Identifying the Economic Implications of Artificial Intelligence for Copyright Policy

Context and Direction for Economic Research

Edited by Brent Lutes, *Chief Economist, United States Copyright Office*

February 2025

Identifying the Economic Implications of Artificial Intelligence for Copyright Policy

Context and Direction for Economic Research

Edited by Brent Lutes, *Chief Economist, United States Copyright Office*

February 2025

Suggested Citation: Lutes, Brent A. ed., *Identifying the Economic Implications of Artificial Intelligence for Copyright Policy: Context and Direction for Economic Research*, U.S. Copyright Office, 2025.

Contents

Contributors

Editor

Brent Lutes *(U.S. Copyright Office)*

Contributing Authors

Joshua Gans *(University of Toronto)*

Shane Greenstein *(Harvard Business School)*

Adam Jaffe *(Brandeis University)*

Brent Lutes *(U.S. Copyright Office)*

Abhishek Nagaraj *(University of California, Berkeley)*

Imke Reimers *(Cornell University)*

Michael D. Smith *(Carnegie Mellon University)*

Rahul Telang *(Carnegie Mellon University)*

Catherine Tucker *(Massachusetts Institute of Technology)*

Joel Waldfogel *(University of Minnesota)*

Additional Contributors

Michael Palmedo *(U.S. Copyright Office)*

Ryan Safner *(U.S. Copyright Office)*

Acknowledgements

The contributing authors would like to thank Andrew Toole, Chief Economist of the U.S. Patent and Trademark Office, and Ben Zhao, Neubauer Professor of Computer Science at University of Chicago, for their helpful insights.

Preface

The discussions in this edited volume reflect the work of an ad hoc committee of economic scholars convened by the U.S. Copyright Office, through its Office of the Chief Economist, to discuss the numerous economic issues at the intersection of artificial intelligence and copyright policy. The committee engaged in several months of substantive discussions, consultation with technical experts, and research, culminating in a daylong roundtable event on January 23, 2024.

The mandate of this committee was to identify the most consequential economic characteristics of AI and copyright, and what factors may inform policy discussions and decisions. The goal of the present discussion is not to provide answers to specific questions or to identify optimal policy choices. Rather, it is to provide a structured and rigorous framework for considering economic evidence. That is, the goal is to help construct a scale upon which to weigh the evidence, leaving it to the broader economic research community to produce the empirical and theoretical evidence necessary to balance that scale.

This volume serves as a platform for articulating the ideas expressed by participants as part of the roundtable. All principal contributors submitted written materials summarizing the group's prior discussions on a particular topic, with editorial support provided by the Office of the Chief Economist. The discussions here presume the reader to have some advanced understanding of economic concepts and complexities; it is not aimed at a general audience. There are eight parts, with the first providing a general discussion and the rest addressing specific issues; many of these seemingly distinct issues are highly interrelated. There are differing reasonable viewpoints on certain issues, not all of which are discussed herein.

Finally, a disclaimer must be made. The many ideas and views that are discussed in this volume do not necessarily represent the views of every roundtable participant or their respective institutions. **The U.S. Copyright Office does not take a position on these ideas for the purposes of this project. The inclusion of any particular idea merely indicates that the idea was discussed by roundtable participants and should not be taken as an endorsement of that idea or opinion by the Copyright Office, the principal contributors of a particular section, or any individual participants. Additionally, all contributing authors are economists and not lawyers. All discussions here are thus from an economic perspective; terms and concepts used here reflect economic definitions, and not legal, industry, or lay-person definitions, unless otherwise noted.**

Suggested Citation: Lutes, Brent A. ed., *Identifying the Economic Implications of Artificial Intelligence for Copyright Policy: Context and Direction for Economic Research*, U.S. Copyright Office, 2025.

1. Introduction

Principal contributor: Brent Lutes

The economic questions around artificial intelligence are vast and far-reaching.[1] The discussions here do not attempt to articulate all of them or provide unconditional answers to any of them. Rather, the intent is to explicate what we believe to be the most consequential economic questions in the specific context of copyright policy and how those questions might matter, and to provide a framework to apply the answers (for which we rely on the broader economic research community) to policy discussions. However, before turning to that task, some context is necessary. To that end, this part provides a primer on the economics of copyright, the scope of our discussion (including clarification of certain, otherwise ambiguous, terms), and some basic technological context.

For ease of exposition, the remainder of the volume is divided into seven parts, each corresponding to particular sets of issues. Parts 2, 3, and 4 discuss matters relating to the output of generative AI systems. These include, respectively, the effects of competition from AI-generated works on human creators and human-generated works, the issues around AI-generated works that replicate human-generated works, and concerns relating to rights of publicity. The latter parts (5 through 8) discuss matters relating to the use of copyrighted works as an input to generative AI systems. These include, respectively, the effects ingestion of copyrighted works may have on creative incentives, the effects of access to copyrighted works on the technological and industrial development of AI, considerations relating to the control of access to copyrighted works for training purposes, and the unintended or peripheral effects of policies towards AI.

Many of these seemingly distinct issues are highly interrelated (for example, market power is a central theme for many of the issues discussed), and the decisions around any one issue may affect the balance for other issues. Thus, it is necessary to consider what follows in a holistic way and not base policy on a siloed view of any one issue.

1.1. Basic Economics of Copyright

The ultimate economic policy objective behind copyright is to enhance long-run social welfare by enabling the consumption of knowledge and creative works and facilitating scientific and cultural innovation. The consumption of creative works can increase contemporaneous social welfare as consumers gain useful knowledge, inspiration, entertainment, or other individual utility. However, beyond the contemporaneous welfare gains, the ability to consume creative works can also produce dynamic welfare gains as existing creative works are inputs in the production of new creative works — new works build upon what already exists (to the extent that creators can access existing works). Thus, enhancing creative output in one period can generate compounding benefits for future periods by increasing the rate of innovation.

1 For a primer, see e.g., Cuntz et al. (2024).

Nonetheless, because the *existence* of works is a prerequisite for the consumption of those works, we must ensure that creative works are produced at optimal quantity and quality levels and that the public can access those works. That is, copyright policy is intended to efficiently balance incentives to create and distribute works on the one hand, with the cost to consumers and future creators of accessing those works on the other — two factors that are mainly at odds with one another.

Copyright policy differs from policies toward more typical goods and services insofar as the economic ideal in those markets is usually to minimize market power to approximate a perfectly competitive market. The economic justification for a policy intervention expressly designed to diverge from that perfectly competitive archetype, at least in the case of copyright, stems from the public good-like characteristics of creative works. Namely, creative works tend to be "non-excludable"[2] and "non-rival"[3] — two features that usually result in market failure, absent intervention.

They are non-excludable as it is difficult, without legal intervention, to prevent someone from consuming the expression of an idea,[4] and they are non-rival as the consumption of a creative work typically does not diminish the ability of others to consume it. These characteristics lead to four relevant features of markets for creative works: (1) producers of original creative works face a relatively high fixed cost of production (e.g., developing a story, writing and recording a song, conducting research for an article, developing software, and so on); (2) copiers of original creative works face minimal fixed production costs, freeriding on the investments of the original creators; (3) both original producers and copiers face a relatively low marginal cost of reproduction (which is even more so the case in the age of digital distribution); and (4) absent legal intervention, copiers face few barriers to entering the market for creative works.

Because copiers incur lower total production costs (fixed costs plus marginal costs) than do original creators, and because copiers can freely enter the market, the market price for creative work will be driven to the copiers' marginal cost of production (absent legal intervention), which in the case of digital publication, is effectively zero. Depending on the speed with which this happens, original creators may be prevented from ever recouping their fixed production costs. This combination of factors may deprive creators of both the incentives and the means by which to produce the creative works, leading to market failure. To the extent this prevents the production of works that would otherwise generate value in excess of production costs, it reduces current social welfare. Moreover, to the extent that creative output reflects a cumulative process (creators build on existing creative works in the production of new works), diminishing the availability of creative works in one period may significantly reduce the potential gains in future welfare by slowing the innovation process.

2 "Non-excludable" refers to a characteristic of a good or service by which, after producing that good or service, it is difficult to prevent any particular consumer from consuming it. For example, national defense is a non-excludable service consumed by the people of a country irrespective of who may have or may not have paid for it.

3 "Non-rival" refers to a characteristic of a good or service by which the consumption of the good or service does not diminish its existence or affect the ability of others to consume that good or service. Digital products are often considered to be effectively non-rival insofar as they can be near infinitely consumed through costless copying without any degradation of the underlying information.

4 The expression of an idea (the intangible good) is distinct from the vehicle used to deliver that expression (the tangible good). The vehicle used to deliver an expression of an idea to consumers is typically a normal, excludable good, whereas the expression of the idea is typically non-excludable (to a degree). For example, it is relatively easy to, without any legal intervention, exclude someone from consuming a specific copy of a book, as such consumption requires physical possession of a specific, tangible, and unitary thing. In contrast, without legal intervention, it is difficult to prevent someone from consuming the content of that book, as there is a profit motive to make and distribute unauthorized copies.

1. Introduction

The excludability enabled by copyright law is one policy solution to the fixed cost recovery problem. It erects barriers for would-be competitors who wish to compete with an original creative work using copies of the work. These barriers confer a degree of market power to the rightsholder of an original work. If a work has a net positive value, that market power can allow the rightsholder to raise prices above marginal cost and potentially recover their fixed production costs, thus mitigating the inherent market failure.

However, conferring market power that allows excessively high prices (beyond what is necessary for full cost recovery) is typically not welfare maximizing.[5] Market power comes at a social cost insofar as it results in a less-than-optimal level of consumption. If we allow more of it than is necessary to solve the fixed cost recovery problem, the outcome will be inefficient. While this is true in the simple producer/consumer static model, the issue is greatly exacerbated by the fact that access to existing works is also a factor in the production of new works. High access prices can increase incentives to create through higher revenues for creators, but they can also increase production costs and thus *diminish* the incentives and capacity to create.

Limits on the degree of excludability (term limits or fair use exceptions, for example) can mitigate the pricing efficiency problem and enable a socially optimal level of access. Identifying the contours of the dynamically optimal balance between incentives (both intrinsic and extrinsic) to produce creative works on the one hand, and efficient access to those works on the other hand, is the implicit backdrop for the economic study of copyright policy.

Nonetheless, the optimal balance between those factors is not a stationary target. Throughout history, new technologies have shifted both the actual balance and what would be the optimal balance through changes in mediums of expression, production costs, distribution mechanisms, or enforcement capacity. Photography lowered the costs of portraiture, digital streaming cut physical distribution costs, file-sharing technology created a piracy crisis for the music industry, and so on. In addition to technological changes, changes in cultural attitudes, institutions, and consumption patterns also continuously alter the landscape for access and distribution. All of these factors change both the degree of market power that is *actually* conferred by existing laws and the amount of market power that is necessary to incentivize the production of new creative works. Artificial intelligence will likely act through many of those same channels, perhaps differing from historic precedents only in magnitude.

While some of the questions raised by artificial intelligence are obvious, others are not. It is not clear that we as a society, or as a research community, have fully articulated what these questions are, how to frame them, or how to begin to answer them in a scientifically rigorous way. The purpose of this volume is to begin to make such articulations and, in so doing, provide a framework for applying economic evidence to the policy debate and guidance for the economic research community on how to contribute useful evidence.

5 Note, "cost" here refers to the economic concept of cost, which is broader than the colloquial use of the term. In addition to explicit financial costs, it also incorporates the implicit opportunity costs faced by creators. Thus, the implication of this statement is simply that paying creators more than what is necessary to incentivize the socially optimal level of creative output is economically inefficient.

1.2. Scope of Discussion

The term “artificial intelligence” is somewhat nebulous, with a meaning that can differ drastically depending on context. In this discussion, we use it to refer to systems that ingest large amounts of (potentially copyrighted) materials so that a machine can identify relational and conditional patterns between data points in order to ultimately perform some complex human-like tasks without explicit instructions. One such task is the production of creative content (e.g., text, images, and sounds). AI systems aimed at producing such content are collectively referred to here as “generative AI” (U.S. Copyright Office, 2023). This subset of AI technology is the primary focus of this discussion.

Generative AI technologies, including text, image, video, and sound generation models, are powered by advanced machine learning techniques (discussed further in the next section). Notable commercially available examples include ChatGPT, Gemini, Copilot, LLaMA, DALL-E, Stable Diffusion, and Midjourney. The core functionality of these models involves generating content in response to prompts provided by users, facilitating widespread commercial and personal use.

These models can take many forms and are likely to develop further in the near future. One of the primary forms of generative AI is a “large language model” (LLM), where a developer “trains” a model on large amount of text data and then makes this pre-trained model (commercially) available for end users, who supply their own desired text inputs (prompts). In the context of text generation, output ranges from facts and information to software code to entire novels, depending on the prompt and the safeguards built in the models. Certain models also permit the use of audiovisual prompts, as well as generate audiovisual outputs. Both these models and LLMs can be described as “foundation models.”[6]

One distinction made throughout this volume is that between “human-generated” works and “AI-generated” works. This is a useful rhetorical and analytical distinction, but it is not always a practical or precise one. There is a broad spectrum of ways in which AI technology can contribute to some creative expression. These range from AI tools that, for instance, assist with spelling and grammar to generative AI systems capable of producing a sophisticated image entirely based on a single-word prompt.

Nonetheless, there is a point along the spectrum at which we start to think of creative work as being more AI-generated than human-generated. These are cases where humans operate the technology but exercise minimum control over the creative output. The discussions here do not attempt to identify the point at which this happens but acknowledges that such a division, however tenuous it may be, exists, and there may be sensible reasons to treat AI-generated works differently from merely AI-*assisted* works. For the most part, our discussion does not distinguish human-generated works that are AI-assisted from those human-generated works that are not AI-assisted.

6 Trained on massive datasets, foundation models (FMs) are large deep learning neural networks that have changed the way data scientists approach machine learning (ML). Rather than develop AI from scratch, data scientists use a foundation model as a starting point to develop ML models that power new applications more quickly and cost-effectively. The term foundation model was coined by researchers to describe ML models trained on a broad spectrum of generalized and unlabeled data and capable of performing a wide variety of general tasks, such as understanding language, generating text and images, and conversing in natural language (Amazon Web Services, 2024).

Separately, it is useful to identify the parties relevant to the discussion. We identify three primary classes of stakeholders. The first is the public — the consumers of creative works and the benefactors of cultural and scientific innovation. In this group, we include not only those who directly consume creative works but also society at large, since even those who do not directly consume creative works are impacted by the scientific and cultural innovation facilitated through those works.

The second set of stakeholders consists of suppliers of generative AI models, to include data suppliers/aggregators, distributors, and developers. "Developer" can loosely refer to many parties along the long development chain of AI systems. Nonetheless, here, we focus on what is arguably the top of the development chain — developers of so-called "foundation" models. We will exclude questions that arise for "downstream" developers, i.e., developers building on top of these foundation models to produce applications targeted at a specific use case or industry. However, the focus on foundation models is meant to encompass a wide variety of emerging models, including those focused on different types of media such as text, images, video, music, and speech, as well as multi-modal models that combine these things. Most of the discussion here will focus on large foundation models developed and released in the last few years (e.g., GPT-4, LLaMA, Dall-E, Stable Diffusion, etc.).

The third general group of stakeholders are "rightsholders," which refers to both the initial human creators and those to whom rights have been transferred. For example, both the recording artist and the record label may be considered rightsholders in this discussion. Importantly, different types of rightsholders may react differently to AI-induced changes in creative incentives. While the recording artist may be more sensitive to the intrinsic incentives related to maintaining control of their work and its artistic integrity, the record label may be more sensitive to extrinsic pecuniary incentives such as generative AI lowering the cost of producing music. Additionally, copyright policy affects rightsholders in two distinct (but linked) ways, insofar as their work is ingested for model training, and the output of generative AI models can compete with rightsholders' works.

The focus of the following discussions is additionally narrowed to what we refer to as "unbounded" AI models, which differ from "bounded" AI models in policy-relevant ways. The rationale behind this distinction is a presumption regarding the ease of negotiating the rights to potentially copyrighted content before it is used in training datasets.

An AI model may be considered "bounded" if its training data sources and their provenance are easily identifiable (i.e., if there are clear and concise bounds around the training dataset and its sources). This is a necessary condition for efficient negotiations over the use of copyrighted works. These bounded models are, for example, specialized AI tools that learn from a specific and clearly bounded data set. For instance, suppose a developer creates a program that can assist with script plots for a specific TV show by training it exclusively on the show's prior scripts and viewership data. This is straightforward because the creator knows exactly what content the AI is learning from and who owns the rights to that content.

In this case, so long as intellectual property rights are clearly defined, we expect efficient negotiations to occur without additional significant policy intervention. The use of content for training can easily be negotiated before it is used. In the same vein, remedies for potentially infringing output can be agreed upon during those *ex-ante* negotiations (Gans, 2024). Slight variations in this scenario, such as models drawing from a few sources, are also straightforward if it is possible to apportion responsibility to each source, and the transaction costs of negotiation are small.

An AI model is considered "unbounded" when the set of content on which it is trained is not clearly distinguishable or rightsholders relevant to that content are not easily identified. For example, a model based on a large volume of text indiscriminately scraped by web crawlers would be considered an unbounded model. Transaction costs associated with negotiating the use of materials for training in unbounded models may be, in some circumstances, very high.

Because bounded models, by definition, rely on the transaction between relatively few and well-defined actors, they are unlikely to produce the degree of economic challenges posed by unbounded models. For that reason, the ensuing parts of this volume primarily focuses on unbounded models.

