” exceptions: research or study, criticism or review, parody or satire, news reporting, and legal advice. That suggests a copyright claim would be easier to win, though overseas technology companies have often tried to claim they cannot be sued in Australia.

Our take: Under the Morrison government, Australia went out ahead of the world and legislated to essentially force Google and Facebook to pay the media (including The Australian Financial Review’s owner Nine) for news on their platforms. It has done little, whether via copyright or anything else, to address the same question for AI.

Some publishers, including Germany’s Axel Springer and the American wire service The Associated Press, have already struck deals with OpenAI. Yet enough publishers and AI makers won’t – especially in smaller markets like Australia where there is an imbalance of market power – that governments will be corralled to intervene.

### 2ac they say: “data poisoning doesn’t cause war”

#### Data poisoning escalates and causes war with state and non-state actors

James Johnson, (2021) - Dr. James Johnson is an Assistant Professor in the School of Law and Government at Dublin City University and a Non-Resident Fellow with the Modern War Institute at West Point. Dr Johnson was previously a Postdoctoral Research Fellow at the James Martin Center for Nonproliferation Studies in Monterey, California. ‘Catalytic nuclear war’ in the age of artificial intelligence & autonomy: Emerging military technology and escalation risk between nuclear-armed states, Journal of Strategic Studies, DOI: 10.1080/01402390.2020.1867541

Disinformation, misinformation, and information manipulation

One rapidly developing and increasingly prominent field of Al-augmented technology that can complement and force multiply existing malicious social manipulation behavior and generate campaigns of manipulation - most notably the spread of misinformation or disinformation - is the ability to generate audio and video images that fabricate events, create fictitious situations and propagate falsehoods.121 Experts expect that it will be possible for non-state actors (or amateurs more generally) to generate photorealistic high definition video, audio, and document forgeries using machine-learning tools - in particular, generative adversarial networks (GANs) - at a low-cost and on a large scale.122 In combination with other tools,123 GANs have already demonstrated that producing realistic images is possible and will likely improve significantly in the next few years.124 Examples of disinformation and manipulation, among other things, include distributed audio of private statements and conversations between political or military leaders (both domestically or with allies and adversaries); video of crises or conflict (see below) designed to incite public outrage and prompt pressure for retribution; or audio or video material calibrated to deflect, deceive, or otherwise distract attention from, an actor's culpability of an aggressor by producing false alternative versions of events.125

As Al technology advances, the quality, cost, and availability of GANs and other tools - especially Al-enhanced audio software will make it increasingly difficult to discern what is real from what is not, eroding public trust in hitherto trust-worthy information sources.126 In 2014, for example, thousands of residents at St. Mary Parish in Louisiana received a fake text message alert via a bogus Twitter account warning of a 'toxic fume hazard' in the area. Further fanning the flames, a fake YouTube video was also posted showing a masked ISIS fighter standing next to looping footage of an explosion.127 Thus, it is not difficult to imagine how these Al-enhanced technologies in the hands of non-state actors with nefarious goals (or 'apocalyptic' world view) might have dangerous consequences or for nuclear security and strategic stability. Deliberate malevolent information manipulation by non-state actors (terrorists, criminals, or state proxies) could destabilize implications on effective deterrence and military planning, both during peace and war. GANs generated deepfakes might also exacerbate the escalation risks by manipulating the digital information landscape, where decisions about nuclear weapons are made. It is easy to imagine unprovoked escalation caused by a malicious third-party (or state-proxy) clandestine false-flag operation in the competitive strategic environment.128 During a crisis, a state's inability to determine an attacker's intent may lead an actor to conclude that an attack - threatened or actual - was intended to undermine its nuclear deterrent.129

Al systems will likely come under inexorable stress from nefarious attacks using counter-AI techniques (e.g., data pollution, spoofing, false alarms, or tricking a system in order to reverse engineering algorithms), which might undermine the confidence in a network, creating new vulnerabilities, errors, and unintentional escalation risks. In the emerging deepfakes arms-race - much like cybersecurity more broadly - detection software will likely lag behind advances in offensive enabling solutions - or offense-dominant ones.130 According to computer science expert Hany Farid, there are probably 100 to 1,000 times 'more people developing the technology to manipulate content than there is to detect [it].'131 Efforts to counter these technologies and regain the upper-hand in the war on digital fakery have been underway for several years (e.g., automated anti-virus software, encryption, and other tools to call-out manipulated images and videos). Today, more sophisticated Al-enhanced techniques coupled with research that suggests pre-existing cognitive schemes, beliefs, and attitudes rather than credulity or gullibility determines whether the public believes particular fakery is real or not - meaning fake images and videos can achieve accepted even though they can be easily debunked.132