1.3. Technological Context

Although a full examination of AI technology is beyond the scope of this project, it is nonetheless helpful to review some of the basic technical details of how generative AI models are trained, as those details are the source of certain economic challenges around copyright policy. Generative AI models, mainly those aimed at text and image generation, are often based on a deep learning technique known as the "transformer architecture" (Turner, 2023). This approach revolves around handling sequential data, like sentences in a text. Unlike previous models that processed data in order, transformers can look at entire sequences simultaneously, allowing for more context and faster learning. They do this using a mechanism called "attention" (Turner, 2023), which lets the model focus on different parts of the input data — for example, correlating words in a sentence based on context rather than just their position.

During training, these models are fed vast amounts of data, which they use to learn patterns and features at multiple levels. In the beginning, essential elements like word associations or simple image patterns are recognized, and as training progresses, the model learns to identify and generate more complex structures. This iterative learning process involves constantly adjusting the model's internal parameters to reduce errors, a well-established method known as "backpropagation" (Rumelhart et al., 1986). The transformer architecture's ability to handle large, complex datasets with this nuanced understanding makes it particularly effective for generative tasks, enabling the creation of text and images that are highly similar to what a human might produce. While many generative AI models do not publicly disclose the data used, some commonly used training datasets include the Common Crawl dataset (which is based on scraped information relating to images and text from a majority of publicly accessible pages on the internet), the Wikipedia corpus, digitized books, and patent records (Baack, 2024; Gao et al., 2020; and Liu et al., 2024). Such a dataset includes copyrighted works.

A critical aspect of these models is that there is not a one-to-one relationship between a given output and a particular piece of training data. Researchers have not yet identified a practical way to determine the impact that a single work in the training data (e.g., a particular novel, song lyric, or fact) has on the model's parameters and thus on what output the model will generate in response to a particular prompt. Training data using these techniques creates instances of models that operate at multiple hierarchical levels. Through this process, the models statistically encode increasingly complex concepts ranging from simple grammar and syntax to social context, meaning, and relationships between entities, all based on statistical similarities. Extracting these complex hierarchical relationships across a wide variety of training texts and encoding them into the model's parameters enables the model to generate outputs in response to apparently unrelated or novel prompts in surprising ways.

References

Amazon Web Services. "What Are Foundation Models." Last accessed November 11, 2024. https://aws.amazon.com/what-is/foundation-models/

Baack, Stefan, and Mozilla Insights. 2024. "Training Data for the Price of a Sandwich: Common Crawl's Impact on Generative AI." Mozilla Foundation. https://foundation.mozilla.org/en/research/library/generative-ai-training-data/

Cuntz, Alexander, Carsten Fink, and Hansueli Stamm. "Artificial Intelligence and Intellectual Property: An Economic Perspective." *World Intellectual Property Organization (WIPO) Economic Research Working Paper Series* 77 (2024). https://www.wipo.int/edocs/pubdocs/en/wipo-pub-econstat-wp-77-en-artificial-intelligence-and-intellectual-property-an-economic-perspective.pdf

Gans, Joshua S. "Copyright Policy Options for Generative Artificial Intelligence." *Journal of Law and Economics* (forthcoming, 2024). https://www.nber.org/papers/w32106

Gao, Leo, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, Shawn Presser, Connor Leahy. 2020. "The Pile: An 800GB Dataset of Diverse Text for Language Modeling." arXiv:2101.00027. https://arxiv.org/pdf/2101.00027

Liu, Yiheng, Hao He, Tianle Han, Xu Zhang, Mengyuan Liu, Jiaming Tian, Yutong Zhang, Jiaqi Wang, Xiaohui Gao, Tianyang Zhong, Yi Pan, Shaochen Xu, Zihao Wu, Zhengliang Liu, Xin Zhang, Shu Zhang, Xintao Hu, Tuo Zhang, Ning Qiang, Tianming Liu, Bao Ge. 2024. "Understanding LLMs: A Comprehensive Overview from Training to Inference." arXiv:2401.02038v2. https://arxiv.org/pdf/2401.02038

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. "Learning Representations by Back-Propagating Errors." *Nature* 323:533–536. https://www.nature.com/articles/323533a0.pdf

Turner, Richard E. "An Introduction to Transformers." arXiv preprint arXiv:2304.10557 (2023). https://arxiv.org/pdf/2304.10557

U.S. Copyright Office. "Artificial Intelligence and Copyright: Notice of Inquiry." *Federal Register* Vol. 88, no. 167. August 30, 2023. https://www.copyright.gov/ai/docs/Federal-Register-Document-Artificial-Intelligence-and-Copyright-NOI.pdf

2. Copyrightability of AI-Generated Works and Demand Displacement

Principal contributors: Imke Reimers and Joel Waldfogel

The purpose of this part is to establish the framework for a rigorous treatment of the questions around copyright protection for AI-generated works in a way that appropriately distinguishes policy objectives from the mechanisms used to achieve them. The main question is how the emergence of generative AI technology affects the optimal provision of copyright protection. Although the technology may change the optimal level of protection for human creators, we primarily focus on optimal copyright protection (or lack thereof) for AI-generated works, noting that the effects of generative AI technology on human creation are a significant determinant of that optimum.

However, it is not yet clear that policies toward AI-generated work can practically diverge from those toward human-generated work. For that reason, we first examine the preliminary and necessary conditions for having divergent policies based on the creative origins of a work. The subsequent sections then examine, in turn, the net social value proposition of AI-generated works and the changing economics of human creation.

2.1. Feasibility of Divergent Policies Toward AI-Generated Works

As discussed above, whether having divergent policies toward works based on the use of generative AI is sensible depends first on whether it is feasible. Copyright policy more generally tends to be agnostic toward many characteristics of creative works.

However, there are occasions when divergent policies do emerge. For example, works made for hire can have a different copyright term than other works. Nonetheless, a key factor in the efficacy of such policies is the degree to which the characteristic on which the policy turns is observable. Indeed, distinguishability is a necessary (but not sufficient) condition for any sort of divergent policies to be effective. Put simply, AI-generated works cannot be *consistently* excluded from copyright protection unless some institution (e.g., the court system) is able to reliably discern the role of generative AI in the production of a work.

For example, the current system is based on registration applicants self-reporting AI usage in the production of works. This allows for claims to be registered for the elements that exhibit sufficient human authorship. Nonetheless, the duty to identify AI-generated aspects falls to applicants, thus, the AI-generated works are only distinguishable from human-generated works if applicants reveal accurate information. Basic economics tells us the conditions under which applicants will have adequate incentives to do so. One of those conditions is that the expected (negative) value of the consequences from failure to disclose are sufficiently high. That, in turn, depends on the likelihood of detection and the magnitude of the consequences conditional on being detected. Note, the likelihood of detection, and thereby the incentives to disclose accurate information, may effectively be zero if an AI-generated work simply cannot be distinguished from a human-generated work in any way, since the owner will face no consequences from failing to provide accurate information and will likely have strong commercial motives to not do so. It is likely too early to gauge practical distinguishability, and it is a matter that is likely to change quickly as both generative AI and detection technologies evolve.

While distinguishing AI content from human content based solely on the work in question is an increasingly difficult task, it is not the only option. Another possibility is that the judicial system establishes standards based on secondary or indicative evidence, for example, contemporaneous documentation of the creative process. This would come with its own set of challenges, costs, and inefficiencies that may outweigh any potential benefits, but it is not entirely without precedent. Moreover, if technological developments in generative AI renders direct detection of creative origins impossible, some sort of system for indirect determination of creative origins will likely be necessary if we wish to continue excluding AI-generated works from copyright protection.

However, even if one could determine the creative origins of a work and identify suitable differential extents of protection for different products, another practical issue arises. There are many different levels of AI involvement in creation, and it is difficult to pick a cutoff level beyond which a work becomes "AI-generated." It is unlikely that such a cutoff could be constructed as a bright line; rather, it would likely need to be based on a set of heuristic rules, potentially creating legal ambiguity, which comes with a social cost. Indeed, the U.S. Copyright Office (2023) recognizes the "case-by-case" nature of this issue in laying out its process for reviewing the registrability of works containing AI-generated material:

> *[The Office] begins by asking "whether the 'work' is basically one of human authorship, with the computer [or other device] merely being an assisting instrument, or whether the traditional elements of authorship in the work (literary, artistic, or musical expression or elements of selection, arrangement, etc.) were actually conceived and executed not by man but by a machine." In the case of works containing AI-generated material, the Office will consider whether the AI contributions are the result of "mechanical reproduction" or instead of an author's "own original mental conception, to which [the author] gave visible form." The answer will depend on the circumstances, particularly how the AI tool operates and how it was used to create the final work. This is necessarily a case-by-case inquiry.*

The following subsections proceed with the assumption that it will be feasible to distinguish AI-generated works from others. However, this remains an open foundational question.

2.2. Net Social Value of AI-Generated Works

Whether it is economically efficient to extend copyright protection to AI-generated works depends on the net value proposition of AI-generated works. If AI-generated works cause a net reduction in value, it is not efficient to further incentivize their production through copyright protection. In contrast, a net positive value may make it efficient to extend copyright protection to AI-generated works (depending on other conditions holding). While it is likely that some AI-generated works will have social value, that value needs to be weighed against the value of human-generated works displaced by the technology. This section examines both the value of AI-generated works and their displacing effects.

With respect to the former, it is possible for generative AI to create services or products whose value to consumers exceeds the value of the content they incorporate. These may be things that are creative or innovative in and of themselves or things that aid in the creativity or innovation of others. However, there are also cases where AI-generated works add no incremental social value. For example, when generative AI platforms produce near-verbatim copies of existing work, those works are no more valuable than the original — except possibly where the copies enhance public access to existing works.

Another important consideration is the extent to which AI-generated works serve as viable substitutes for human-generated works, either directly or indirectly. The substitution effect is easy to conceptualize in the case of AI producing near-verbatim copies: Consumers may choose to consume an AI-generated work over a human-generated work if the former is more accessible or cheaper than the latter and sufficiently satisfies the same need. This displaces the demand for human-generated works and the value that rightsholders can capture.

At one extreme, we can analogize the effects of AI-generated works to those of piracy, to the extent that generative AI models can produce near-verbatim unauthorized copies of valuable — and costly-to-create — content. In this context, one would expect AI-enabled access to works to substantially displace human creators' revenue. Each use in which a consumer would have been willing to pay the price for the pre-existing product (but now obtains it via generative AI) would transfer or eliminate revenue.

While the substitution effect is easy to conceptualize in the case of AI producing near-verbatim copies, it exists (perhaps to a lesser degree) even when AI output bears no substantive resemblance to a particular human-generated work. For example, recording artists often buy or commission musical works from songwriters. While songs are certainly not a fungible commodity, if a model can produce songs of sufficiently high quality for substantially cheaper than a human songwriter, the AI generated materials may function as substitutes. In this scenario, it would be rational for some recording artists to opt for the AI option, thus diminishing the demand for and revenue of human songwriters. In this way, AI may reduce the production of human-generated musical works.

While generative AI is likely to have a meaningful substitution effect on human-generated works, not all of its uses would supplant revenue for human creators. Some uses will reduce deadweight loss, replacing it with consumer surplus by allowing for additional consumption that otherwise would not occur. For example, a user who produces a verbatim copy of some literary work may have never been willing to purchase the original at its market price but nonetheless gains some utility from consuming it. In this sense, generative AI creates value through a sort of price discrimination mechanism, facilitating consumption in instances in which users are unwilling to pay the content creators' prices.

Similarly, the aspiring recording artist who may not be able to license a high-quality musical work from a human songwriter may, with the help of generative AI, be able to produce a record that would not have otherwise been feasible. In this case, in addition to any direct net value of the AI-generated song, the work will also lead to a knock-on effect, insofar as it facilitates the production of a (human-generated) record that may not have otherwise been produced.

Empirical research is needed in order to better understand the relative magnitude of the value created and the value displaced by generative AI. For example, understanding the efficiency gains around the creative process facilitated by AI will help determine its value relative to human-generated works. Similarly, developing an understanding of the extent to which AI allows users to circumvent copyright restrictions (e.g., by producing near verbatim copies of works), and how much of that consumption is incremental to what would have otherwise occurred, will help identify the potential displacing effects of generative AI.

There is also a second negative impact AI-generated works may have on human content creators, beyond the diversion of revenue: it may decrease the value of the underlying human work. A generative AI model may provide consumers with a tarnished version of copyright-protected content (for example, by providing content the model presents as being associated with the *New York Times* but now with misinformation). To the extent that this diminishes the perceived value of the original (e.g., because the quality of the tarnished version is being conflated with the quality of the original) consumers may be less willing to pay for the original. That is, in addition to generative AI models being able to take a piece of the pie through the substitution effects, they can also reduce the size of the entire pie by degrading perceived quality, further displacing demand for human-generated creative works. Again, further research is needed to assess the extent to which value of human authored works may be diminished and whether this issue will ebb as the quality of AI works improves.

There is also a third, and perhaps more complex, way in which the displacement of human creators may impact long-run scientific and cultural innovation (Yang and Zhang, 2024). Increases in the number of new creative products generate welfare benefits in part because of the experimentation process. When humans create, there is a large degree of substantive and stylistic variance in the output. Given the variety of the output, unanticipated successes arise often and open new paths of innovation. For example, there may be many musicians producing music within a particular genre, each producing different variations on the genre's core characteristics. Given enough variations, the market will seize on the experimental outcomes that are most pleasing or innovative, which can then cause the development of an entirely new genre. The success of this process depends critically on variation and experimentation. It is unclear whether AI-generated output can ever engage in the same sort of experimentation and innovation as humans.

Perhaps AI can only help authors create formulaic work, causing lower substantive and stylistic variance than human work. The lower cost, as reduced by AI technology, may then cause these formulaic works to crowd out more risky and costly experimental creations that sometimes lead to valuable innovation. However, it is also difficult to rule out the alternative possibility that AI could function as a complement to high-variance creation. Perhaps the high-variance aspect of creative products comes at conception, and the generative AI allows faster and cheaper realization of concepts. Then AI-assisted creations might hasten the creation and discovery of valuable content.

Additionally, the displacement of human creators may, at some point, also slow the progress of generative-AI technology. Human-generated works, on which foundation models are trained, fuel generative AI. The output of these models is limited by the volume and quality of the input (human-generated works). The progression of the technology is currently contingent on ever-growing bodies of training materials (Sevilla et al., 2022; Villalobos et al., 2022). Diminishing incentives for human creators may thus degrade the long-run capabilities of the technology to the extent that it limits the fuel needed to advance AI technology. Insofar as developers are creating and capturing value, and their future ability to do so depends on access to new high-quality human-generated works, some market between human creators and AI developers may arise. Whether that market would equilibrate to the socially efficient outcome is an open question that warrants a close examination, both from technological and economic perspectives.

Generally, offering copyright protection to AI-generated works will further incentivize their production, exacerbating the effects, both positive and negative, on human creative output. This is not necessarily a reduction in creative output but does represent a shift in who (or what) is producing that output. Whether this represents a diminishment of creative value (or welfare more generally) requires a judgment of the relative value of human-generated works compared to AI-generated works.

However, even if the net value proposition of AI-generated works is positive and substantial (i.e., their value is greater than the value they displace), offering copyright protection to those works may not be socially optimal. As previously discussed, the economic function of copyright is to correct the fixed cost recovery problem with creative works, which would otherwise lead to underproduction. Recall, however, that copyright inherently limits public access to existing works and thus produces a social cost. Whether the underlying market failure (the fixed cost recovery problem) remains substantial enough to warrant the social costs associated with copyright protection is an important but, to date, largely unstudied question.

Production costs for AI-generated works may be meaningfully lower than those for human-generated works. Hence, we must consider both the incremental, quality-adjusted gain in creative output that copyright may produce *and* the degree to which public access to works is limited by copyright. Copyright protection only serves its economic objective if the social value of the former outweighs that of the latter. If the fixed production costs of AI-generated works are sufficiently low, the additional incentives of copyright are not necessary for reaching optimal production levels, thus, offering copyright protection would be suboptimal. Empirical research on the degree to which the fixed cost recovery problem exists for AI-generated works will help us better understand the need for copyright protection of AI-generated works.

2.3. The Changing Economics of Human Production

While it is possible that offering copyright protection to AI-generated works may give those works an inefficient competitive advantage over human-generated works, the effects on the incentive structure of human creators may also be tempered by the changing economics of human creativity. This is because AI may not only deliver products and services that compete with existing content (causing a shift in the demand for human-generated works) as discussed in the previous section; it may also facilitate creation in ways that make creators more productive (a potentially countervailing shift in the supply curve for human-generated works). For example, AI may make it easier for writers of news articles, books, and screenplays to do their work, allowing them to produce more or higher-quality work with less time and fewer resources. Viewed from this perspective, AI resembles other technological changes, such as digitization, from which insights may be gained.

Digitization affected both demand for and supply of creative products. With respect to demand, digitization facilitated copying and distribution, including unauthorized distribution, or piracy (for example, unauthorized file sharing). Piracy as a substitute for authorized sales substantially diminished the market power of creators. For instance, recorded music revenue was cut in half. Nonetheless, despite the demand shock, production of new music did not subside. This is likely because digitization also shifted the supply curve with substantial reductions in the costs of production, distribution, and promotion (noting that some of those cost savings may have been offset by new costs, such as enforcement of digital rights). Similar effects can be observed with books, movies, television, and photography, among other categories (Waldfogel, 2017 and 2019). Moreover, there is evidence that the quality of the new products (in terms of usefulness to consumers) was high relative to recent historical standards (Waldfogel, 2012). Ultimately, the effects of digitization serve as a reminder that technological changes can have simultaneous and sometimes countervailing effects, for example, on revenue *and* production costs. If this is the case with generative AI, considering one, or even some limited subset, of its effects in isolation will be misleading.

Putting aside for now the previously discussed substitutional effects generative AI may have on human-generated works, we also expect the technology will reduce production costs for human creators. The technology increases human productivity thus reducing labor cost per quality-adjusted unit.[7] This will have dynamic effects on the market for human-produced creative works and on the labor market for human creators.

When the costs of production factors (in this case, labor) decrease, the supply curve shifts to the right, decreasing price and increasing the quantity produced. Since the quality of creative works is also a function of labor costs, and because quality is a significant competitive factor in the market for creative works, we expect that the supply curve shift will result in some combination of lower prices for accessing creative works, increased quality of works produced, and increased production volume. We can think of this as a decrease in quality-adjusted prices and an increase in quality-adjusted volume.

These decreased prices, increased quality, and increased creative output are unambiguously beneficial to consumers, at least from a static perspective (again, putting aside the previously discussed substitution effects). However, the degree to which this improves welfare depends heavily on the predictability of product quality at the time of entry (Aguiar and Waldfogel, 2018). In a context where the quality or success of a creative work is highly predictable, the welfare gains from reduced production costs will be modest. This is because works are only produced if their predicted value

7 The increased efficiency could, depending on the characteristics of the supply and demand functions, manifest itself as an increase in average quality rather than a decrease in per-unit cost, since the increased efficiency would make attaining higher quality levels less costly. Nonetheless, on a constant-quality basis, per unit labor costs would likely decrease.

exceeds the cost of producing them. In a scenario where value can be reliably predicted, every work that could produce sufficiently high value is already being produced. A reduction in costs would facilitate the production of works with a predicted value that would have previously been too low to warrant production. Hence, in a world of perfect predictability, cost reductions allow for additional entry of only marginally worse products. Such additional products may not be very consequential for consumers or for revenue.

Nonetheless, predictability is typically low — an understanding codified, for example, in the movie industry, by the expression "nobody knows anything" (Goldman, 1989). In that sort of environment, a cost reduction that facilitates the entry of additional products will deliver products throughout the realized quality distribution, including some high-value "winners." Indeed, in the extreme case in which value is entirely random and unpredictable, the additional products made possible by a cost-reducing technology will be as good on average as the pre-existing products. Thus, the welfare gains that follow cost reductions may be substantial in certain contexts. The predictability within various markets for creative content will, therefore, be a significant factor in determining how the effects of AI may differ from one context to another.

While the static benefits of an AI-induced reduction in production costs to consumers are clear (at least in direction, if not in magnitude), the effect on human creators is less predictable without sufficient information about the demand curve and cost function for creative works. The increased productivity of creators means that fewer creators are required in order to produce a given volume of creative works. All else constant, this decreases the labor market demand for human creators. However, increased productivity is not the only relevant factor. As discussed above, the volume of creative works demanded will increase in the face of lower production costs, which would increase demand in the labor market for creators, all else constant. Said differently, the labor required to produce a unit of output decreases, but the volume of output increases with an ambiguous net effect on labor.

Which of these two effects will dominate is an empirical question. It depends on the slope of the demand curve for creative works and the sensitivity of the supply of creative works to changes in labor productivity.[8] If marginal production costs (and hence, price) for creative works are sufficiently sensitive to changes in labor productivity, and demand for creative works is sufficiently elastic, we would expect to see an increase in the demand for human creators and an increase in their earnings.[9] That is, in terms of labor demand, the increase in the quantity of works demanded may outweigh the decreased labor needed to make each work. Conversely, if the price of creative works is relatively insensitive to changes in labor input costs, or the demand for creative works is inelastic, the increase in productivity will lead to a reduction in the demand for human creators and a decrease in their earnings.

In the latter scenario, fewer creators may be able to support themselves through their creative activities. This is clearly not good news for human creators, as many will be pushed to alternative

8 The cost function tells us by how much prices will fall as labor productivity increases. The slope of the demand curve determines the magnitude by which the volume demanded will change in response to a change in prices.

9 An increase in earnings is contingent on an upward sloping labor supply curve.

forms of employment. However, putting aside substantial adjustment costs for those affected, this is likely to increase consumer welfare. Such is a common pattern with technological innovation. For example, consider the advent of the camera. The new device put many portrait artists out of business, pushing them to other occupations. At the same time, cameras facilitated new forms of creative expression and paved the way for further innovations like cinematography. The camera, despite its displacing effects, clearly led to advancements in the production of creative works and generally improved social welfare.

References

Aguiar, Luis, and Joel Waldfogel. “Quality Predictability and the Welfare Benefit of New Products: Evidence from the Digitization of Recorded Music.” *Journal of Political Economy* 126, no. 2 (2018): https://doi.org/10.1086/696229

Goldman, William. *Adventures in the Screen Trade: A Personal View of Hollywood and Screenwriting*. Warner Books, 1989.

Sevilla, J., L. Heim, A. Ho, T. Besiroglu, M. Hobbhahn, and P. Villalobos. “Compute Trends Across Three Eras of Machine Learning.” In *Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN)*, Padua, Italy, 2022, 1–8. https://doi.org/10.1109/IJCNN55064.2022.9891914

U.S. Copyright Office. “Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence.” *Federal Register* Vol. 88, no. 16190. March 16, 2023. https://www.copyright.gov/ai/ai_policy_guidance.pdf

Villalobos, P., J. Sevilla, L. Heim, T. Besiroglu, M. Hobbhahn, and A. Ho. “Will We Run Out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning.” *arXiv preprint* arXiv:2211.04325. https://arxiv.org/pdf/2211.04325

Waldfogel, Joel. “Copyright Protection, Technological Change, and the Quality of New Products: Evidence from Recorded Music Since Napster.” *Journal of Law & Economics* 55, no. 4 (November 2012): https://doi.org/10.1086/665824

Waldfogel, Joel. *Digital Renaissance: What Data and Economics Tell Us About the Future of Popular Culture*. Princeton University Press, 2018.

Waldfogel, Joel. “How Digitization Has Created a Golden Age of Music, Movies, Books, and Television.” *Journal of Economic Perspectives* 31, no. 3 (2017): 195–214. https://www.aeaweb.org/articles?id=10.1257/jep.31.3.195

Yang, S. Alex, and Angela Huyue Zhang. “Generative AI and Copyright: A Dynamic Perspective.” *arXiv preprint* arXiv:2402.17801 (2024). https://arxiv.org/pdf/2402.17801

3. Copyright Infringement by AI Output

Principal contributor: Joshua Gans

Generative AI can, intentionally or otherwise, closely replicate existing works. When and whether this constitutes infringement is a legal question. Nonetheless, from an economic perspective there is a guiding consideration in defining the optimal scope of what output is infringing: copyright protection from infringement should balance the incentives to produce and the ability to consume creative works. This arises from the fact that the mechanism used for incentivizing the production of new works (exclusive rights pertaining to the usage of a work) also limits consumers' access to existing works. Similarly, it also limits access by follow-on creators.

On the one hand, the breadth of protection from potentially competing works, such as similar novels or music tracks, directly affects the market power conferred through copyrights. In that sense, the broadest possible scope of protection could offer substantial economic rewards to those who successfully create a desirable, non-infringing work. On the other hand, all new creative work is built upon existing works, whether those existing works inspire the new work, provide an archetype, or aid the creator's learning process. Thus, it is unlikely that any creator could produce a new work that bears no resemblance to the existing works that the creator previously consumed. Therefore, the broadest possible scope of protection could also effectively hinder new creative output for fear of liability.

Likewise, the scope of protection offered to a particular work may inversely affect the public's access to that work. For example, suppose paraphrasing is outside the scope of infringement protection so that consumers can access a non-infringing summary of a textbook. In that case, the textbook publisher may need to lower its prices to induce consumers to buy the book rather than the summary. Those lower prices translate into greater public access (conditional on the textbook being produced in the first place), which serves the policy objective of copyright — to promote the creation and dissemination of creative works. Still, the lower profit margin may also make the publisher less likely to produce more textbooks in the future, which would impede future access.

Ultimately, the way to optimally balance public access to and consumption of existing works with the incentives to create new, socially valuable works is by conferring a limited degree and scope of market power to creators through copyright protection so that they may appropriate some of the social value their works generate. This ability to appropriate value is an incentive to create works, but the underlying market power used to do that is inherently access-limiting.

Thus, the balance, or constrained optimization problem, at the heart of copyright questions can be expressed as an exercise in achieving the optimal degree and scope of market power to permit creators. Analyzing this problem requires first identifying the optimal level of market power that we wish to confer to rightsholders in the context of competing AI-generated works.

For the present purpose, we assume that the optimal degree and scope of market power in this context is the same as that of current policy toward potentially infringing human-generated works. While it is unclear whether this analog is at an optimal level or scope (a subject outside the bounds of the current debate), it provides a valuable reference point.

With an established target, it is then necessary to consider the factors that contribute to the balance in the human analog, how those factors may differ in the case of AI output, and how to bring them back into alignment, if necessary. Notably, numerous factors may affect market power. Those factors and their impact likely differ across types of works and the business models built around commercializing those works. Our discussion here does not attempt to articulate every factor or expound upon the nuanced ways they may be salient in different industries. Rather, it discusses only some of the infringement factors that influence market power.

3.1. Thresholds for Infringement

One of the most salient factors determining the market power conferred through copyright is the threshold used to determine whether one work infringes upon another. A higher threshold for infringement, such as strictly requiring precise verbatim copying, decreases the market power of rightsholders, while a lower threshold, such as requiring only passing similarity, increases their market power.

One of the legal conditions for infringement is that the alleged infringer must have copied (and not have independently created) some element of the copyrighted work. How parties are required to demonstrate copying, or lack thereof, will also affect market power. Currently, at a very high level of generality, rightsholders are required to show that an alleged infringer had the opportunity to access and copy a work; then the burden usually falls to alleged infringers to demonstrate that they did not engage in copying, thereby proving independent creation.

Because the copying requirement can be understood as a consumption requirement (or, more precisely, an opportunity for consumption, or access, requirement), the way it is applied may differ between in the case of infringing AI-generated works. In practice, access may be harder to dispute in the context of AI output, depending on the type of model that produced it.

In some instances — such as those in which the model developer uses licensed data sets or otherwise tracks what works are used in training — determining whether a model consumed the copyrighted work in its training may be relatively easy. For others, such as those trained on massive data sets scraped from the internet, it may be much more difficult to make such a determination. Indeed, because such models may have ingested much of what has ever been put online, affirmatively demonstrating that the data used to train it does or does not include a specific copyrighted work may be extremely challenging. Because access to evidence about the contents of a training set may impact the likely success of a copyright infringement claim, it also affects the market power of rightsholders compared to their market power in the human infringer analog.

How the access requirement will be handled in the context of potentially infringing AI output is a nuanced legal and policy question. Nonetheless, it is crucial to recognize that the human analog case may affect the market power of rightsholders differently, potentially warranting some recalibration of that market power through other mechanisms. Empirical evidence can help us anticipate the degree to which this will be a consequential issue. In particular, a measure of the importance of “independent creation” defenses in infringement claims against human creators would be highly informative. If “independent creation” is a seldomly successful defense, then it being marginally less likely to succeed may only make a trivial difference in the case of infringing AI works. Conversely, if “independent creation” is a sufficiently successful defense, then the difference between the human and AI infringement cases may be substantial.

3.2. The Value of Infringement Claims

The interpretation of infringement thresholds may determine whether a rightsholder can exercise market power against a particular work. Still, it does not tell us the potential value that can be extracted through that market power. The value of asserting infringement claims is determined by the magnitude and form of remedy available and the costs of pursuing those remedies. Remedies are typically determined by statute, judicial precedent, and the circumstances of a particular infringement. Policy, judicial institutions, and technology largely determine the cost to rightsholders of pursuing those remedies.

One policy that has a major effect on the cost of pursuing infringement claims relates to the type of parties that can be held liable for infringing works. In particular, parties that demonstrate a higher concentration of infringement often prove less costly to pursue, on a per-infringement basis. That is, given a fixed number of infringements, pursuing a small number of high-frequency infringers will often be less costly than pursuing a large number of low-frequency infringers. Identifying and then bringing claims against many small-scale infringers is expensive. Moreover, those small-scale infringers may have fewer resources to pay damages than larger-scale and higher-frequency infringers. This means that a set of infringements may be worth pursuing if they can be bundled together and brought against a consolidated infringer, but that same set of infringements may not be worth pursuing if each infringement must be individually adjudicated.

An example of this arose in the 1990s as internet platforms for sharing (often copyrighted) content began to develop. It was then widely argued that holding internet service providers (ISPs) and hosting sites accountable for infringing materials accessed or distributed by their users would stifle the growth of such platforms and technological development (Bello and Aufderheide, 2021). For that reason, a “safe harbor” provision was built into the Digital Millennium Copyright Act (DMCA) that provided a limitation on liability for eligible ISPs and hosting sites from damages and injunctive relief so long as certain conditions were met (Nimmer, 2000).

The safe harbor provision of the DMCA almost certainly profoundly impacted how ISPs, hosting sites, and their business models developed over the subsequent twenty years. Without knowing what would have transpired in the absence of the DMCA, it is difficult to speculate on the ultimate social value created (or lost) by this law. However, one clear fact is that the DMCA’s enactment substantially changed the practical enforceability of copyrights compared to the alternative. This is because, instead of being able to pursue a low number of high-frequency infringers (ISPs), infringement liability became more decentralized, scattered among a high number of low-frequency infringers (individual users uploading or downloading infringing content).[10] The increased transaction costs of infringement claims and the reduced average payout to rightsholders meant that relatively fewer incidents would be litigated, thus effectively diminishing the level of copyright enforcement.[11]

In the case of infringing AI-generated output, liability could logically fall to the individual AI user who created the infringing work with prompts to an AI model, to entities responsible for giving users the ability to create infringing work (e.g., developers of the AI models), or to both. Limiting liability to only individual users would likely increase the transaction costs faced by rightsholders wishing to enforce those rights, thus reducing the effective level of enforcement and decreasing rightsholders' market power. In contrast, assigning liability to developers would likely decrease the transaction cost per infringement, thus increasing the effective enforcement level and rightsholders' market power. The converse is true for AI development costs — developer liability would lead to greater developer costs, and liability limited to users would lead to lower developer costs.[12]

Which liability regime would be optimal is a matter that will rely on empirical characteristics of the market and technology. The decisions around who may be held liable for infringement will almost certainly have implications for enforceability. The magnitude of those implications depends on how diffuse the output-generating population is and the degree to which it has the resources to cover damage awards. For example, if infringers are typically small-scale web content creators, that would constrain enforcement more than if infringers were sophisticated marketing firms creating large volumes of content for advertisers. A highly diffuse or under-resourced class of potential infringers will make any policy decision around developer liability all the more consequential concerning practical enforceability. A measure of how the frequency of infringement claims and the magnitude of damages change in response to differences in the diffusion and resources of liable parties would be informative in this context.[13] Such a measure and a qualitative understanding of AI developers and users could allow us to better anticipate the impact of decisions around liability assignment.

3.3. The Cost Function of Infringers

Another factor that can affect the level of market power is the cost of producing infringing works. Producers of infringing works (competitors of original content creators) face a production cost function, just like any other producers. Two components of that cost function are the likelihood of the

10 There are some circumstances where this was not the case; for instance, pirate websites that fell outside the safe harbor of the DMCA could still be held liable.

11 The DMCA also offered at least one countervailing factor by establishing a notice-and-takedown system by which rightsholders could request that ISPs remove infringing content. This serves as a relatively inexpensive way to block specific instances of infringement, but usually does not provide direct monetary transfers from infringers to rightsholders. Thus, while the notice-and-takedown system likely reduced the cost of policing infringements, its overt deterring effects are likely limited.

12 This trade-off is considered by Gans (2024) building on early work in Gans (2015). Depending on whether the volume of content that is, for example, used in training data for generative AI, the likelihood of transaction costs being important can change. Thus, there may be a difference between large language models that are trained on a high volume of content and other more tailored models, such as retrieval augmented generation (or RAG) models trained on specific content.

13 Gans (2024) proposes such a measure that compares the importance of content to the efficacy of the AI model versus the potential expected harm to content holder's commercial interests if the AI model is available.

rightsholder successfully pursuing an infringement claim against the infringer and the consequences when that happens. However, an even more basic component of that cost function is the cost of manufacturing infringing works, whether that is photocopying a book, painting a reproduction of some artwork, entering prompts into a generative AI platform, or something else.

The effects of simply changing the production and distribution costs of infringing works were broadly observed in the digitization of works that began in the 1990s. Digitization eliminated direct costs for copying and distributing some works, increasing the supply of infringing works. This surge of infringing works (often perfect substitutes for original works) significantly reduced rightsholders' market power.

As a general economic principle, infringing works will be produced and distributed when the value the infringer can capture exceeds the total costs of producing infringing works (including both direct production costs and the expected penalty for infringement). If the production and distribution costs are reduced, more infringing works will meet that threshold and be distributed. The more infringing works against which a rightsholder must compete, the lower their market power (and incentive to create) will be.

Whether generative AI technology reduces the costs of producing infringing works and the degree to which it may do so are empirical questions. In particular, it is necessary to measure the cost savings AI generates in terms of copying and how that affects the supply of infringing works. However, to the extent that generative AI does lower such costs and other factors remain constant, more infringing works will likely be produced and distributed. In that way, the market power of rightsholders may be lower in the context of infringing AI works than in the context of infringing human works.

3.4. Indirect Value Appropriation

Another critical question about infringing AI output revolves around the degree to which the value of infringing output can be indirectly appropriated by rightsholders prior to infringement. So far, the discussion in this part has primarily focused on applying market power at the point where infringing output is produced or distributed — the point at which the value of infringing works can be *directly* appropriated. However, there is another point where negotiations may occur, and market power could be applied. This is at the point of original access. To produce some infringing output, the infringer must first, in some form, access the original work. To the extent the owner of that original work can restrict access before any infringing output is produced, they can, under certain circumstances, extract some of the value of future infringement through negotiations around access. This is known in the economic literature as "indirect appropriation."