Al-enhanced fake news, deepfakes, bots, and other malevolent social media campaigns could also influence public opinion - creating false narratives or amplifying false alarms - with destabilizing effects on a mass scale, especially in times of geopolitical tension and internal strife.'33 In 2017, for example, a deepfake video was circulated on Russian social media, alleging a U.S. B-52 bomber had accidentally dropped a 'dummy nuclear bomb' on a Lithuanian building.'34 A state or non-state actor could, for instance, generate an image or recording of a military commander obtained from open-sources to generate and disseminate a deepfake containing false orders, intelligence, or geospatial imagery that at best generates confusion that, in the worst-case scenario, aggravates a tense situation or crisis between rival nuclear powers - this idea is examined below. For example, in the world's first reported Al- enabled theft in 2019, non-state actors used Al voice mimicking software to generate a fake recording of a British energy executive.'35 In short, Al- augmented technology is rapidly becoming another capability in non-state actors' toolkit to wage campaigns of disinformation and deception - one that both sides may have used against them.'36

The democratization of ever more sophisticated technology will likely amplify documented human pathologies,137 explaining people's attraction to novel and negative information, memes, falsehoods, and filter bubbles - the so-called 'information cascade' phenomenon - which deepfakes are adapt to perpetuate.138 In the case of cognitive availability heuristics, people tend to skew their judgments to more recent (or memorable) information, situations, or experiences, thus making new or novel opinions biased toward those that can be more easily recalled.139 During times of high-pressure crises, decision-makers tend to interpret unusual circumstances as threatening, even if an adversary's behavior has not changed. Routine activities (e.g., troop movements) scrutinized in the context of an early-warning alert may be considered as more menacing than they might otherwise be.140

In 2017, for example, South Korean counterintelligence officials received fake mobile and social media alerts with orders for U.S. military and DoD personnel to evacuate the Korean Peninsula.141 Information attacks such as this suggest that non-state actors, state proxy actors - and perhaps state actors - will inevitably attempt to use social media as a tool of war to provoke nuclear confrontation for political-ideological, religious, or other malevolent goals; and with increasing levels of sophistication, stratagem, and Al-enhanced subterfuge.142 Al might also enable non-state actors to automate, accelerate, and scale synthetic social media accounts and content to support malevolent disinformation operations.143

These potentially catalytic dynamics will likely be compounded by human cognitive bias. A motivated non-state actor would be well-positioned to use Al-augmented tools (e.g., 'fake news' or deepfake generated propaganda) to exploit this psychological weakness, ensuring control of the dissemination of false narratives and opinions. Threat assessments of catalytic escalation must, therefore, include both the likelihood that a third-party can execute a particular attack and the probability that the falsehoods or fakery that transpires leads to escalation.144 Strategic decisions made by isolated nuclear-armed authoritarian regimes under these circumstances - especially regimes that believed their survival was threatened - could trigger dangerous accidental nuclear escalation dynamics.

## answers to science research da

### 2ac vs science research da

#### 1. Non-unique—science research and innovation are declining

Lawler and Colleen 23 (Daniel Lawler, health and science reporter for AFP. Juliette Colleen, reporter for AFQ. “Rate of scientific breakthroughs slowing over time: Study” 1/4/2023. Accessed 5/29/2024. https://phys.org/news/2023-01-scientific-breakthroughs.html)

The rate of ground-breaking scientific discoveries and technological innovation is slowing down despite an ever-growing amount of knowledge, according to an analysis released Wednesday of millions of research papers and patents.

While previous research has shown downturns in individual disciplines, the study is the first that "emphatically, convincingly documents this decline of disruptiveness across all major fields of science and technology," lead author Michael Park told AFP.

Park, a doctoral student at the University of Minnesota's Carlson School of Management, called disruptive discoveries those that "break away from existing ideas" and "push the whole scientific field into new territory."

The researchers gave a "disruptiveness score" to 45 million scientific papers dating from 1945 to 2010, and to 3.9 million US-based patents from 1976 to 2010.

From the start of those time ranges, research papers and patents have been increasingly likely to consolidate or build upon previous knowledge, according to results published in the journal Nature.

The ranking was based on how the papers were cited in other studies five years after publication, assuming that the more disruptive the research was, the less its predecessors would be cited.

The biggest decrease in disruptive research came in physical sciences such as physics and chemistry.

"The nature of research is shifting" as incremental innovations become more common, senior study author Russell Funk said.

Burden of knowledge

One theory for the decline is that all the "low-hanging fruit" of science has already been plucked.

If that were the case, disruptiveness in various scientific fields would have fallen at different speeds, Park said.

But instead "the declines are pretty consistent in their speeds and timing across all major fields," Park said, indicating that the low-hanging fruit theory is not likely to be the culprit.

Instead, the researchers pointed to what has been dubbed "the burden of research," which suggests there is now so much that scientists must learn to master a particular field they have little time left to push boundaries.

This causes scientists and inventors to "focus on a narrow slice of the existing knowledge, leading them to just come up with something more consolidating rather than disruptive," Park said.