A seminal example of this in the economic literature is that of photocopying scholarly journals after photocopying technology became widely available (Liebowitz, 1985). The example demonstrated that the ability to create unauthorized copies of a journal (an infringement) increased the value of authorized copies to the original purchaser. Once photocopying became a ubiquitous technology, the value of a journal to the purchaser was no longer constrained to a single reader; it could be photocopied, providing value to many more readers to whom the original purchaser (usually an academic institution) granted access. This, on its own, was not ideal for publishers because many of those readers would have otherwise purchased an authorized copy (or their institutions would have purchased additional authorized copies). However, institutional purchasers of the journals implicitly recognized the secondary value from their purchase of authorized copies (in the form of unauthorized photocopies). This increased their willingness to pay for the authorized copies (anticipating the additional value from subsequent photocopying) and resulted in publishers being able to charge more for the original authorized copy. While publishers likely would have preferred to directly monetize infringements, they were able to dampen the negative effects by appropriating some of the value of future infringing copies at the point of original sale.

In the context of generative AI, the analogous theory is that the value of future infringing output could, in principle, be captured through negotiations for access to training data. The ability to create infringing AI output presumably provides some value to users of generative AI platforms. The AI platform can capture that value in various ways, depending on its particular business model. That captured value increases the developers' willingness to pay for the inputs necessary for creating and maintaining the platform, including the creative works used to train generative AI models. Depending on the balance of bargaining power, that can allow owners of works likely to be the subject of infringing output to capture some portion of the value generated by future infringement through negotiation for access to training data.

The necessary factors that led journal publishers to adopt and be successful with a model of indirect appropriations were: (1) publishers could no longer easily prevent unauthorized copying (making the *ex-post* assertion of market power difficult), (2) the primary purchasers of authorized copies were able to capture the value of secondary users of unauthorized copies and incorporate that into their demand function, and (3) it was relatively easy for publishers to restrict access to original copies of journals. Because the factors that make indirect appropriation a favorable, or even viable, model do not universally exist, it is helpful to consider how those factors may or may not hold in the context of generative AI.

Concerning the first condition, it will likely be challenging for rightsholders to prevent the production and distribution of infringing AI output. However, how challenging this will be depends critically on other policies. For example, the assignment of liability for infringing output to AI platforms would make it relatively easier to prevent infringing output compared to a scenario where liability is limited to technology users. The harder it is to assert rights at the point of infringement, the more appealing models of indirect appropriation become.

The second necessary condition is that infringing AI output creates value for the infringer, and developers can capture a relevant portion of that value. There is likely value to users of AI platforms in the production of infringing works, but the extent to which that is true will be context specific. For some uses, the production of infringing output may be incidental, providing the user with no more value than would be derived from non-infringing output. In other circumstances, the ability to create a work substantially similar to some other copyrighted work will be of critical value to the user. Under normal circumstances, we would expect developers to be able to capture some relevant portion of this value.

The third necessary condition (rightsholders being able to restrict access to their works for training purposes) is the most uncertain and pivotal in generative AI. This condition requires that rightsholders can determine whether the model ingested their work and that they have sufficient recourse when unauthorized ingestion does occur. The first part of this may depend on the nature of the model. As previously discussed, tracking input may be possible in the case of bounded models, but it may be impossible or prohibitively expensive in the case of unbounded models. What recourse is available is ultimately a matter of policy, but may range from no available recourse (e.g., if training is universally treated as a fair use) to potentially crippling levels of recourse (e.g., if full statutory damages are applied to every instance of every work ingested by the model). Where policy falls along that spectrum will be a crucial determinant in whether models of indirect appropriation are viable.[14]

Another consideration in the context of access is transaction costs and intermediation. Indirect appropriation occurs through negotiations at the point of access. That means developers would need to negotiate with a large, diffuse set of rightsholders associated with the millions or billions of works ingested by generative AI models. Such negotiations conducted individually would create transaction costs that dwarf the value at stake, making some sort of collective intermediation necessary.[15]

Rigid forms of intermediation, such as statutory licensing, will likely diminish the viability of indirect appropriation, given that statutory licensing constrains the bargaining power of rightsholders. Similarly, intermediation that does not allow for some value distinction between the works being ingested will likely lessen the viability of indirect appropriation. This is because the vast majority of works that a foundation model ingests will never be associated with some infringing AI output. Those works, as a whole, may make valuable contributions to the model by providing context or calibrating associations within the model but still be works that no one wishes to replicate in terms of output. Those works will not generate the secondary value that could be indirectly appropriated. An intermediary that does not or cannot distinguish between works that are likely to be the subject of infringing output from those that are not will license all works at a rate reflecting the average. That average will be influenced mainly by the vast majority of works that do not lead to value in the form of infringing output.

Ultimately, when considering how much market power we wish to confer upon rightsholders, it is essential to do so in the contexts of both *ex-post* assertions of market power (claims against infringing output) and *ex-ante* assertions (indirect appropriation before infringement). The latter, when viable, can act as a backstop for the former. Conversely, when the *ex-ante* assertion is not feasible (for example, if model training is treated as a fair use), decisions around *ex-post* market power will carry greater consequences. Thus, it is useful to develop an understanding of the contexts in which indirect appropriation may be a viable mechanism for rightsholders to capture value from infringing works.

14 Policies around access to training data are discussed in more detail in subsequent parts.

15 This would clearly be a less pertinent issue for the “bounded” models discussed in Part 1. Later parts discuss the various forms this intermediation could take.

References

Bello, Bryan, and Patricia Aufderheide. “The Public Interest and the Information Superhighway: The Digital Future Coalition (1996–2002) and the Afterlife of the Digital Millennium Copyright Act.” *Information & Culture* 56, no. 1 (2021). Available at https://muse.jhu.edu/pub/15/article/784295.pdf

Gans, Joshua S. “Remix Rights and Negotiations over the Use of Copy-Protected Works.” *International Journal of Industrial Organization* 41 (2015): 76-83. Available at https://doi.org/10.1016/j.ijindorg.2015.06.004

Gans, Joshua S. “Copyright Policy Options for Generative Artificial Intelligence.” *Journal of Law and Economics* (forthcoming, 2024). Available at http://dx.doi.org/10.2139/ssrn.4707911

Liebowitz, S. J. “Copying and Indirect Appropriability: Photocopying of Journals.” *Journal of Political Economy* 93, no. 5 (1985): 945–957. https://www.journals.uchicago.edu/doi/abs/10.1086/261343

Nimmer, David. “A Riff on Fair Use in the Digital Millennium Copyright Act.” *University of Pennsylvania Law Review* 148, no. 3 (January 2000). https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=3355&context=penn_law_review

4. Commercial Exploitation of Name, Image, and Likeness

Principal contributor: Shane Greenstein

AI may affect creative incentives by also infringing on other intellectual property rights that are copyright-adjacent but are typically not thought of as copyrights. These rights, often loosely characterized as "rights of publicity," pertain to one's name, image, or likeness (NIL), which can cover a range of intangibles from product endorsements to faked imagery to one's voice or even, in some sense, style.

Rights of publicity are often thought of in the context of fraud protection, as bad actors can and do use AI-generated imitations of people or their work to harm either the subject of imitation or some third party. For example, so-called "deepfakes" can be used to portray someone in an unfavorable light causing them social harm. This sort of nefarious use of AI technology unambiguously diminishes social welfare, and there are many reasons to protect against it.

Because the effects of nefarious use are unambiguous and the remedies are economically (if not technologically) straightforward, this section focuses on non-nefarious use of AI technology in the context of rights of publicity, which can have implications for the incentives to create and innovate.[16] For example, suppose one can create a high-quality musical work in the style of a famous popstar with a few prompts in a generative AI platform. This was the case with the recent duet between musicians Drake and The Weeknd, which became an internet hit before it was revealed, to much surprise, that the song was an AI-generated fake in which neither of the featured musicians actually participated. In such cases, it is reasonable to wonder if the growing capability to replicate a singer's voice could change the demand for their existing work and their incentives to produce new work.

4.1. Background

The saliency of this concern has grown in recent years as the technology behind "deepfakes" has grown. It is now possible to generate online representations with plausible imitations of voice, tone, cadence, and synchronization between video and voice. The technology has evolved to the point that many listeners and viewers cannot distinguish between real and fake. Moreover, the availability of technology has expanded from very recently being the exclusive domain of a handful of sophisticated developers to being widely accessible to the public.

16 The heuristic distinction between "nefarious" and "non-nefarious" is not intended to signal the moral or ethical virtues of some uses over others, but rather to distinguish uses with primarily commercial goals and commercial consequences from other uses. For the purposes of this discussion, we consider a usage to be non-nefarious if its primary goal is commercial in nature, and the harm caused by that usage, if any, is primarily commercial. For example, producing and publishing a new song in the style of a particular artist is likely to be a non-nefarious usage. The primary goal is likely to earn money through the sale of the song, and the primary harm done to the artist is commercial insofar as it might reduce sales of their original works. The artist may suffer secondary non-commercial harm as well if, for example, the song based on their style somehow diminishes the artist's perceived artistic integrity. Nonetheless, the primary harm is likely to be commercial.

Rights of publicity are currently covered by an array of state laws. Most states allow for some rights; however, NILs are treated differently in different states (U.S. Copyright Office, 2024). Currently, many celebrities do not solely rely on state rights of publicity laws to protect their NIL. Many use trademarks and other national IP laws to protect the brands they build around their NIL, perhaps due in part to ambiguities and inconsistencies across state laws (Leaffer, 2007).

There are numerous different contexts in which rights of publicity may arise, many of which may have differing implications for optimal policy. For example, consider the use of AI in voice acting. Ad agencies hire voice actors for commercial purposes, and production may require recording in a studio. That necessitates a visit by the actor. In post-production activities, however, changing a word, altering a sound, or inserting additional material may be required to achieve a desired outcome. Today's technology makes imitating an actor's voice much more manageable. It also makes this post-production work much easier and cheaper insofar as it obviates the need to coordinate with the voice actor to record retakes in the studio. This is a relatively straightforward scenario where simple contracting will likely lead to the efficient outcome. Given that the voice actor will need to contract with the ad agency for the use of their voice in the initial recordings, and their voice cannot be easily replicated without those initial voice recordings, the value of future imitations of her voice can simply be built into the contract.

Another example may be the virtual resurrection of a dead actor or the creation of younger versions of older actors with the crafty use of pre-existing recordings. This was the case for a recent work produced by Will Sasso and Chad Kultgen. Sasso and Kultgen trained an AI voice on many recordings of George Carlin that were produced before his death in 2008. They wrote some material and had the AI read it, adding a laugh track. This became a video called "George Carlin: I'm Glad I'm Dead." They loaded it on YouTube, which enabled them to collect a portion of the ad revenue associated with the video. George Carlin's estate sued Sasso and Kultgen for copyright infringement (Kuo, 2024).

This scenario may be more challenging to resolve through contracting than the voice actor example. Since the process uses existing recordings (made before AI imitation was a possibility), the actor (in this case Carlin's estate) would not have had the opportunity to build the value of future imitation into the contract for the original recordings. Moreover, recordings on which the AI could be trained may have diffuse or ambiguous ownership and be available through disparate channels. Nonetheless, despite the challenges, it is possible that, given sufficiently clear legal rights, contracting alone may produce the efficient outcome. Indeed, the Carlin example has recently come to a contractual resolution (Sisario, 2024). Whether it is the economically efficient resolution remains to be seen.

One thing the voice actor and the Carlin examples have in common is that the AI involved with producing the imitations was trained on bounded sets of works (new recordings of the voice actor or existing recordings of a dead actor). That is, they resemble the "bounded models" discussed in Part 1. However, this will not always be the case.

A contrasting example is when some large unbounded set of NILs are used to create a composite work. Such situations have become more frequent in high-end filmmaking for things like crowd scenes or military sieges. Generative AI can simulate many faces in movie crowd scenes, for example, and the technology for doing so may draw on many sources and NILs, necessitating many negotiations. This example may present intractable contracting issues without some sort of collective intermediation. This is because it would be impossible to attribute the likeliness of any one person in the crowd scene to a set of specific actors (or other individuals) whose images were used to train the AI. Nonetheless, the studio is gaining value from those images, and the actors are losing value (to the extent the studios do not need to hire live actors for the crowd scenes, diminishing the studio's costs and the potential earnings of such actors). Because of the contracting difficulties, the extent to which this sort of AI-augmented production can continue will depend largely on how rights of publicity are defined and enforced.

These examples involve varying degrees of contracting and negotiation today. Norms and standards are already emerging, for instance, in the market for voice acting. Similarly, the recent strike in Hollywood focused on developing contracting norms and standards for compensation among screen actors (MacGillis, 2024). It is sensible to expect these contracting practices to naturally evolve as the technology evolves. That frames questions around whether these negotiations are sufficient to efficiently cover most situations and how the introduction of a new federal right might alter the evolution of those practices.

4.2. The Efficient Assignment of Rights

There is a recent legislative push in the United States to supplement state right of publicity laws with federal laws. For example, the NO FAKES Act that was introduced in 2024 in the 118th Congress (S. 4875, H.R. 9551) seeks to address rights relating to digital replication of an individual's voice or visual likeness. However, before considering the issues around harmonization, it is necessary to consider the social implications of these laws in the first place.

As a preliminary matter, much of the discussion relating to rights of publicity revolves around some notion of fairness rooted in something analogous to moral rights.[17] There are also democratic concerns woven into these discussions as rights of publicity can be viewed as limiting free expression, with First Amendment implications. While there is utility in aligning policy with a society's normative values, and democratic concerns should be accounted for, here we consider rights of publicity in the context of explicit intellectual property policy objectives. We consider whether these rights enhance scientific or cultural innovation.

This context presents two competing factors, similar to those that arise in copyright: providing market power to the owners of these rights may incentivize some socially desirable behavior (e.g., the production of creative works), but it also limits the public's access to those things (to include access by future creators who may wish to build upon earlier works).

For example, giving a singer the exclusive rights to their voice (including replicas of their voice) would likely benefit the singer, but it is not clear that it would necessarily be socially beneficial. To the extent that people other than the singer wish to use imitations of their voice to create new valuable works,

17 Moral rights are the rights "to claim authorship of the work and to object to any distortion, mutilation or other modification of, or other derogatory action in relation to, the said work, which would be prejudicial to his honor or reputation" (Legal Information Institute, 2024).

they will face the transaction costs of identifying and contracting with the singer. Then, they will face the monopolistic pricing enabled by the rights. Additionally, this could bring about certain, potentially inefficient, legal ambiguities around what constitutes a copy of the singer's voice. These issues could hinder creation that would otherwise be of social value. Alternatively, if we do not allow such rights (or, equivalently, assign nonexclusive rights to the public), that can diminish the singer's incentives to create their original works since they will not be able to appropriate as much value from their voice.

Extending this reasoning to the Carlin example, it is unclear, given the aversion of the Carlin estate, whether Sasso and Kultgen would have been able to produce their work had they not believed they had the right to imitate Carlin's voice. Notably, the work was based on new artistic expression put into the voice and style of Carlin, and the work has been at least moderately popular amongst consumers. Thus, it is plausible that Sasso and Kultgen's production generates a net positive social value. On the other side of the equation, Sasso and Kultgen's work will not affect George Carlin's incentives to produce new creative works, as he is no longer doing so. However, the perception regarding whether Sasso and Kultgen ultimately had the right to imitate Carlin's voice may affect the incentives of living creators insofar as it signals how their NILs may eventually be treated.

This problem can be considered as the efficient assignment of property rights. In that sense, there are two potential claimants to whom rights can be assigned: the subject of the NIL can be assigned the right to exclude use of NILs or the public can be assigned the right to use them (which is equivalent to having no rights of publicity). The question is, which assignment of rights would generate more social welfare? The Coase theorem tells us that under a specific set of circumstances, total social welfare is invariant to how property rights are initially assigned — so long as property rights are clearly defined, they will eventually flow to their highest value user through bargaining (Coase, 1960).

For example, if we assign voice rights to a singer, third parties wishing to replicate their voice can negotiate for such use. If the value created by the third party in its use of the singer's voice is greater than the cost of such use to the singer, then the parties will successfully negotiate a contract, and the third party will use their voice. Otherwise, no bargain will be struck, and the singer will retain their rights. Alternatively, if we initially assign the rights to the public (i.e., declared NILs to be public domain) and the singer wishes to prevent the use of their voice by third parties, they will have the opportunity to negotiate with them to prevent such use. Again, if the value to the third party of using their voice exceeds the cost (in terms of both extrinsic and intrinsic costs) to the singer, the third party will retain the rights; otherwise, they will be sold back to the singer. In either case, if the conditions of the Coase theorem hold, the rights will eventually go to the highest-value user and maximize social welfare. The only change is in the distribution of wealth between the singer and the third parties wishing to use their voice (the party to whom rights are initially assigned is necessarily better off as a result, so long as the rights have some positive value).

However, a core implication of the Coase theorem is that there are (commonly occurring) circumstances in which the assignment of property rights to a specific claimant will, in fact, generate greater social welfare than alternative arrangements of property rights. Thus, it is helpful to examine those conditions and their implications in the context of the rights of publicity.

A preliminary (and common) issue concerning Coase and the efficient assignment of rights revolves around transaction costs. In particular, one of Coase's key findings is that welfare is invariant to the initial assignment of rights *only* if transaction costs are sufficiently low. When transaction costs are non-trivial, the implication is that welfare is maximized by assigning rights to the party with the highest value for those rights. In the context of negotiating rights of publicity, there are likely many scenarios where transaction costs are high. Here, whether that implies assignment to subjects of NILs or to the public is an empirical question.