#### 2. No link—restricting commercial ai use doesn’t spill over to research

AAP 23 (Association of American Publishers. “Artificial Intelligence and Copyright” comments before the US Copyright Office by the Association of American Publishers 10/30/2023. Accessed 6/4/2024. https://www.regulations.gov/comment/COLC-2023-0006-9070) wtk

While a few jurisdictions have adopted a copyright exception for text-and-data mining, on the premise that it is necessary to AI research and development, AAP does not believe any exceptions to permit the unlicensed use of copyrighted works to develop AI technologies are necessary or desirable. We note that TDM exceptions were adopted before the rise of Gen AI, and the original aim was to facilitate computational analysis of large amounts of text as may be embodied in scholarly and scientific articles for research purposes. 10 TDM uses do not equate with the use of copyrighted works to train Gen AI models. Licensing solutions remain the better tool for facilitating AI development. Through licensing arrangements, AI systems developers can legitimately access the copyrighted works necessary to training trustworthy and reliable AI models while authors, publishers, and other copyright holders and licensees are appropriately compensated for the use of their works.

Although we do not believe a TDM exception is necessary, below we outline the approach taken in four jurisdictions of which the Office should be aware. We note that the TDM exception defined in the EU Directive on copyright in the Digital Single Market (DSM) differentiates TDM for commercial and non-commercial purposes (specifically, for scientific research). The exception for commercial TDM activity applies unless the rights holder prohibits the use of their copyrighted works by expressly reserving their rights “in an appropriate manner, such as machine-readable means in the case of content made publicly available online.” 11 Lawful access to the copyrighted works is required.

#### 3. No link—Licensing solves TDM research

Kupferschmid 23 (Keith Kupferscmid, CEO of the Copyright Alliance, “Artificial Intelligence and Copyright: Comments of the Copyright Alliance” comment before the U.S. Copyright Office. 10/30/2023. Accessed 5/25/2024 from https://www.regulations.gov/comment/COLC-2023-0006-8935) wtk

There is already a high demand for corpuses of copyrighted works to train AI systems, and copyright owners have already entered into licensing agreements (or are offering licenses) for text and data mining (TDM) uses.155 The licensing activity in the TDM markets (which we discuss in more detail in responses to question six and its subparts) is evidence of existing markets for the use of copyrighted works for AI training and development, and it is important that the conditions of those licenses are respected and that they are not undermined by new, unwarranted exceptions that excuse unauthorized uses. In contrast to earlier forms of AI, these new generative AI models are ingesting copyrighted works to generate works that compete in the same market as the ingested works. In some cases, the output could qualify as derivatives of the ingested, copyrighted works.156 Preserving opportunities for licensing and authorization of copyright works for ingestion could help to mitigate or prevent harm to copyright owners and creators arising from AI output that supersedes or supplants the market for the ingested works.

#### 4. No Internal Link—Research isn’t key to scientific progress or innovation

Park et al. 23 (Michael Park, PhD from University of Minnesota, Carlson School of Management. Erin Leahey is Professor of Sociology at the University of Arizona. Russell J. Funk is an associate professor in the Strategic Management and Entrepreneurship group at the University of Minnesota's Carlson School of Management. “Papers and patents are becoming less disruptive over time” *Nature* 613, 138–144 (2023). https://doi.org/10.1038/s41586-022-05543-x) wtk

We also considered how declining disruptiveness relates to the growth of knowledge (Extended Data Fig. 9). On the one hand, scientists and inventors face an increasing knowledge burden, which may inhibit discoveries and inventions that disrupt the status quo. On the other hand, as previously noted, philosophers of science suggest that existing knowledge fosters discovery and invention3,6,7. Using regression models, we evaluated the relationship between the stock of papers and patents (a proxy for knowledge) within fields and their CD5 (Supplementary Information section 3 and Supplementary Table 2). We find a positive effect of the growth of knowledge on disruptiveness for papers, consistent with previous work20; however, we find a negative effect for patents.

Given these conflicting results, we considered the possibility that the availability of knowledge may differ from its use. In particular, the growth in publishing and patenting may lead scientists and inventors to focus on narrower slices of previous work18,46, thereby limiting the ‘effective’ stock of knowledge. Using three proxies, we document a decline in the use of previous knowledge among scientists and inventors (Fig. 6). First, we see a decline in the diversity of work cited (Fig. 6a,d), indicating that contemporary science and technology are engaging with narrower slices of existing knowledge. Moreover, this decline in diversity is accompanied by an increase in the share of citations to the 1% most highly cited papers and patents (Fig. 6a (i),d(i)), which are also decreasing in semantic diversity (Fig. 6a (ii),d (ii)). Over time, scientists and inventors are increasingly citing the same previous work, and that previous work is becoming more topically similar. Second, we see an increase in self-citation (Fig. 6b,e), a common proxy for the continuation of one’s pre-existing research stream47,48,49, which is consistent with scientists and inventors relying more on highly familiar knowledge. Third, the mean age of work cited, a common measure for the use of dated knowledge50,51,52, is increasing (Fig. 6c,f), suggesting that scientists and inventors may be struggling to keep up with the pace of knowledge expansion and instead relying on older, familiar work. All three indicators point to a consistent story: a narrower scope of existing knowledge is informing contemporary discovery and invention.

#### 5. No impact—science doesn’t prevent extinction

Ellis 18 (Erle C. Ellis is a professor of geography and environmental systems at the University of Maryland, Baltimore County. “Science Alone Won’t Save the Earth. People Have to Do That.” 8/11/2018. Accessed 6/3/2024. https://www.nytimes.com/2018/08/11/opinion/sunday/science-people-environment-earth.html) wtk

One thing is for sure. A better future won’t be realized through unquestioning faith in the safety of scientifically defined environmental limits or in unlimited technological capacities to avoid environmental consequences. When there is no single optimal solution, no amount of rational debate, or even computational intelligence, can find one. Science does not, cannot and should not have all the answers — not for earth’s limits, nor for human futures. A future governed solely by rationality and scientific evidence offers no safe space in these times.

The problem is not us; it is that there is no “us.” Just as one future will never be best for all people, no single way of thinking, believing or acting will ever be enough to forge our better futures together on this one planet. Decisions informed by scientific evidence will, of course, create better outcomes for people and the planet. But no amount of scientific evidence, enlightened rational thought or innovative technology can resolve entirely the social and environmental trade-offs necessary to meet the aspirations of a wonderfully diverse humanity — at least not without creating even greater problems in the future.

For this reason and others, putting expert scientific narratives at the center of decision making, like “nonnegotiable environmental limits,” rather than focusing on opportunities for collective betterment, has led only to increasing divisions over which experts to trust. If we are to continue improving the human world, while retaining the nature we love, it will be necessary to get beyond polemics and expertise, scientific or otherwise. In the end, it is people, and their institutions — not science — that will decide the future.

### 1ar licensing solves

#### Licensing is frequently used now for scientific research

Rowland and Kaufman 23 (Catherine Zaller Rowland, VP and General Counsel for the Copyright Clearance Center. Roy S. Kaufman, Managing Director of Business Development for the Copyright Clearance Center. “Artificial Intelligence and Copyright” comments before the US Copyright Office by Copyright Clearance Center. 10/30/2023. Accessed 6/4/2024. https://www.regulations.gov/comment/COLC-2023-0006-8601) wtk

10. If copyright owners’ consent is required to train generative AI models, how can or should licenses be obtained?

Copyrighted materials can be licensed for AI use directly from rightsholders and collectively through collective management organizations such as CCC. In science publishing and in other fields where open licenses are used, copyright owners frequently license reuse for AI under the terms of such licenses.

10.1. Is direct voluntary licensing feasible in some or all creative sectors?

Direct voluntary licensing is feasible and commonly utilized in text and images. Among many others, the prominent copyright holders Associated Press, Getty Images and vAIsual all offer licenses. We offer no opinion as to other sectors.

10.2. Is a voluntary collective licensing scheme a feasible or desirable approach? Are there existing collective management organizations that are well-suited to provide those licenses, and are there legal or other impediments that would prevent those organizations from performing this role? Should Congress consider statutory or other changes, such as an antitrust exception, to facilitate negotiation of collective licenses?

CCC already offers market-based, global non-exclusive voluntary licenses to support AI in the commercial research, schools, and education technology sectors. These licenses were built with rightsholders and users based on agreed understandings of needs and market conditions. We are well suited to provide and have proven experience of providing these and other AI licenses for the text sectors at a minimum.

### 1ar research not key

#### Research isn’t key—it overstretches scientists

Wilcox 23 (Christie Wilcox, science journalist, “ScienceAdviser: Scientists are publishing too many papers—and that’s bad for science” 11/16/2023. Accessed 5/29/2024. https://www.science.org/content/article/scienceadviser-scientists-are-publishing-too-many-papers-and-s-bad-science) wtk

Academia is often described as a “publish or perish” environment. So much of a scientist’s career is measured in research papers—generally the more, the merrier. And that’s harming the scientific enterprise as a whole, not to mention researchers’ mental and physical health, according to the authors of a recent arXiv preprint.

In recent years, the number of papers being published has “grown exponentially,” the team explains. In 2016, about 1.92 million papers were indexed by the Scopus and Web of Science publication databases. In 2022, that number had jumped to 2.82 million. And this leap happened even as the number of newly awarded PhDs leveled off and declined. That means that, on average, each scientist is writing, editing, and reviewing more papers, they say—a problem they dub “the strain on scientific publishing.”

According to the team’s data, a handful of publishers—MDPI, Elsevier, Frontiers, Springer-Nature, and Wiley—account for more than 70% of the increase in articles per year. In fact, MDPI alone is responsible for 27% of the increase. There’s a single reason for that, according to a Bluesky post from first author Mark Hanson: “I could be nuanced (it's in the paper!). But let’s be frank: it’s special issues.”

Special issues, where researchers are encouraged to contribute papers related to a topic, used to be few and far between. Now, they’re the primary way research is published in MDPI and Frontiers journals, Hanson and colleagues write. And special issues have lower rejection rates and faster turnaround times—potential indicators that the papers are lower quality, the team says. “We’re all overworked, and we can’t let this go on. Our metrics tell us this growth isn’t rigorous science,” Hanson writes in another post .