A complication in the matter of transaction costs also arises insofar as transaction costs here are not invariant to how rights are initially assigned. Assigning rights to subjects of NILs would likely result in lower transaction costs than the alternative. Returning to the earlier example, if rights were assigned to a singer, transaction costs would arise as all third parties wishing to use the singer's voice would need to identify them as the rightsholder, seek them out, and engage in negotiations. Conversely, if non-exclusive rights were initially assigned to the public (i.e., no rights of publicity), the singer would have to seek out every third party wishing to use their voice, decide whether they wish such use to proceed, and if not, contract with the parties to not engage in that use. In most instances, it will be less costly and more practical for a set of third-party prospective licensees to seek out a singular rightsholder than for that singular rightsholder to identify an indefinite set of third parties attempting to use their voice. Additionally, if usage rights are assigned to the public, there may be issues with holdup and otherwise non-productive rent-seeking (e.g., content with no social value that is created solely to extort payment from the subjects of NILs). This implies that, if the value of the rights to the singer and the value to third parties is sufficiently similar, assigning rights to subjects of NILs will be welfare maximizing.

Another potential issue is the asymmetry in bargaining power between the subject of NILs and those who wish to use the NILs of others. Suppose a singer is well-known and has a highly distinctive voice. They may have significantly greater bargaining power than third-party producers who wish to use their voice since there is one seller (the singer) and many potential buyers (anyone wishing to use their voice). If rights are initially assigned to them, they will charge something approaching the monopoly price, which is necessarily more than the cost they incur by letting others use their voice. Alternatively, if (non-exclusive) usage rights are initially assigned to third-party users of their voice (i.e., if NILs are public domain), the singer may be willing to pay some third parties not to use their voice, but the price they are willing to pay will be no more than the cost they incur by letting them use their voice. The marginal cost pricing of the second scenario will produce a higher quantity of output (in terms of third-party productions using the singer's voice) than the monopolistic pricing of the first scenario. Taking this factor in isolation suggests that, in scenarios where subjects of NILs have sufficient market power, welfare would be greater if usage rights were initially assigned to the public.

However, as mentioned previously, this also has implications for the distribution of wealth. The party to whom rights are initially assigned will command a greater portion of the value generated by the singer relative to the alternative. While static measures of social welfare are usually not a function of wealth distribution, that distribution may have meaningful implications for the incentives faced by creators and, thus, for dynamic welfare.

Not assigning rights of publicity to NIL subjects (or, equivalently, assigning non-exclusive rights to the public) diminishes their potential wealth, thereby weakening their incentives to create or otherwise invest in their NILs. A singer whose voice can be freely appropriated by others will have fewer incentives and resources for creating original music. At the same time, having NILs in the public domain will increase the volume of follow-on creation by third parties. Thus, whether this factor favors assigning rights to NIL subjects or to the public depends on the marginal value of "original" works compared to the marginal value of follow-on works.

Ultimately, in assessing the welfare effects of rights of publicity, we face several inextricable factors with countervailing implications. Which of those implications will dominate — or the sign of the composite effect — is likely context-dependent, highlighting the need for empirical exploration of several key factors. First, we need to understand the value of the creative works produced by the subjects of NILs relative to that of third-party works based on the NILs of others. Then, we need to know the degree to which rights of publicity affect the production of the two. This involves understanding the incentives laws may have concerning NIL subjects, the degree of market power the laws would confer on them, and the demand characteristics of third parties wishing to use the NILs of others.

Another consideration that arises if rights of publicity are enacted is legal ambiguity. There is a trade-off between legal ambiguity and settling legal norms in this area. This trade-off is particularly acute because creators could inadvertently produce works that risk being seen as too closely related to someone else's "style," for example. If "style" lacks a legally precise definition or the legal tolerance for stylistic similarity is ill-defined or inconsistent, ambiguity emerges. Ambiguity in property rights leads to socially inefficient outcomes. Under certain circumstances, the clarity of property right laws can be more important than the party to whom they are assigned.

This issue may be a salient factor when considering the national harmonization of rights of publicity. The patchwork of state-level rights of publicity that currently exist doubtlessly brings about legal ambiguity in jurisdictional coverage (Rothman, 2024). However, much of the ambiguity in the scope and definitions within state statutes has been worked out through state judicial systems over several decades. National harmonization would necessarily resolve much of the jurisdictional ambiguity. Still, a new statute would introduce new definitional imprecisions that would need to be worked out by the federal judicial system, likely over a long period. In assessing the merits of national harmonization, we need to measure the adverse effects of the current jurisdictional ambiguity and compare that to the potential new statutory ambiguity that would exist for some time after national harmonization.

References

Coase, Ronald H. "Law Economics." *Journal of Law and Economics* 3 (1960): 1-44.

Coons, Chris. "No Fakes Act One-Pager." Last updated July 31, 2024.
https://www.coons.senate.gov/imo/media/doc/no_fakes_act_one-pager.pdf

Kuo, Christopher. "George Carlin's Estate Sues Podcasters Over A.I. Episode." *The New York Times*, January 26, 2024.
https://www.nytimes.com/2024/01/26/arts/carlin-lawsuit-ai-podcast-copyright.html

Leaffer, Marshall. “The Right of Publicity: A Comparative Perspective.” *Albany Law Review* 70, no. 4 (2007): 1357-1374.

“Moral Rights.” *Legal Information Institute*. Last accessed November 20, 2024. https://www.law.cornell.edu/wex/moral_rights

MacGillis, Alec. “Hollywood Writers Went on Strike to Protect Their Livelihoods from Generative AI. Their Remarkable Victory Matters for All Workers.” *Brookings*, November 16, 2024. https://www.brookings.edu/articles/hollywood-writers-went-on-strike-to-protect-their-livelihoods-from-generative-ai-their-remarkable-victory-matters-for-all-workers/

Rothman, Jennifer. *The Right of Publicity: Privacy Reimagined for a Public World*. Harvard University Press, 2024.

Sisario, Ben. “George Carlin’s Estate Reaches Settlement After A.I. Podcast.” *The New York Times*, April 2, 2024. https://www.nytimes.com/2024/04/02/arts/george-carlins-settlement-ai-podcast.html

U.S. Copyright Office. *Copyright and Artificial Intelligence: Part 1—Digital Replicas Report*. 2024. https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf

5. The Effects of AI Ingestion on Rightsholders' Incentives

Principal contributors: Michael D. Smith and Rahul Telang

Foundation models are known for ingesting huge volumes of data, often including copyrighted materials. In the context of generative AI, those models are often ingesting creative works, many of which fall under copyright protection. This process involves producing a digital copy, often without the permission of, or payment to, rightsholders (Library Copyright Alliance, 2023; Sag, 2023). There are many debates around the legality of this practice (Congressional Research Service, 2023). However, putting aside questions of how to interpret current laws, a more fundamental question behind the legal and economic debate is to what degree does allowing or restricting this practice serve the economic objectives of copyright.

One of the goals of copyright is to facilitate cultural and scientific innovation — a goal that requires balancing the economic rewards that can be captured by producers of creative works with their ability to access existing works as part of the creative process. For example, a writer may read existing works to gain a knowledge base or inspiration for writing their new work. That cumulative creative process is the foundation of innovation, and generative AI is thought of by some as doing a similar thing as it ingests existing works and produces something ostensibly new (with the significant distinction that humans traditionally pay rightsholders, in one form or another, for what they ingest, and the extent to which developers have paid for what their algorithms ingest varies greatly across different types of content). In that sense, the algorithms are engaging in the sort of innovation process that copyright policy aims to encourage. However, it is important to consider how AI may shift the balance between incentives and access and what policies could recalibrate that balance if necessary.

There are two general aspects to consider here. First, what social benefits come from developers having access to training materials, and second, what are the implications for the incentives of human creators to produce works? Here we deal with the latter, leaving the former for Part 6. Nevertheless, policymakers should not take either of these aspects in isolation; welfare maximization requires optimally balancing the two.

5.1. Impact on Commercial Incentives to Create

The impact that ingestion has on the commercial incentives of human creators largely depends on the extent to which such use is incremental to, or a substitute for, other anticipated uses. Some argue that the ingestion of creative works crowds out other commercially valuable uses of those same works — a substitutional effect. For example, consider news reporting. To the extent that developers gain access to a newspaper's content for free and incorporate that into their output, that might impact a news organization's ability to maintain revenue from its subscribers. If a user could ask an LLM to summarize a newspaper's stories for that day and then ask it to expand on particular stories, the user would be less motivated to continue their subscription to that newspaper. It may also have negative effects through secondary channels. For example, if an LLM can act as a news aggregator, it would compete with other news aggregators who license content from the newspaper. That would likely diminish the newspaper's licensing revenues.

In contrast, others argue that the use of works in training AI models is incremental and does not directly harm the commercial interests of rightsholders. This is, in some ways, analogous to the arguments made by Napster in its dispute with the music industry over file sharing.[18]

In the extreme, and perhaps unlikely, case that use by developers is entirely incremental, unpermitted ingestion would not diminish existing commercial incentives to create (this does not preclude it from affecting intrinsic incentives, as discussed below). While, in this scenario, rightsholders are made no worse off commercially, the use of human creative works in AI systems can provide a mechanism to increase creators' incentives beyond what they would be in the absence of ingestion. To the extent that rightsholders are able to share in some of that newly created value, (e.g., through licensing agreements with developers) it will increase the overall value of the work to the creator, which can enhance incentives to create, thus increasing overall creative output. Moreover, allowing rightsholders to share in all the various ways in which their work creates value may help align production incentives with demand.

The other extreme (and also unlikely) case is that ingestion is entirely substitutional for other commercially valuable uses. If this is done without compensation to rightsholders, it will serve purely as a transfer of wealth from rightsholders to developers (and possibly to consumers), clearly diminishing human creators' incentives. Diminished incentives would lead to a reduction in new human-generated works (in terms of quantity, quality, or both).

The more likely case lies somewhere between the two extremes, where ingestion is sometimes an incremental use and sometimes a substitutional use. This is similar to early debates around internet piracy, where one side maintained that piracy diminished rightsholders' commercial interests (a substitutional use), and the other side contended that consumers of pirated materials would not have purchased an authorized version in the absence of piracy (an incremental use).[19] There are numerous empirical studies showing that the unauthorized use of copyrighted materials reduces sales through legal channels but at a rate of less than one-to-one (i.e., many, but not all, instances of piracy would have resulted in a commercial sale if piracy were not an option).[20] Notably, related research has also demonstrated that reduced revenue from piracy, in turn, reduces the quantity and quality of creative production (Danaher et al., 2017; Telang and Waldfogel, 2018). Therefore, to understand where and how ingestion may affect human creators' incentives, it is necessary to empirically examine the degree of substitution under various circumstances. This relates to the discussion in Part 2 regarding labor displacement.

---

18 See for example, Napster's opening brief before the U.S. Ninth Circuit in 2000: "This case is not about any diminution in the value of Plaintiffs' copyrights; none has occurred or is reasonably foreseeable as the result of Napster. This case is about whether Plaintiffs can use their control over music copyrights to achieve control over Napster's decentralized technology and prevent it from transforming the Internet in ways that might undermine their present chokehold on music promotion and distribution" (Los Angeles Times, 2000).

19 For a review of earlier economic literature on internet piracy and its impact on creative industries, see Handke (2011).

20 See, e.g., Danaher et al. (2020).

A second way in which unpermitted model training may negatively affect welfare is through distribution incentives. If developers are not subject to *legal* restrictions around the use of copyrighted materials, then rightsholders may subject them to *practical* restrictions. Rightsholders who openly provide access to their content will have strong incentives to cease such practices. For example, rightsholders who make their content open to all and monetize it by selling advertisements may be incentivized to switch to a subscription model. If content is behind a paywall, developers will not be able to freely take it, forcing licensing negotiations. This will not only diminish access by model developers (the intended consequence) but also limit access by consumers (an incidental consequence) thus diminishing consumer welfare. This effect is not contingent on whether ingestion is incremental use or substitutional use. The fact that rents can be more easily extracted from developers by limiting public access will serve as a sufficient incentive to do so.

A third, more subtle, way in which unfettered ingestion may affect rightsholders' incentives and welfare is through the potential degradation of quality and credibility. As discussed in Part 2, AI may produce tarnished replicas of the works they ingest. For example, an AI-generated summary of a news story may mischaracterize key facts, contain incoherent statements, or otherwise produce content of low quality. To the extent that the low quality of the AI-replica is conflated with the quality of the original, rightsholders' credibility and brand value can be diminished. This, in turn, diminishes rightsholders' incentives to invest in their reputation or brand through high-quality works since they lose a degree of control over those things. This may also have negative democratic implications to the extent that public trust in certain sources of information is undermined.

Ultimately, it is likely that in some instances, the commercial incentives of human creators to produce creative works will be diminished as a result of generative AI model ingestion. It is also likely that ingestion will cause rightsholders to limit public access to their works to some degree and result in diminished incentives around brand and reputation building. It is important to develop empirical evidence around the scale of these problems in order to weigh the welfare harm against the potential welfare gains discussed in Part 6.

5.2. Impact on Intrinsic Incentives to Create

Another consideration is how ingestion by foundation models might affect the intrinsic incentives of creators. One characteristic of creative works that complicates economic analyses of them is that creators often face strong intrinsic incentives to create. Humans have a natural drive to be creative and fulfilling that drive generates utility for the creator. The utility that comes from creativity in and of itself serves as an incentive to create works. When a creator is deciding whether to produce something, they will implicitly consider both the intrinsic incentives (the internal satisfaction they will gain through creating the work) and the extrinsic incentives (the commercial value derived from others consuming the work).

The discussion in this part has so far focused on how ingestion may affect extrinsic incentives. However, in terms of influencing behavior, intrinsic incentives operate no differently than extrinsic incentives: the higher the reward for engaging in an activity (whether the reward derives from within or is captured through some external transaction), the more people will engage in that activity.

Thus, it is useful to consider how ingestion may also change intrinsic incentives for creators to the extent that such changes will affect creative output. There is a (relatively small) body of literature on the role of intrinsic incentives and how they interact with and compare to extrinsic incentives (Johnson, 2012). However, here, we focus narrowly on how intrinsic incentives might change as a result of ingestion.

Two factors may contribute to changes in intrinsic incentives. First, some of the intrinsic utility derived from creativity is contingent on recognition. Some creators may be distributing works in part because they wish to be recognized as creators. Thus, whether attribution is granted in the use or distribution of a work may affect the creator's intrinsic incentives to create that work. Second, some creators may be concerned with the artistic integrity of their work or the messages their works are used to communicate. For example, if a song is manipulated to convey extreme social or political messages that differ from the creator's intended message, that may decrease the intrinsic utility that the creator gains from producing the song. In this context, the courts have historically given rightsholders a degree of deference in determining how their creations are used.[21]

Avoiding a use that is contrary to the creator's intended message requires that creators be able to exercise some degree of control over how their works are used; this is usually facilitated by license contracts. When that control is eroded and such outcomes become more likely, the net intrinsic utility that a creator can expect to gain through creation is diminished, thus diminishing their intrinsic incentives to create.

The challenge in the context of model training is that once content enters the training corpus it is difficult to place guardrails around how it will be used in output and therefore difficult for creators to anticipate whether and at what price they might be willing to license their content. Similarly, it is difficult after ingestion to identify the contribution of any particular work to some output. This is both a technical and a practical limitation. Developers currently assert that, because of how the technology functions, it is not possible to measure the impact that any one work on the ingestion side has on any one work on the output side, thus contributing creators cannot actually be identified. It is unclear whether this is impossible or only cost prohibitive.[22] Nonetheless, developers maintain that tracing contributions is not an option.

Moreover, even if it were technically possible to identify contributing creators, foundation models work by assigning weights to millions or billions of existing works, and they can use any given work for multiple purposes at various stages of training. Every work in the model with a non-zero weight (which, again, can be millions or billions) could reasonably be considered as a contributor. Identifying

21 For example, in *Soderberg v. Cleanflicks*, there was a dispute around whether a third party (Cleanflicks) had the right to edit films to remove coarse language or adult content. Cleanflicks was compensating the rightsholder (by purchasing the original unedited DVD) and distributing the edited content to an audience that, purportedly, would not have purchased the unedited version. The court decided that even though the editorial changes Cleanflicks made to movies arguably would not hurt the rightsholders' market, that financial consideration was outweighed by the rightsholders "right to control the content of the copyrighted work." See https://casetext.com/case/clean-flicks-of-colorado. Note, that subsequent to this ruling Congress passed the Family Movie Act of 2005 amending section 110 of Title 17 to allow certain modifications to films under specified circumstances.

22 See, e.g., the recent "Safety by Design" agreement among various generative AI companies (including Amazon, Anthropic, Google, Meta, Microsoft, OpenAI, and Stability AI) to, among other things, "[d]evelop state of the art media provenance or detection solutions" to reduce the incidence of AI generated child sex abuse materials (Thorn, 2024).

and crediting millions or billions of creators is likely to be impractical. Nonetheless, even when the contributing works are obvious, there appears to be an unwillingness to include some sort of attribution in AI-generated works.[23]

5.3. Licensing as a Potential Solution

Beyond measuring the scale of effects discussed in the previous sections, it is also useful to consider potential solutions. If policymakers wish to maintain creative incentives at existing levels, there are many potential ways of doing so. Almost any aspect of copyright law could (at least in principle) be changed in a way that increases or decreases the amount of value rightsholders can appropriate from their work. In fact, if one can appropriately measure the impact, it could be compensated for in ways entirely unrelated to ingestion.