Why the push for so many papers? Hanson and colleagues didn’t have access to revenue data for the publishers they examined, but they suspect it’s profit-driven. A back-of-the-envelope calculation from co-author Dan Brockington estimates that MDPI made roughly twenty times as much from author processing charges in 2021 than in 2015. And a recent analysis in Quantitative Science Studies found that article processing charges for making papers open access drove massive income gains for for-profit publishers ( Science is published by the non-profit AAAS ) in recent years. The authors of that paper estimated that scientists paid more than $1 billion in open-access article fees to five of the biggest publishers—Elsevier, Sage, Springer-Nature, Taylor & Francis, and Wiley—between 2015 and 2018. And they note that these companies have “an obscenely high profit margin,” pointing to previous calculations.

Unsurprisingly, MDPI and other publishers disagree with Hanson et al.’s analyses. A representative from MDPI told El Pais that the data is “questionable and imprecise” and stood by the increase in special issues, as it “reflects our dedication to advancing scholarly publishing and meeting the changing needs of the research community. ” A spokesperson for Frontiers similarly defended special issues, saying the analyses have “notable limitations and possible biases” and that Frontiers journals “ensure meticulous scrutiny of scholarly work before publication.”

Regardless of the motivations behind special issues or the increase in paper production, the strain on scientists is real, Hanson and colleagues say, and reducing it will require fundamental changes to academic culture and scientific publishing. “If this continues, we may lose the most important thing about science, which is scientific rigor,” Pablo Gómez Barreiro, another of the paper’s co-authors, tells El Pais. “If you lose that, you lose everything.”

#### Bogus papers undermine the application and trust of research

McKie 24 (Robin McKie is science and environment editor for the Observer. “‘The situation has become appalling’: fake scientific papers push research credibility to crisis point” The Observer 2/3/2024. Accessed 5/29/2024. https://www.theguardian.com/science/2024/feb/03/the-situation-has-become-appalling-fake-scientific-papers-push-research-credibility-to-crisis-point) wtk

Tens of thousands of bogus research papers are being published in journals in an international scandal that is worsening every year, scientists have warned. Medical research is being compromised, drug development hindered and promising academic research jeopardised thanks to a global wave of sham science that is sweeping laboratories and universities.

Last year the annual number of papers retracted by research journals topped 10,000 for the first time. Most analysts believe the figure is only the tip of an iceberg of scientific fraud.

“The situation has become appalling,” said Professor Dorothy Bishop of Oxford University. “The level of publishing of fraudulent papers is creating serious problems for science. In many fields it is becoming difficult to build up a cumulative approach to a subject, because we lack a solid foundation of trustworthy findings. And it’s getting worse and worse.”

The startling rise in the publication of sham science papers has its roots in China, where young doctors and scientists seeking promotion were required to have published scientific papers. Shadow organisations – known as “paper mills” – began to supply fabricated work for publication in journals there.

The practice has since spread to India, Iran, Russia, former Soviet Union states and eastern Europe, with paper mills supplying ­fabricated studies to more and more journals as increasing numbers of young ­scientists try to boost their careers by claiming false research experience. In some cases, journal editors have been bribed to accept articles, while paper mills have managed to establish their own agents as guest editors who then allow reams of ­falsified work to be published.

“Editors are not fulfilling their roles properly, and peer reviewers are not doing their jobs. And some are being paid large sums of money,” said Professor Alison Avenell of Aberdeen University. “It is deeply worrying.”

The products of paper mills often look like regular articles but are based on templates in which names of genes or diseases are slotted in at random among fictitious tables and figures. Worryingly, these articles can then get incorporated into large databases used by those working on drug discovery.

### 1ar science doesn’t solve extinction

#### No science impact—it isn’t reflected by policy choices

Leshner 21 (Alan I. Leshner is chief executive officer, emeritus, of the American Association for the Advancement of Science. He previously has served as the director of the National Institute on Drug Abuse at the National Institutes of Health. “Trust in Science Is Not the Problem” *Issues in Science and Technology* Volume 37, Number 3, Spring 2021. Accessed 6/5/2024. https://issues.org/trust-in-science-is-not-the-problem-engagement-leshner/) wtk

It is clear, then, that the public continues to place a high degree of trust in science and scientists. But this does not mean people will follow science-based recommendations on specific issues, and that can be a problem. In fact, a 2020 Pew survey showed that fewer than half of Americans believed they should rely primarily on “experts” to tell them how to deal with various societal issues. Although it can be disquieting to scientists, this should not be surprising. Indeed, science and public policy experts have long taught that important decisions, such as policy decisions, are rarely, if ever, made solely on the basis of science, but are based on both facts and values, or on facts and personal experience. Moreover, long-held beliefs or core values often win out over scientific evidence when policy decisions are being made.

The bottom line is that the public makes decisions based on an array of inputs, including but not limited to scientific facts. This is particularly true when the issue or problem is controversial—as demonstrated in another Pew survey that compared the views held by scientists who were members of the American Association for the Advancement of Science with those of a sample of the general population about a variety of contentious issues. According to that survey, AAAS members were far more likely to hold beliefs consistent with the preponderance of scientific evidence than were members of the broader public. However, the amount of disagreement varied substantially across topics. A large gap occurred, for example, between scientists and the general public concerning the belief that humans evolved over time, where 98% of scientists believed in evolution while only 65% of the public did so. A smaller gap occurred in belief about the safety of vaccines such as MMR, where 86% of scientists thought them safe compared with 68% of the public.