However, there are fewer available policy instruments for combating the perverse incentives for rightsholders to further limit public access to their works in response to ingestion. Indeed, the only potentially viable solution to that particular problem may be a licensing requirement for ingestion. For that reason, this section focuses on issues of licensing as a solution to both the creative incentives problem and the access-limiting incentives problem. Nevertheless, this solution comes with its own social costs (as discussed in Part 6) as well as practical limitations.

The licensing of copyrighted training materials is controversial. Rightsholders argue that developers are infringing on their rights under existing copyright law and on their ability to appropriate a sufficient portion of the value of their work. In contrast, developers argue that they should not have to pay to use copyrighted content that is freely accessible online because (1) it is being made freely available to anyone else who wishes to access it; (2) use in training is not harming the commercial interests of rightsholders; and (3) having to pay license fees would be prohibitively burdensome and expensive, hindering the development of AI technology.

These polarized positions are unlikely to be universally true (despite being argued by both sides in generalities). It is, therefore, helpful to consider the circumstances in which a licensing requirement may be a viable and sensible solution and the circumstances in which it is not. A natural starting point is to look at where licensing agreements for training materials are already being made.

For example, in the context of audio, the generative AI platform Stable Audio licensed "over 800,000 audio files containing music, sound effects, and single-instrument stems" from AudioSparx for use in its training corpus. In the context of news, OpenAI signed licensing deals with the Wall Street Journal, Axel Springer (Politico, Business Insider, Bild, Welt), Hearst (Esquire, Cosmopolitan, ELLE), and the Associated Press for use in OpenAI's training corpus (Bruell, 2024; Coster, 2023; David, 2023; Hearst, 2024; and O'Brian, 2023). Microsoft has also signed deals with Axel Springer, Hearts, Reuters, the Financial Times, and USA Today for use of their content in its Copilot Daily feature (Microsoft, 2024). Similar licensing deals have been struck between developers and Shutterstock, Getty Images,

23 Rightsholders have shown instances where a model's output quotes a rightsholder's work nearly verbatim, without attribution to the original author. See, e.g., Grynbaum and Mac (2023).

and Reddit (Davis, 2024; Getty Images, 2024; and Shutterstock, 2024). These examples appear to demonstrate that there may be circumstances where licensing is logistically feasible, and mutually acceptable royalty rates that will not shutter AI platforms can be negotiated.

One key aspect of these deals is that developers typically license copyrighted content when they have no easy way to get it from the public web, for example, because the copyrighted content is behind a paywall or is otherwise difficult to scrape. That is, developers tend to license when copyrights are relatively easy to enforce. Thus, a preliminary (if obvious) condition for effective licensing to occur is that rightsholders can practically restrict access to their work. As discussed in the previous section, and evident from the above examples, a natural way to do this is to put content behind a technological barrier such as a paywall. However, because that may be a socially undesirable outcome, it is useful to consider whether there are other ways of restricting access to generative AI models without diminishing public access for human consumers.

Some alternative technological solutions are currently being experimented with. For example, Nightshade is a system that can be applied to images so that if those works are ingested by a model, they will inflict (potentially existential) harm on it (Heikkiläarchive, 2023). While that harm does not directly benefit rightsholders who employ Nightshade, it can raise the relative cost of unlicensed ingestion so much that licensing becomes the least costly alternative. Note, this solution does not prevent human consumers from accessing the work. One limitation of technological solutions such as this is that they must be continuously improved upon to stay ahead of counter technologies aimed at circumventing access limits. It may be a perpetual arms race between developers and rightsholders, which itself will impose a social cost.

Aside from technological solutions, the enforcement of legal restrictions on a model's access to training data can, in principle, have the necessary effect. This, of course, raises the question: even if rights are clearly defined to exclude unpermitted ingestion, would such rights be enforceable from a practical perspective? For enforcement to be practically effective, infringement needs to be detectable and come with a sufficiently large consequence. However, detectability may be a limiting factor, at least under the current regime, as many developers are unwilling to reveal what goes into their models (Nasr et al., 2023). Thus, for legal restrictions to be a viable solution, some amount of transparency around training data would need to be mandated. Despite the current challenges around transparency, some degree of it appears to be technically feasible. One company, Fairly Trained, for example, offers a certificate verifying that a generative AI model has been trained without copyright infringement (Fairly Trained, 2024). The feasibility of enforcement is further discussed in Part 7.

Separate from the issues around limiting access by developers, there remains a question of whether a market could arise to facilitate licensing, even if licensing were the only option for access. As noted above, developers argue that they cannot afford to license training materials.

As a preliminary matter, in order for transactions to occur, the bargaining set must not be empty—the maximum willingness to pay of the licensee must be equal to or greater than the minimum willingness to accept of the licensor. That is, developers' value of gaining the rights must be greater than rightsholders' value of retaining the rights for a licensing transaction to occur. It is not clear how often this will be the case. Foundation models ingest millions or billions of works, and developers assert that the value of any one work is, on average, trivial. Such a trivial value may not exceed the transaction cost of licensing. In those instances, no willing licensing transactions will occur. Absent any unaccounted-for externalities, the lack of a market under these circumstances is the economically efficient outcome. The implication of an empty bargaining set is that generative AI models would not be creating enough value to compensate rightsholders for whatever costs they face concerning ingestion. Forcing transactions in this scenario would reduce welfare.

Similarly, another necessary condition for a market to arise is that transaction costs be sufficiently low. Transaction costs are likely to be substantial in some cases. As discussed above, developers assert that the value of any one work to the model is generally trivial. If licensing negotiations occurred on an individual basis, even equally trivial transaction costs could outweigh the value derived from the deal. If the transaction costs associated with seeking out and negotiating with individual rightsholders are higher than the value of the content to the licensor, individual licensing will rarely occur, even if a transaction (absent transaction costs) is mutually beneficial.

This leads to another key aspect of the licensing examples discussed above: developers were likely not negotiating directly with individual creators. Rather, they were licensing from entities that controlled a high concentration of rights. In all likelihood, bargaining took place around the average value of rights within a portfolio, not around individual values. Therefore, while there were surely transaction costs associated with those negotiations, the transaction cost on a per-work basis was low enough to facilitate a license deal.

While centralized ownership of rights can facilitate efficient markets, it is not a strict requirement. It is possible that the effects of centralized ownership can be mimicked by collective intermediation. For example, collective rights organizations often help reduce transaction costs around large-scale licensing in certain industries, such as music. Many of these organizations offer blanket licenses and have well-established methods for distributing royalties among their members. Nonetheless, foundation models ingest a highly diffuse body of works, and collective intermediation may be better suited for some subsets of those works than it is for others. Developing an economic understanding of the conditions under which collective intermediation is likely to arise and facilitate the formation of markets can help policymakers better understand the impact that licensing requirements may have.

One technical challenge that a licensing requirement may bring is that such a requirement would force developers to figure out which pieces of information on the public web are copyrighted and which are not, and that hurdle might be so large as to bias what types of information are used to train models. This problem is unlikely to be entirely resolved through collective intermediation because there will always be some works that are not represented by an intermediary, making copy permissions ambiguous. Potential solutions to this issue are discussed in Part 7.

A middle-ground solution might be to allow rightsholders to opt out of having their copyrighted data used to train models. Indeed, the recent EU AI act mandates that rightsholders have the ability to opt out. This could be analogous to the "noindex" tag that some websites use to prevent their content from being scraped and archived by search engines (Google, 2024). However, such an opt-out system would need to have three characteristics to be effective. First, it would have to be broader than web-based "noindex" tags. Instead of relying on individual websites to maintain "noindex" tags, we would likely need to have some sort of centralized database of content that rightsholders do not want to be ingested, which may present substantial logistical hurdles. Second, it would need some mechanism whereby rightsholders could verify that their content had been removed from the training corpus. That poses a technical problem, which is likely solvable, but could be costly. Finally, the system would need a mechanism, either legal or economic, to ensure that developers abide by the stated wishes of website and content owners.

It is also worth noting that an opt-out regime for the ingestion of training materials would meaningfully change the nature of current copyright protections by shifting the burden for action from the user of the copyrighted material to the copyright owners. While shifting that responsibility to rightsholders would not be ideal for many rightsholders, an opt-out system may have certain economic advantages over an opt-in system depending on the circumstance. Opt-in and opt-out systems are further discussed in Part 7.

References

Bruell, Alexandra, Sam Schechner, and Deepa Seetharaman. "OpenAI, WSJ Owner News Corp Strike Content Deal Valued at Over $250 Million." *The Wall Street Journal*. Last updated May 22, 2024. https://www.wsj.com/business/media/openai-news-corp-strike-deal-23f186ba

Coster, Helen. "Global News Publisher Axel Springer Partners with OpenAI in Landmark Deal." *Reuters*. December 13, 2023. https://www.reuters.com/business/media-telecom/global-news-publisher-axel-springer-partners-with-openai-landmark-deal-2023-12-13/

Congressional Research Service. *General Artificial Intelligence and Copyright Law*. September 29, 2023. https://crsreports.congress.gov/product/pdf/LSB/LSB10922

Davis, Wes. "Reddit AI Training License Deal for User Content." *The Verge*. February 17, 2024. https://www.theverge.com/2024/2/17/24075670/reddit-ai-training-license-deal-user-content

Danaher, Brett, Michael D. Smith, and Rahul Telang. 2020. "Piracy Landscape Study: Analysis of Existing and Emerging Research Relevant to Intellectual Property Rights (IPR) Enforcement of Commercial-Scale Piracy." *USPTO Economic Working Paper* 2020–02. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3577670

Danaher, Brett, Michael D. Smith, and Rahul Telang. 2017. "Copyright Enforcement in the Digital Age: Empirical Evidence and Policy Implications." *Communications of the ACM* 60 (2): 68–75. https://dl.acm.org/doi/pdf/10.1145/2979673

David, Emilla. "Stability AI Releases AI Audio Platform." *The Verge*. September 13, 2023. https://www.theverge.com/2023/9/13/23871635/stability-ai-generative-audio-model-platform

Getty Images. "Getty Images Launches Commercially Safe Generative AI Offering." *Getty Images Newsroom*, October 3, 2024. https://newsroom.gettyimages.com/en/getty-images/getty-images-launches-commercially-safe-generative-ai-offering

Grynbaum, Michael, and Ryan Mac. "New York Times Sues OpenAI and Microsoft Over Use of Copyrighted Work." *The New York Times*. December 27, 2023. https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html

5. The Effects of AI Ingestion on Rightsholders' Incentives

Handke, Christian. "The Effects of Unauthorized Copying." Chapter 4 in *Economic Effects of Copyright: The Empirical Evidence So Far*, Report for the National Academies of the Sciences. April 2011. https://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga_063399.pdf

Heikkiläarchive, Melissa. "This New Data Poisoning Tool Lets Artists Fight Back Against Generative AI." *MIT Technology Review*. October 23, 2023.

Johnson, Eric E. 2012. "Intellectual Property and the Incentive Fallacy." *Florida State University Law Review* 39: 623–680. https://heinonline.org/HOL/Page?handle=hein.journals/flsulr39&div=29&g_sent=1&casa_token=&collection=journals

Microsoft Corporate Blogs. "An AI Companion for Everyone." *The Official Microsoft Blog*. October 1, 2024. https://blogs.microsoft.com/blog/2024/10/01/an-ai-companion-for-everyone/

Nasr, Milad, Nicholas Carlini, Jonathan Hayase, Matthew Jagielski, A. Feder Cooper, Daphne Ippolito, Christopher A. Choquette-Choo, Eric Wallace, Florian Tramèr, and Katherine Lee. "Scalable Extraction of Training Data from (Production) Language Models." *ArXiv* abs/2311.17035 (2023). https://arxiv.org/pdf/2311.17035

O'Brian, Matt. "ChatGPT-Maker OpenAI Signs Deal with AP to License News Stories." *Associated Press*. July 13, 2023. https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a

Sag, Matthew. "Copyright Safety for Generative AI." *Houston Law Review*, Vol. 61, No. 2, 2023. https://heinonline.org/HOL/Page?handle=hein.journals/hulr61&div=15&g_sent=1&casa_token=&collection=journals

Shutterstock, Inc. "Shutterstock Partners with OpenAI and Leads the Way to Bring AI-Generated Content to All." *Shutterstock Press Room.* Accessed November 20, 2024. https://www.shutterstock.com/press/20435

Telang, Rahul, and Joel Waldfogel. 2018. "Piracy and New Product Creation: A Bollywood Story." *Information Economics and Policy* 43: 1–11. https://www.sciencedirect.com/science/article/abs/pii/S0167624517301348

Thorn. *Safety by Design for Generative AI*. 2024. https://info.thorn.org/hubfs/thorn-safety-by-design-for-generative-AI.pdf

6. Developers' Access to Training Data

Principal contributor: Abhishek Nagaraj

Generative AI systems require a vast amount of computing and technical expertise to build, and importantly, they need to be trained on large corpora of text, visual, and multimedia information to achieve their current level of performance. The need for more expansive corpora is likely to intensify as developers vie for ever more complex and sophisticated models. To a large degree, training data, often in the form of copyrightable works, serve as the fuel driving advances in AI technologies. There will likely be improvements in the proverbial fuel efficiency of AI as developers fine-tune their machinery, but at this point it seems likely that the demand for training materials will continue to grow (possibly to the point of exhausting the existing supply). Moreover, even as developers become more efficient at exploiting smaller sets of higher quality data, the volume of data available from which to select a training set will likely remain a significant factor in technological advancement. Additionally, the demand for training data will intensify as more new market entrants compete to offer specialized models.

The legal and ethical debates around generative AI, its utility, and its access to training data abound. Analogous to broader debates about the role of intellectual property in the age of AI, the economic questions here are complex. While unrestricted access would almost certainly be an advantage to developers (at least in the near term, putting aside any potentially negative long-run dynamics), this can also impose a cost to rightsholders and have negative implications for creative incentives. Similarly, the strict prohibition of use without explicit permission is likely to impose meaningful transaction costs, which could slow technological development. As is often the case, the efficient policy is likely at some point in the space between these extremes; however, that is a wide and multidimensional space.

Carefully evaluating the pros and cons of different policy responses across numerous dimensions will require a clear articulation of the critical forces at work and further research on vital empirical margins. The policies ultimately chosen will determine not only the social value that generative AI systems will create (or destroy) but also influence which companies and developers will shape its development. This part provides a broad framework that helps progress these points and articulates research questions that are likely to arise.

Generally, we discuss the implications of different policies on the performance of generative AI systems, acknowledging that what performance means is still unresolved. Currently, performance is assessed by developers and researchers through quantitative and qualitative methods (including using AI in the evaluation).[24] While performance is pragmatically assessed using easy-to-measure

24 For example, for text-based models, one standard metric is the BLEU score (Bilingual Evaluation Understudy), which compares the model-generated text to a set of high-quality reference texts, assessing the closeness of the match. In image generation, metrics like Fréchet Inception Distance (FID) are used to measure the distance between feature vectors of authentic and generated images, indicating how similar they are to authentic images in distribution. In both realms, human evaluators are employed to assess output quality (Heusel, et al., 2017; Microsoft, 2024).

characteristics of output, those measures are not always well founded in the needs or preferences of generative AI users. Nonetheless, in the long run, the performance of these systems will be determined by realized outcomes such as adoption, use, and revenue.

Performance assessment issues aside, it is well understood that high-quality foundation models currently require vast amounts of training data and that, generally, more extensive and higher quality training data sets are quite clearly associated with improved performance (Liu, 2024). Therefore, barring any copyright restrictions or limitations from rightsholders, a developer would typically want access to all possible data, even if they intend to only select a subset of it for training. However, as discussed in Part 5, there are many reasons why rightsholders may wish to restrict developers' access to their works. For any policy regime that does restrict access at some level, developers will face a tradeoff between the cost of access and the performance of their models. The specific characteristics of such policies will determine the contours of that tradeoff. Thus, the meta-question for economists as it relates to developers and the performance of foundation models is quite simple — how do different access regimes affect the eventual aggregate performance of generative AI models?

A sensible way to approach this question is to understand the contingencies that influence this overall tradeoff. That understanding will help us develop a framework sensitive to the factors of technological advancement. Three key contingencies are particularly significant: (1) heterogeneities in the quantity, diversity, and access regimes for training data; (2) heterogeneities in developer type (e.g., corporate versus startup or closed source versus open source); and (3) heterogeneities in user needs and content types. The remainder of this part discusses each of these contingencies in turn.

6.1. Quantity, Diversity, and Access Regimes for Training Data

While more training data is generally better for model performance, not all data contributes equally to the performance of a model (Liu et al., 2024). Training data also demonstrates diminishing returns to scale, such that, after a certain point, each additional piece of training data has a smaller and smaller effect on model performance (Liu et al., 2024). These facts bring about a number of critical empirical and theoretical questions.

First, it will be useful to know the marginal value of different types of training data to performance. For example, if the NY Times archive was removed from certain LLMs, by how much would performance degrade, if at all, and does that differ from the marginal value of, say, online restaurant reviews? Relatedly, it will also be useful to know how the marginal value may differ for different types of uses or user prompts.

Separately, to establish a baseline, it will be useful to determine the performance achievable by models trained only on public domain content compared to models trained on copyrighted works. Here again, understanding how or if that performance differs for different uses or user prompts is important. Taking this a step further, it is also useful to look not only at the differences in performance but also the differences in consumers' willingness to pay for access to models of varying performance levels. Differences in performance may not translate into differences in willingness to pay. That relationship is contingent on consumers first perceiving differences, and second, on them valuing improved performance in a particular dimension. There may be applications where users are perfectly content with "pretty good" output such that any marginal improvements in performance would be lost on them. For example, if an author is using ChatGPT to brainstorm ideas for something they intend to eventually create on their own, they may not care very much if the grammar or syntax of ChatGPT's output is slightly muddled. In contrast, there may be use cases, such as high-end film production, where the quality of the AI output is of paramount importance.