Some failure to follow scientific recommendations results from a broader lack of understanding about the nature and processes of science. For example, there typically is some uncertainty in scientific evidence, and there are very few situations where all scientists agree about what the data are showing or how the data should be interpreted. Scientists understand that there can be scientific consensus behind the recommendations they make to the public in spite of some uncertainty and disagreement—but that can be disquieting to nonscientists. Another example is the evolving nature of scientific theories. Scientific theories often are revised or even replaced as additional information is acquired. Scientists accept these revisions as a normal part of the scientific process, whereas the public may consider them signs of lack of authority or expertise.

In the same vein, it also is important to acknowledge that not all science is equally definitive. Some scientific theories are weaker than others and some data sets are more reliably interpreted than others. Patients, for example, often seek second opinions on their clinical tests or recommendations because even experts can disagree on the meanings of their findings. Such disagreements or lack of certainty become particularly problematic when people—scientists or nonscientists—distort or deny the meaning of data or findings on topics, such as the effectiveness of vaccines or the evidence for human-caused climate change, that have been more completely established.

## answers to opt-out cp

### 2ac vs opt out cp

#### 1. Permutation: do both—we can allow people to opt out of AI training AND require AI companies to get licenses

#### 2. Opt-out fails—lack of technology

AAP 23 (Association of American Publishers. “Artificial Intelligence and Copyright” comments before the US Copyright Office by the Association of American Publishers 10/30/2023. Accessed 6/4/2024. https://www.regulations.gov/comment/COLC-2023-0006-9070) wtk

An “opt out” approach turns the copyright framework on its head. Copyright secures exclusive rights to the copyright owner and determinations regarding how and by whom the rights holder’s works are used lie solely with the copyright holder (see also Response to Question 9).

In addition, there are practical issues with an opt out approach. As Gen AI developers have routinely sourced books, scientific articles and other copyrighted materials from piracy sites, relying on technical tools to prevent scraping and collection of copyrighted works as training material under an opt out approach would be ineffective. Piracy site operators reproduce and distribute copyrighted works without the consent of authors and publishers, and publishers do not have an opportunity to embed the technological tools that would enable an opt out.

Furthermore, there is as yet no technology that can detect or flag when an AI-bot disregards or ignores opt out instructions or “do not train/scrape” tags. Thus, absent a transparency requirement that identifies and discloses the copyrighted works used for training purposes, rights holders would be unable to verify compliance with the opt-out instructions.

There is also a risk that adequate tools for an “opt out” approach would not be developed in a timely fashion, would not be accessible to large and small copyright holders alike, would not be adequately implemented, or would impair online access/discoverability for purposes unrelated to AI training. The inability to develop “standard technical measures” under 17 U.S.C. § 512(i)(1)(B) in the 25 years after the DMCA went into effect serves as a cautionary tale.63

#### 3. It would be impossible to detect copyright infringement

Coffey and Smith 23 (Danielle Coffey, President and CEO of News/Media Alliance. Regan Smith, Senior Vice President and General Counsel for News/Media Alliance. “Artificial Intelligence and Copyright” Comments of the News/Media Alliance before the U.S. Copyright Office. Docket No. 2023–6. 10/30/2023. Accessed 5/21/2024. <http://www.newsmediaalliance.org/wp-content/uploads/2023/10/Final-NMA-USCO-AI-NOI-Response-Submitted-10.30.23.pdf>) wtk

Publishers are also concerned that opt-out systems must be efficient at scale. In an opt-out regime, developers may have incentives to make opt-out difficult for publishers (or at least, expend the minimum compliance efforts required), whereas with an opt-in regime, developers are incentivized to seek efficient licensing solutions. For example, DALL-E’s opt-out system requires the “owner or rights holder . . . to submit an individual copy of each image they'd like removed from DALL-E's training dataset, along with a description.” 166 This is obviously impractical for more than a de minimis number of images. The Copyright Office need only recall its years-long DMCA study to predict the difficulties with this system.

Moreover, an opt-out regime puts the burden on copyright owners to find out who is using their material. Not only does this incentivize non-disclosure, but developers commonly train their systems on material acquired from sites that have been identified by the U.S. government as notorious markets for piracy,167 necessitating that copyright owners enforce rights against infringers as a prerequisite—a burden that is impossible to achieve.