As a corollary to this, an examination of the performance of models trained on "clean" sources like Adobe's Firefly system compared to rivals like Midjourney, where data provenance is murkier (Longpre et al., 2024), will be similarly useful. These clean-sourced models have demonstrated that foundation models are possible to develop without the unauthorized use of training materials. However, it is critical that we understand the differences in utility, performance, and user value between these differently sourced models. It may be that so called "clean" sourcing is viable, but only for models aimed at certain use cases. Determining the bounds within which these models are viable will help us more accurately anticipate the potential effects of different policy regimes.

Finally, we need to understand how developers would react to the imposition of different access regimes in terms of substituting between various forms of training data. For example, if access were restricted to, say, newspaper archives (hence requiring a license and a fee), could developers simply rely exclusively on online current event discussion forums (e.g., Reddit)? And if so, what, if any, degradation in performance would result from such a switch? This is important because it will heavily influence licensing markets. If certain types of materials are carved out for restriction, the substitutability between the restricted and unrestricted materials will determine the market power of rightsholders in restricted materials. Returning to the example, if newspapers are restricted but online forums are not, the degradation in performance that would come from switching to only training on online forum materials will determine how much bargaining power newspapers will have in license negotiations with developers.

6.2. Variations in Types of Developers and the Policies That Advantage Some over Others

When considering the impacts of different copyright regimes on generative AI, it is useful to recognize the broad spectrum of the organizational forms and funding models for innovation that exist in this space. Any copyright policy (or lack thereof) toward training data access will inevitably advantage some types of developers and disadvantage others. Knowing where those advantages and disadvantages may arise will allow us to determine policy in a more deliberate way.

Competitive advantages and disadvantages will likely arise along many dimensions — some surprising and others predictable. Within the latter category, two dimensions seem particularly relevant: (1) open-source versus closed-source and (2) large incumbent developers versus startup developers.

In terms of open-source versus closed-source, models vary in how transparent they are; some share source code and reveal training data (open-source),[25] others do neither (closed-source), and some

25 Open-source here pertains to both volunteer communities of developers (e.g., the group that develops the Linux kernel) as well as academic contributors who are not in the business of developing practical AI systems, but instead testing, validating, and proposing new innovations that could shape future development.

fall somewhere in between. Open-source models are transparent about all aspects of model design and training, including, notably, the training data they use. Because such models are both open to scrutiny and permit a diverse set of follow-on users to modify them as they please, developers of such models need to be much more vigilant about copyright violations and the use of training data of questionable origins.

Further, open-source models are often volunteer-funded and have very limited resources to acquire and license training data. Even if they did, such licenses would need to support not just the use of the training data in the model but the wide release of the underlying data for follow-on use as well. Greater restrictions on access to training data will likely shift the relative advantage to closed-source models. Thus, in determining the degree of restrictions, it is useful to understand the relative value of open-source and closed-source models. This is a question in need of empirical and theoretical inquiry.

A similar question arises when considering the heterogeneity between larger incumbent firms (e.g., Google or OpenAI) versus startups seeking to compete in the domain. Unlike open-source models, both organizations can presumably license data, but startups will have more limited resources. Further, their bargaining power with larger rightsholders might be relatively lower. A separate issue is that large incumbent firms often have internal data from other aspects of their business that they can use to train their own models while restricting access for competitors. For example, Google has access to the billions of emails in its Gmail repository — a corpus that most other firms do not have access to. Similarly, Adobe's Firefly model does not use potentially copyrighted images, but it can build on its vast internal repositories of stock images. Startups are less likely to have these stocks of internal training data.

For these reasons, stronger restrictions on access to training data may tilt the balance of power in favor of well-resourced incumbent developers. To understand the degree to which this will happen, empirical research into the relative elasticity of supply of startup versus large firms in the market for generative AI systems is needed. Relatedly, to the extent restricted access to training data acts as a barrier to competition, empirical and theoretical research will help us understand the implications for welfare and innovation.

6.3. Variation Across Users, Prompts, and Media Types

As previously discussed, much of the existing performance comparisons across models do not meaningly reflect the fact that model outputs are valuable only insofar as they improve productivity for downstream users (Brynjolfsson, 2023). Comparison metrics tend to be divorced from actual use cases. However, we know there is immense variation in the value of these downstream use cases and the ability of different models to add value to them.

Therefore, it is necessary to develop an economic framework that can evaluate the performance of foundation models under different copyright regimes while accounting for this diversity. Such a framework would measure variations in different types of prompts that users are likely to pose as well as the value of alternate machine-generated responses to these prompts. If, for example, most use cases involve a small set of prompts that are well answered with a public domain-based model, then the degradation in performance that comes from restricting access to copyrighted works will mainly be relevant only for less-frequent prompts, making the diminished performance more acceptable. Developing performance metrics sensitive to real-world use and value distributions would, therefore, be a foundational component of evaluating the impact of copyright policy on the development of generative AI models.

Finally, much of the previous discussion lays out the general questions without necessarily delving into the specific types of media users will demand from generative AI systems. For example, past work has shown how copyright policies tend to bind more for images as compared to text (Nagaraj, 2018), and it is likely that similar variations will be found across the spectrum of media types at issue. Empirical and theoretical work that understands and outlines these contingencies is needed. Even if divergent copyright policy based on medium of expression were not feasible, such an understanding will allow policymakers to weigh the differential effects and their social value implications.

References

Brynjolfsson, Erik, Li Danielle, and Lindsey R. Raymond. "Generative AI at Work." *NBER Working Paper* No. 31161. April 2023. Revised November 2023. https://www.nber.org/papers/w31161

Heusel, Martin, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium." *Advances in Neural Information Processing Systems* 30 (2017). https://proceedings.neurips.cc/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf

Liu, Yiheng, Hao He, Tianle Han, Xu Zhang, Mengyuan Liu, Jiaming Tian, Yutong Zhang, Jiaqi Wang, Xiaohui Gao, Tianyang Zhong, Yi Pan, Shaochen Xu, Zihao Wu, Zhengliang Liu, Xin Zhang, Shu Zhang, Xintao Hu, Tuo Zhang, Ning Qiang, Tianming Liu, and Bao Ge. 2024. "Understanding LLMs: A Comprehensive Overview from Training to Inference." *arXiv:2401.02038v2*. https://arxiv.org/pdf/2401.02038

Longpre, Shayne, Robert Mahari, Naana Obeng-Marnu, William Brannon, Tobin South, Jad Kabbara, and Sandy Pentland. 2024. "Data Authenticity, Consent, and Provenance for AI Are All Broken: What Will It Take to Fix Them?" *An MIT Exploration of Generative AI*, March 27, 2024. https://doi.org/10.21428/e4baedd9.a650f77d

Microsoft. "What Is a BLEU Score? - Custom Translator - Azure AI Services | Microsoft Learn." Last accessed November 20, 2024. https://learn.microsoft.com/en-us/azure/ai-services/translator/custom-translator/concepts/bleu-score

Nagaraj, Abhishek. 2018. "Does Copyright Affect Reuse? Evidence from Google Books and Wikipedia." *Management Science* 64(7): 2973-3468. https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2017.2767

7. Controlling the Use of Copyrighted Materials in Training

Principal contributor: Adam Jaffe

The discussion in the previous parts outlines some of the pros and cons of controlling access to copyright-protected training materials. However, this raises a yet unaddressed question: given that developers have used and will need to continue to use copyrighted works, is control practical, and if it is, how might our choices around implementing control best serve copyright policy objectives? In general, it is important to understand how our treatment of developers' past and future use of copyrighted works may affect institutional integrity, incentives, value appropriation and distribution, and transaction costs, among other things.

Many rightsholders assert that almost all (or maybe all) foundation models are based at least in part on training data that included copyrighted materials, which were taken and copied without permission from, or compensation to, rightsholders (The Authors Guild, 2023). Currently, it does not appear to be technically feasible to identify in a systematic or general way how a particular work within the training set contributes to the model parameters or outputs.

To crystalize the issues, we first consider whether a market could endogenously arise under current rules. Subsequently, we consider several simple approaches, which are intended collectively to span the range of possible policies. These approaches are considered without regard to their legal or political feasibility, and we do not advocate for any one approach over another, nor do we make specific policy proposals. Rather, the purpose here is to illustrate the economic issues and identify the points of comparison.

7.1. Could a Solution Emerge Endogenously in the Marketplace?

Although, as discussed in Part 5, some deals around training data have been struck, at the moment, incentives do not seem to be pushing toward a more generalized market solution. Many model developers seem disinterested in disclosing exactly how copyrighted material is used in training (which is a rational and predictable response to the current legal ambiguity around the issues). If the courts or Congress moved toward the strict liability option described below, that might create enough pressure for new approaches to emerge. However, the technical and transactional challenges in crafting a solution under current rules are considerable.

To the extent that a broad scale endogenous market solution is unlikely to emerge,[26] it is useful to understand why. There are two potential explanations for this (which are not mutually exclusive). First, the market for these transactions may not possess the necessary structural conditions for producing

26 Although some narrowly focused licensing solutions have already emerged, it remains to be seen whether these test cases are long-run economically viable or whether they can be translated to a broader scale.

efficient outcomes. It may be, for example, that legal rights are not sufficiently clear, transaction costs are too high, or there are externalities not being priced into transactions.

Second, if the market were structurally capable of producing efficient outcomes but still failed to do so, it may be that the technology simply is not producing enough value to entice rightsholders to license their works. That is, there may be no license fee at which developers and rightsholders would both be willing to transact, even after correcting all market failures. This is a reasonable concern if the value of any particular work is truly trivial, as claimed by developers.

These two explanations have different policy implications. The first implies that policies should be crafted to fix the existing market failures, facilitating transactions and an efficient market outcome. This may entail, for example, clearly defining rights, establishing institutions to ease transactional frictions, or appropriately compensating for externalities. The second requires that we recognize that the failure of a market to form *is* the efficient outcome. If developers are producing insufficient value to compensate rightsholders in a way that compares to their next best alternatives, then forcing transactions will be welfare reducing. Thus, the first step for researchers along this front should be to evaluate the reasons why markets are not forming in more general ways.

7.2. Unrestricted Access to Copyrighted Materials

An approach at one extreme of the spectrum is to allow developers unrestricted access to copyrighted training materials without the need for permission or compensation (i.e., declaring all training to be fair use). This approach would satisfy the desire of developers for maximum flexibility at the expense of depriving historical rightsholders of rents that they may have reasonably anticipated receiving and potentially undermining long-held presumptions about how copyright operates.

Note that this approach need not eliminate all rights of existing copyright owners relative to generative AI — only the rights relating to the use of training data. For example, if someone used an LLM to create a novel in the style of Harry Potter, the training access granted to developers does not necessarily need to change J.K. Rowling's ability to assert rights to that output as a derivative work or an otherwise infringing copy of her work. However, writers of magical fiction in general could not claim any right to generic magical fiction produced by a model trained on their works.

This approach would have the lowest transaction costs among the various options, and it would maximize the speed of AI development (at least in the near-term). On the other side, this option may be the most detrimental to the incentives of human creators. To the extent that generative AI serves as a substitute for human creative output, anything that makes the development of or access to that technology less costly without proportionally increasing the benefits to human creators in some other way will diminish human creators' market power.

Another dynamic issue to consider is that developers are relying on ever-increasing volumes of training materials. Once demand for works to be used in training exceeds the supply of existing works, further development will be tied to the production of new works by human creators. In that way, diminishing the incentives of human creators may have negative dynamic effects on AI development in the long run.

There are several conditions under which unrestricted access to copyrighted materials could theoretically be the optimal option. First, the short-run social value of developing foundation models must be sufficiently high. Second, the harm inflicted on rightsholders, and the dynamic integrity of the copyright system, must be sufficiently low. Third, the transaction costs in alternative approaches must be sufficiently high to prevent the endogenous formation of reasonably efficient markets. The first condition likely cannot yet be evaluated given our current uncertain understanding of the technology's potential. However, the second and third conditions are testable.

7.3. Full Retrospective and Prospective Restrictions on Access for Training Purposes

An approach at the other end of the spectrum involves strictly enforcing restrictions on developers' access to copyrighted training materials, both for past and future use. This approach would hold developers liable for the past unauthorized use of copyrighted material in training existing models and require them to gain authorization for any such use moving forward. However, whether this approach is feasible obviously depends on our ability to identify works used in training, both retrospectively and prospectively. As previously discussed, this may not be possible.

Given the purported impossibility of tracing model inputs based on model outputs, any retrospective identification would require records about training data. It is possible that sufficient records do not exist, in which case this option is unlikely to function as intended. Prospective identification seems technically feasible but likely will come at a nontrivial cost. This aspect of the approach is not contingent on the existence of past records, but we would need to establish some sort of record keeping and verification requirements or create technical solutions that can prevent ingestion for rights to be enforceable.

Setting aside practicality issues, consider the theoretical implications of this approach. First, note the potential availability of statutory damages (whereby damages are due for infringement without any requirement to prove actual harm). These damages would apply to works with copyright registrations, which is likely a very small subset of the copyrighted materials ingested. Nonetheless, statutory damages apply on a per-work, per-infringement basis, and could add up to staggering sums. These damages would potentially bankrupt some current players and significantly increase development costs for those who could bear them.

This would create an incentive for developers to pursue some kind of settlement, presumably offering significant amounts of money to rightsholders. But the transaction costs would be high because they would in principle need agreement from many different rightsholders. It is unclear whether those rightsholders would agree to a settlement, and it would certainly create a lot of uncertainty while it was being worked out. Additionally, the prospective license requirement may also involve excessive transaction costs. As discussed in previous parts, these costs can be mitigated through some sort of collective intermediation, but it is likely that transaction costs will always be a significant source of inefficiency in markets for training data.

The conditions under which this approach may be the optimal option are the opposite of those discussed in the previous section. First, the social value of developing foundation models must be sufficiently low, given that this approach will slow or prevent development. Second, the harm inflicted by ingestion on rightsholders and the dynamic integrity of the copyright system must be sufficiently high. Third, the transaction costs must be sufficiently low to allow for the formation of reasonably efficient markets.

7.4. Grant Retrospective Conditional Amnesty and Enforce Access Restrictions Prospectively

The first two approaches represented opposing ends of the spectrum. There are many options between those extremes. One such option would be to grant amnesty for past use without permission up until a certain date but strictly enforce the right to control copying going into the future. This option takes into account the challenges with enforcing retrospective liability and the potentially existential implications it may have for current developers. It may also limit harm to the long-run integrity of the system and the incentives for continued human creation without forcing developers to completely abandon the progress so far made in generative AI development.

This approach is clearly not ideal for all developers or for all rightsowners. However, it is better for rightsholders than the open-access option, and it is better for developers than the strict retrospective and prospective enforcement option. Moreover, because the retrospective enforcement of rights may bankrupt many developers (who may be future licensees), at least some rightsholders would end up better off under this approach than under the strict liability approach. It has been suggested that realizing the full potential of existing models will require that they be trained on significant amounts of new data. If this is correct, developers' willingness to pay for the right to use new data could be significant.[27]

This approach reduces uncertainty and transaction costs, which is socially beneficial, but does it in a way that is prejudicial to rightsholders relative to historical expectations. One mechanism that could be employed to adjust for this may be to establish some sort of costly condition for amnesty that somehow benefits otherwise uncompensated rightsholders. For example, to gain retrospective amnesty, each model developer could be required to make some kind of one-time payment into a compensation fund, which would then be distributed in some way among historical rightsholders.

The consequences of such an approach would be very heterogeneous across different historical rightsholders, with active creators winning and inactive creators of older works losing. These differences could be mitigated, but probably not eliminated, if some kind of compensation was collected and distributed.

7.5. Create a New Statutory Blanket License

Another solution that has been suggested is that Congress create a new blanket license for training materials, for example, analogous to that administered by SoundExchange for the streaming of sound recordings. However, the conditions that economically justify compulsory licensing are rare.

27 Open-source here pertains to both volunteer communities of developers (e.g., the group that develops the Linux kernel) as well as academic contributors who are not in the business of developing practical AI systems, but instead testing, validating, and proposing new innovations that could shape future development.

As a preliminary matter, one should not conflate the advantages of compulsory licensing with the advantages of collective intermediation more generally. The primary advantage of collective intermediation is that it reduces transaction costs by centralizing negotiations. Compulsory licensing shares this advantage because it is one form of collective intermediation (other forms, as previously discussed, are collective rights organizations like those that exist in the music industry). However, that advantage arises from the centralized nature of negotiations, not from the compulsory nature of the regime. Therefore, the reduction of transaction costs is not, on its own, a sufficient justification for compulsory licensing, since the same effect can be achieved through other, less intrusive forms of collective intermediation.

Nonetheless, there are conditions that can justify compulsory licensing. One such justification is the existence of externalities with substantial value implications that would not otherwise be considered in licensing negotiations. For example, suppose generative AI models have some positive knock-on effect even for those not consuming the service. That social value may not be a factor for either developers or rightsholders when negotiating license terms, in which case licensing costs would be higher than the socially optimal level.

Compulsory licensing may also be warranted if other forms of collective intermediation result in excessive market power for rightsholders. This, in fact, was the stated justification in the early 1900s for today's compulsory licensing regime around mechanical reproduction rights for musical works (Peters, 2004). However, compulsory licensing may seem like an odd solution to issues of market power in intellectual property. Intellectual property rights are intentionally designed to confer market power to rightsholders. If there is an overprovision of market power, the more natural solution may be to recalibrate the scope and breadth of the intellectual property rights that are leading to that market power in the first place. However, when more efficient approaches are unavailable, compulsory licensing remains one solution to the issues of excessive market power.