#### 4. Permutation: do the counterplan

#### 5. Links to the net benefit because it still allows people to prevent their work from being used to train AI

### 1ar opt out fails—detection

#### The cp empirically fails and puts the onus on publishers to detect

Ortiz 23 (Karla Ortiz, Concept Artist, Illustrator, and Fine Artist, “Artificial Intelligence and Intellectual Property – Part II: Copyright” expert testimony before the US. Senate Judiciary Subcommittee on Intellectual Property. 7/7/2023. Accessed 5/16/2024. https://www.judiciary.senate.gov/download/2023-07-12-pm-testimony-ortiz) wtk

There is also a push in the generative AI industry to set opt-out (choose not to participate in something) as a standard. I must note that presently, creative persons like myself have never been offered the opportunity to truly opt our work out of training sets, even though this narrative has persisted. Currently, techniques to “unlearn” specific samples of data are extremely costly even for moderate or small models, and completely intractable for state of the art generative AI models. Due to this it is near impossible for algorithms to forget data once it’s been trained upon, outside of destroying the algorithms and training sets and starting from scratch.12 Thus, while Machine “Unlearning” is still a nascent stage, current “optout” procedures are largely ineffective and unreliable.

Even if opt-out procedures were perfected, this still misses the mark. Opting out puts the onus on the public to police the use of their data. Given the multitudes of AI models that have proliferated in just one short year, asking a member of the public to constantly monitor the use of their own data by AI companies that are constantly multiplying seems to me to place the burden on the wrong party. For starters, it places an undue burden on people who may not be well versed with the technology, people who may not know the language, people who may have physical impediments, people who may have time limitations, or people who are not on the internet, are unaware their data is in the possession of technology companies and being used for commercial profit.

On a practical level opt-out makes no sense. There is a serious resource imbalance between ordinary citizens and Generative AI companies. Generative AI companies often are well-funded, and generate millions if not billions in profit. Even though they profit richly, Generative AI companies do not want to spend any of those profits in using data responsibly, they instead want to shift the responsibility onto the ordinary citizen. These models are trained upon an immense amount of data, and models are constantly updated. Does this mean the public has to keep up to date every time a new update to a model comes out? What about every time you publish something new online, does one have to verify if it is used by Generative AI companies? What about when third party users upload data without your consent? How does one keep track if their work was utilized? How does one find said work in the millions of billions of data points? What if you miss the opt-out period, and then become an unwilling participant? Technology companies cannot claim ownership of all the data on the internet that exists for our benefit without paying heed to copyright or privacy rights. Opt-out would be an ineffective and inappropriate standard for most if not all consumer facing industries, Generative AI should not be the exception. The onus should be on the technology companies who devote their immense wealth to build these models to ensure they are doing it fairly and ethically, not on the common citizen who may not even be aware their data has been taken. Explicit opt-in is the only way forward, as it helps ensure that all AI/ML models are built with right holders and the public’s explicit authorization for commercial use of their data from the beginning.

#### The cp fails—it excludes content from search traffic and can’t detect copyright infringement

Coffey and Smith 23 (Danielle Coffey, President and CEO of News/Media Alliance. Regan Smith, Senior Vice President and General Counsel for News/Media Alliance. “Artificial Intelligence and Copyright” Comments of the News/Media Alliance before the U.S. Copyright Office. Docket No. 2023–6. 10/30/2023. Accessed 5/21/2024. <http://www.newsmediaalliance.org/wp-content/uploads/2023/10/Final-NMA-USCO-AI-NOI-Response-Submitted-10.30.23.pdf>) wtk

The scraping of publisher websites is systematic and generally takes place without a license or authorization, in violation of publishers’ terms of service, and with no real way for publishers to opt out of such scraping. Even where opt-out measures are offered or respected, they are insufficient at best. While some developers now provide publishers with the option to opt out, this is not a common practice and such opt-outs only apply to the specific developer in question, making opting out impractical and burdensome for media publishers. Similarly, while some developers have indicated that publishers can use robots.txt exclusion protocol going forward to indicate their unwillingness to be scraped for AI training purposes, the use of the protocol has traditionally meant being excluded from even simple search results by search engines—reducing publishers’ visibility and discoverability to the public. There is also no requirement for developers to comply with the voluntary opt-out signal or for scrapers to accurately identify themselves, allowing bad actors to continue scraping publisher content without authorization. Further, and more fundamentally, publishers should not have to affirmatively opt out from generative AI uses to prevent the commercial consumption of their protected material—it is antithetical to the guiding principles of U.S. copyright law and the exclusive rights afforded to rightsholders. Such opt-out solutions are also “too little, too late,” considering the vast scraping and copying of publisher content that has already taken place to bring generative AI models to the point of commerciality.

### 1ar opt out fails—tech

#### Only the plan solves—too many technical barriers to opt-outs

AAP 23 (Association of American Publishers. “Artificial Intelligence and Copyright” comments before the US Copyright Office by the Association of American Publishers 10/30/2023. Accessed 6/4/2024. https://www.regulations.gov/comment/COLC-2023-0006-9070) wtk

As noted in the preceding section, the progress for developing the technical tools to implement an opt-out mechanism has been slow, and the tools may prove to have limited scope/effectiveness. As a practical matter, the absence of technologies that would alert a rights holder when its opt-out instructions or “do not train/do not scrape” tags are disregarded, coupled with the lack of transparency as to which copyrighted works have been acquired by a Gen AI systems developer or dataset creator without a license, would result in rights holders having little recourse when their exclusive rights have been violated.

At the same time, it is feasible to obtain in advance rights holder consent to use their copyrighted works for training, and this is what the law requires. Licensing is occurring across all publishing sectors. The claim that the volume of works used for training makes it burdensome for a Gen AI systems developer to seek permission is not an excuse for infringing on the copyrights and livelihoods of the thousands of authors, publishers, and other artists. Copyrighted works are as integral to Gen AI technologies as they are to services like Netflix and Spotify, both of which license the works made available through their platforms. Developers of Gen AI are capable of doing the same and rights holders are willing to work with them to effect such licensing. Given the valuations of AI developers — for example, it has been reported that OpenAI could see its valuation at an estimated $80 billion66 — licensing deals with the publishing industry would represent a fraction of the cost. To give a sense of scale, AAP has approximately 130 members across trade, education, and professional and scholarly publishing. For training purposes in any given sector, negotiating licenses with some percentage of the industry is quite feasible. Voluntary collective licensing approaches may also emerge. (See Response to Question 10.2.)

Publishers invest significant resources in bringing authors’ works to market — taking risks on new works from new and existing authors. These investments in professional writing, peer-reviewed reporting of scientific research, and separating fact from fiction will only be more important as Gen AI technologies indiscriminately trained on unvetted material accelerate the spread of misinformation and bias. Requiring Gen AI developers to obtain licenses is essential to ensuring that publishers continue to have proper incentives to invest in new works that benefit society as a whole.

### 1ar opt out fails—search traffic

#### It cuts off search traffic to publishers

Coffey 24 (Danielle Coffey, President & CEO of News/Media Alliance. “Oversight of A.I.: The Future of Journalism” Questions for the Record, part of the U.S. Senate Committee on the Judiciary’s Subcommittee on Privacy, Technology, and the Law. 1/10/2024. Accessed 5/20/2024. https://www.judiciary.senate.gov/download/2024-01-10-qfr-responses-coffey) wtk

Some AI developers, including OpenAI and some others, have provided publishers with opt-out measures, including honoring the robots.txt exclusion protocol on a go-forward basis, and most publishers also prohibit AI scraping through their terms of service. Increasingly many companies have opted out or blocked AI web crawlers—over the course of three weeks in late September at least 250 top websites blocked OpenAI’s GPTBot while 14 percent of the 1,000 most popular websites block Common Crawl’s CCBot.

Overall, opt-out measures, including technical measures such as robots.txt, are blunt and flawed instruments when it comes to protecting publishers from infringement in practice. Without cooperation from GAI developers, there is no easy, standardized way to block scraping for AI purposes. And blocking for AI training can often have the undesirable effect of also blocking crawling for search and other desirable, mutually beneficial uses. There is also no requirement for developers to comply with the opt-out signal or for scrapers to accurately identify themselves, allowing bad actors to continue scraping publisher content without authorization. Further, and more fundamentally, publishers should not have to affirmatively opt out from generative AI uses to prevent the commercial consumption of their protected material—it is antithetical to the guiding principles of U.S. copyright law and the exclusive rights afforded to rightsholders. Such opt-out solutions are also “too little, too late,” considering the vast scraping and copying of publisher content that has already taken place to bring GAI models to the point of commerciality.

### 1ar opt out fails—onus

#### Err heavily aff—the cp is a big-tech ruse

Coffey and Smith 23 (Danielle Coffey, President and CEO of News/Media Alliance. Regan Smith, Senior Vice President and General Counsel for News/Media Alliance. “Artificial Intelligence and Copyright” Comments of the News/Media Alliance before the U.S. Copyright Office. Docket No. 2023–6. 10/30/2023. Accessed 5/21/2024. <http://www.newsmediaalliance.org/wp-content/uploads/2023/10/Final-NMA-USCO-AI-NOI-Response-Submitted-10.30.23.pdf>) wtk

Discussions around opt-out are more relevant in countries and regions that, unlike the United States, may already have a statutory text and data mining (TDM) exception that allows some or all users to engage in TDM for limited AI training purposes. It is rare indeed to have a sweeping exception for TDM that extends to highly commercial uses without the ability to opt out. In addition to raising other potential concerns, including compliance with international agreements, retaining the ability to opt out of such exceptions is important in those countries or regions. The United States, however, has not adopted an exception to our copyright laws for TDM. N/MA opposes the creation of a new or expanded exception to copyright law that would change the status quo to permit AI training without the rightsholder’s authorization.

To date, current tools present a Potemkin village of a solution, providing limited benefits to publishers while creating a patina of responsibility to justify positions that copying is legal absent affirmative opt-out. It is inappropriate industrial policy to place the burden on a copyright owner to remedy a potentially infringing act, rather than on a generative AI developer or deployer who already possesses the right and ability to control what material is used for training (whether by selecting, cleaning, or fine tuning a dataset, licensing content, or by paying a low wage to someone overseas to mitigate the worst violations). And the necessary act of choosing what copyrighted works an AI system is trained on distinguishes these developments from the architectures that gave rise to the section 512 safe harbor.