Nevertheless, even if substantial externalities or issues of excessive market power do exist, those conditions on their own are also not sufficient to justify compulsory licensing. The reason that compulsory licensing is used sparingly is that such regimes can impose substantial economic inefficiencies, which have the ability to vastly exceed the inefficiencies they are intended to resolve. This is because compulsory licensing regimes require us to supplant the collective wisdom imbedded in market forces with the limited knowledge of some central planners. The central planner must decide on a licensing rate, but to arrive at the socially optimal rate, the planner must have a comprehensive and explicit understanding of the value of rights to all stakeholders (knowledge that would otherwise be implicitly factored in through market forces). Since such knowledge cannot be possessed in the absence of a market, central planners are forced to set licensing rates with incomplete information. Rates that are too high or rates that are too low will result in an inefficient allocation of resources and deadweight loss. Compulsory licensing is only economically defensible if the inefficiencies that it resolves sufficiently exceed the different set of inefficiencies it imposes.

Aside from the general efficiency issues with compulsory licensing, several other concerns arise in the specific context of training data. For example, previous statutory licensing systems have dealt with a well-defined type of work and mostly identifiable rightsholders. Foundation models ingest any and all types of works that can be expressed digitally. That adds the extra complication of having to equate contributions from very different types of works and different types of rights. Additionally, models ingest things they find on the internet. It will be difficult to identify rightsholders for much of that content, with many not even knowing they are rightsholders. This sort of ambiguity would make the administration of such a system complex and costly. It is possible that administrative costs could account for the vast majority of revenues collected by the system. Relatedly, it is likely that most works ingested would get trivial amounts of money given the volume of training sets. However, if the overall blanket license fee is set to collect significant revenue and it can be apportioned to rightsholders in a way that is somehow proportional to the value of their work, popular and successful authors could get a significant new revenue stream.

Putting aside, for now, the efficiency concerns with compulsory licensing, implementing such an approach requires the determination of (1) a formula for the royalty owed by model trainers and (2) a formula for the distribution of royalties to different rightsholders. One candidate structure for both sides of the problem is to use revenue as the scaling variable.

With respect to the setting of developers' obligations, basing the royalty on revenue earned has the consequence that developers would only owe royalties once they start selling or licensing some kind of product. Whether this is a feature or a bug depends on one's point of view. Start-up firms do not normally have the luxury of paying for their inputs *only* after they start earning revenue. However, because this particular input is intangible and not consumed by use, there may be an efficiency argument for allowing developers to escape royalty obligation if they never commercialize.

Because the contribution of particular works to the parameters of the models is hard (or impossible) to determine, there does not seem to be any obviously "correct" method for determining the share of royalties that should go to each rightsholder. However, this means that a simple approach based on widely available information will likely be as good as any effort to target the money more precisely. The revenues associated with a work could serve as a rough proxy for a work's relative value to developers. As a general tendency, works will have more impact on a model the more frequently they appear in the training corpus (Liu et al., 2024), and the more distinctive is their use. Sales and subscription revenue seem a reasonable, if imperfect, proxy for these factors, as the per-unit price captures something about distinctiveness and the number of sales proxies for frequency of use.

The need to calculate royalties owed and factors of allocation are the same requirements that arise in all statutory licensing systems and in any sort of collective intermediation generally. Notably, no existing system has a perfect mechanism for determining these things. We are always forced to rely on rough approximations and proxies for things that we cannot directly measure (e.g., using the number of streams to measure a track's marginal contribution to subscriber revenue). Thus, the threshold for whether a statutory licensing system for training data is sensible is not whether the system can assess and allocate royalties in a way that perfectly reflects value and marginal contributions. Rather, the threshold should be that the inefficiencies that come from misallocations and the need to administer a complex system are smaller than the inefficiencies that come from any other alternative.

This approach also brings about questions of opt-in/opt-out capabilities. The efficiency of a statutory license system is gained by users knowing with certainty that they have the right to use all works so long as the license fee is paid. Allowing for non-universal coverage by a statutory license diminishes some of that efficiency because users are still left to determine which works they can or cannot use. Thus, whether having optional participation from rightsholders within the statutory licensing system is sensible depends on how easily users can distinguish between participating rightsholders and non-participating rightsholders. In the present case, this is a technological question. If works can be programmatically identified as participating or not participating, for example, by some tag in a webpage's source code, then allowing optional participation may not diminish efficiency by much and may in fact be welfare improving to the extent that it accounts for the heterogeneous preferences of rightsholders.

However, if the system is optional, we will have to make potentially consequential decisions about what the default should be. That is, should all rightsholders start as participants and be given the opportunity to opt out, or should they start as non-participants and be given the opportunity to opt in? If there were no transaction costs, there would be no difference in terms of efficiency between the two options. However, there will be transaction costs, and those costs will often be much greater than the value at stake. For example, for many rightsholders, the payout for participating will be very close to zero. In those cases, taking even five minutes to opt in would impose more cost on the rightsholder than they stand to gain. In all likelihood, the vast majority of rightsholders (many of whom do not even know they are rightsholders) would not bother to opt in, even if they are indifferent to their work being ingested. Likewise, those same rightsholders would be unlikely to opt out if the default was to include them.

Transaction costs play a central role in determining the most efficient system in terms of opting in or out. Consider three groups of rightsholders. First, we have rightsholders who would be willing to sell access to their works at a price, net of transaction costs, that developers are willing to pay. This group is effectively opting in regardless of the default. They do nothing in the opt-out case, and they choose to opt in in the opt-in case. The only difference between the two cases is whether rightsholders incur the transaction cost of taking action (opting in).

Second, we have rightsholders who value restricting access to developers more than developers value gaining access. In contrast to the first group, this group will always be opted out, regardless of the default. They will do nothing in the opt-in case, and they will opt out in the opt-out case. Again, the only difference is whether rightsholders incur the transaction cost of taking action (opting out). Considering only these two groups, which is the efficient system, depends on which group is larger (assuming the transaction cost of opting in is, on average, the same as that of opting out). In principle, we want to set the default as to minimize the number of people who have to take action.

The third group of rightsholders includes those for whom the potential value at stake is less than the cost of taking action. This includes those who are entirely indifferent to their work being ingested, those who have a slight preference against ingestion at the given royalty rate, and those who have a slight preference for ingestion at the given royalty rate. In these cases, the value of access to developers may be positive, but not enough to entice these rightsholders to opt in. Similarly, the net value to rightsholders of allowing access may be negative, but not enough to entice them to opt out.

Considering only this group, the most efficient system will be the one that allocates rights most often to the party that would have received them absent any transactions costs. If, for example, these rightsholders have no interest in restricting access to their works and developers have some positive value for accessing those works, in the absence of transaction costs, the rights would end up with developers. Thus, in this scenario, the efficient system would be an opt-out system that gives developers default access. If, in contrast, rightsholders in this group value restricting access more than developers value gaining access (but still less than the transaction cost of opting out), then the efficient system would be an opt-in system.

Ultimately, determining the overall most efficient system depends on the relative size of the three groups and the value distribution within the third group. The third group has the potential to be the most consequential because it is the only group bound by the default. Ensuring low transaction costs for rightsholders when opting in or opting out will reduce the size of this group and thus reduce the difference between an opt-in system and an opt-out system in terms of efficiency. Nonetheless, determining the degree to which rightsholders are, more or less, indifferent to their works being ingested would be a helpful empirical exercise.

7.6. How to Evaluate Different Options

Choosing among the options is challenging because their effects are potentially difficult to quantify *ex ante*, and there are multiple, somewhat incommensurable objectives at stake. However, here, we propose a set of criteria to consider when assessing options:

- Maintaining the integrity of the copyright system. Copyright is a dynamic system. We give copyright protection to current creators to signal to future creators that they too will receive such benefits, thus incentivizing them to produce creative works. When there are systematic differences between the benefits creators expect to receive and what they actually receive, the system will deteriorate, and the signal we intend to send to future creators will lose some degree of credibility. The perception of widespread infringement will, therefore, diminish the efficacy of copyright as an incentive.

- Maintaining incentives for human creation. Credibility issues aside, how we treat AI ingestion will have implications for the level of incentives faced by human creators. The new level of incentives will need to continue to strike the appropriate balance between the production of new works and access to existing works.

- Mitigating the costs associated with legal ambiguity. Inefficiencies arise when legal rights and the consequences of violating them are uncertain. First, the uncertainty causes stakeholders to be reluctant to act in ways that may otherwise be productive, thus slowing the innovative process. Second, when rights are not clearly and universally understood, we turn toward expensive adjudication processes to resolve disputes. The more certainty there is in a particular area, the more predictable the outcome of adjudication will be, thus making it less necessary. When making policy decisions, if the candidate policies would produce sufficiently similar outcomes, the cost of indecision is high, and the quality of relevant information is unlikely to significantly improve in the near-term, then swiftly choosing one, even with a fair chance of choosing the less-than-optimal option, will be more efficient than prolonging the decision. In contrast, if policy outcomes are likely to be highly consequential, the cost of legal ambiguity is low, and relevant information is likely to improve in the near term, prolonging the decision-making process may be sensible.

- Distinguishing value creation from value transfer. Some portion of the value perceived to come from generative AI is in fact value that developers have newly created. However, some portion also represents a mere transfer of value from other stakeholders, such as authors whose works are used for training. Policies toward AI ingestion need to be compared on the basis of the former and not the latter. Value creation is welfare improving; value transfer is not. Thus, a policy that results in high-value models is not necessarily the optimal policy if much of that value is a mere transfer from other sources that would have still existed in the models' absence.

- Cost of government oversight and implementation. For some policy options, administrative costs have the potential to be very large, possibly even outweighing the value at stake. Within the spectrum of possible policies toward access to training data, unrestricted access is the least costly in terms of oversight and administration. A statutory license system is likely to be the costliest to administer. Any system reliant on private licensing may not face the same administrative costs but may indirectly impose costs insofar as it increases the burden on the court system, which would be tasked with sorting out disputes. When comparing policies toward ingestion, it is important to ensure that the cost of a solution does not outweigh the cost of the problem it is intended to solve.

- Transaction costs in the private sector. Transaction costs should play a central role in policy comparisons. These costs are likely the largest barrier to the endogenous formation of licensing markets. The option that presents the lowest transaction costs is unrestricted access, followed by a statutory licensing system. Systems that require private negotiation will have the highest transaction costs. However, as previously discussed, some of that can be mitigated through collective intermediation.

Some of these factors clearly favor certain policy solutions over others, but there is no single solution that is superior in all factors. Absent empirical evidence on the full welfare effects of each of these factors, decision-makers will likely need to weigh each in a heuristic or subjective way. However, quantifying each factor in the context of different policy options will allow for better-informed decisions.

References

Liu, Yiheng, Hao He, Tianle Han, Xu Zhang, Mengyuan Liu, Jiaming Tian, Yutong Zhang, Jiaqi Wang, Xiaohui Gao, Tianyang Zhong, Yi Pan, Shaochen Xu, Zihao Wu, Zhengliang Liu, Xin Zhang, Shu Zhang, Xintao Hu, Tuo Zhang, Ning Qiang, Tianming Liu, and Bao Ge. 2024. "Understanding LLMs: A Comprehensive Overview from Training to Inference." *arXiv:2401.02038v2*. https://arxiv.org/pdf/2401.02038

Peters, Marybeth. "Section 115 Compulsory Licenses." Statement before the Subcommittee on Courts, the Internet and Intellectual Property, House Judiciary Committee. March 11, 2004. https://www.copyright.gov/docs/regstat031104.html

The Authors Guild. "More than 15,000 Authors Sign Authors Guild Letter Calling on AI Industry Leaders to Protect Writers." *The Authors Guild*, last modified June 20, 2023. https://authorsguild.org/news/thousands-sign-authors-guild-letter-calling-on-ai-industry-leaders-to-protect-writers/

8. Potential Socioeconomic Biases of AI Policy

Principal contributor: Catherine Tucker

The discussions so far have focused mainly on the direct links between copyright policy toward AI and the specific policy objectives of copyright. That is, it has focused on the *intended* consequences of AI-relevant copyright policy. Nonetheless, all policy decisions in this realm (explicit or implicit) will almost certainly be followed by a complex network of *unintended* outcomes, some of which may be consequential for social welfare. Although there will likely be countless unintended and, in some cases, unforeseeable consequences of copyright policy toward training materials, this part focuses on the socioeconomic biases that may arise from those policies.

One source of bias arises because the output of generative AI will depend on the nature of the data that is used to train the model. That supply of data (creative works) is going to be subject to certain distortions that will flow through to the output of such models. These distortions will reflect the ease and cost of accessing any particular work. In general, we expect these distortions to come from four sources: (1) variations in the financial resources of creators; (2) variations in the works that are, and are not, subject to copyright protection; (3) variations in digitization; and (4) variations in the scale of available data.

With respect to variations in the financial resource of creators, certain biases can be exacerbated by giving developers unrestricted access. In Web 1.0, we worried about lower direct rewards for artists dulling incentives. In Web 2.0, we found out that if there are indirect rewards (such as YouTube fame), this can compensate. Now we have a world where indirect rewards are also challenged because human creators will ultimately become more anonymous, at least insofar as they contribute to the synthetic output of generative AI.

It is uncertain what compensating behavior may arise to mitigate this issue. However, it could mean that art becomes the domain of those who have completely non-pecuniary incentives. Some may laud this change as stripping away commercial concerns and leaving the production of art for the sake of art. Nonetheless, it is important to remember that commercialization is not just a matter of pecuniary incentives; it is also a matter of providing creators with the means by which to practice their art and produce their works. All the artistic motive in the world does not change the fact that a creator must have food, shelter, and the tools of their craft to produce anything.

If the means to acquire those can no longer be captured from the production of creative works, it must be acquired from another source. For those of lesser resources, this means seeking employment unrelated to the production of their art, which not only reduces the volume of their creative output but also reduces the time available for improving one's skills, thus diminishing the quality of their output. Putting aside the quality-diminishing effects that occur when artists are forced to shift from professionals to hobbyists, this will ultimately influence who is able to produce art. In particular, it means that creative production will shift to those who have other sources of income or preestablished wealth.

Taken together, these issues imply that widespread generative AI use could produce a future with a more homogeneous idea of what art and culture are, which unconsciously reflects the degree of privilege among contributors to the training materials. This impulse weakens the ability of copyright to promote originality. Ultimately, policy decisions around access to training materials will shape how the technology develops and is used. The social effects of the technology will be unevenly felt amongst socioeconomic groups. While these distortions may be unavoidable in certain contexts, they can probably be mitigated through careful policy intervention.

With respect to variations in copyright protection, developers will likely respond in predictable ways to the risks in front of them. One of those risks, depending on policy outcomes, may be copyright infringement based on the use of certain training materials. If a developer wants to avoid potential litigation risk, they may focus, for example, on creative works already clearly in the public domain. However, this may introduce certain distortions. Whether a work is in the public domain is a complex legal question depending on factors that are not always readily known to a potential user. Nonetheless, it is clear that in 2025, all published works in the United States from 1929 and before are in the public domain (U.S. Copyright Office, 2024). Therefore, as a risk mitigation tactic, it might make sense for a developer to train a model on works published prior to 1929, which will reflect the societal norms of the time and not necessarily be representative of modern norms or of the current population.

For example, in the realm of literary works, the ability to be published was, historically, often a matter of resources and access to leisure. Published authors tended to come mainly from the wealthier classes of society. While the publication of works still portrays certain socioeconomic biases, those biases were orders of magnitude larger 100 years ago when publication was the near-exclusive domain of white males born to a privileged class. While exceptions certainly exist, they were relatively rare and special cases. For example, it has been argued that Jane Austen would never have been published if her father had not formed a contract with her publisher to obtain the copyright or if she had been married (House, 2023).

Therefore, it seems likely that we may end up with generative AI models that reflect the degree of privilege that was needed to publish works prior to 1929. This is, in a sense, a simple case of sample selection bias. Certain groups and views will be underrepresented in the 100-plus-year-old training sample and thus be underrepresented in the output. As the prevalence of generative AI content grows, the degree of underrepresentation will be amplified.

While it seems unlikely that developers literally train their models on *only* pre-1929 works, they will take into account the variable infringement risk associated with the works they consider for training. Holding all else constant, developers will elect to use lower risk works when possible. Infringement risk decreases with the age of a work; thus, when training activities are subject to copyright restrictions, the average age of works ingested will be higher than the alternative.

A related issue is that of open access licenses, such as those facilitated by Creative Commons. These licenses signal to developers (or any other users) that infringement risk is very low, thus making such works an attractive option when developers are subject to infringement liability. The body of open access works likely does not suffer from the age distortion of public domain works (at least not to the same degree), but there may be other implicit biases in who elects to offer an open-access license. In particular, creators of such works are people who may not need to directly appropriate the value of their works. In some cases, open access may be a path for indirect appropriation (e.g., providing notoriety to facilitate the commercialization of future works), but in many cases, it also reflects differences in creative motives and resources. Creators who can support themselves through other means while producing creative works likely exhibit systematic socioeconomic differences compared to creators who must rely on the commercialization of their works to support themselves.

With respect to the digitization of works, developers will clearly be drawn to works that already exist in digital form over those they themselves would have to physically access and then digitize. In modern times, nearly all new media and information is digital and thus potentially accessible as training materials. This is, however, not universally true, and the areas where it is not are likely to cause some degree of bias in output. For example, much of the cultural heritage of certain indigenous peoples exists exclusively as oral tradition. In those cases, generative AI tasked with producing something related to such cultures would lack key perspectives. As reliance on AI increases, the absence of such perspectives will erode the preservation of cultural heritage and diminish the influence certain cultures have on society.

While this would be a relatively narrow (but significant) issue if foundation models were trained exclusively on modern materials, if they begin to disproportionately train on public domain materials, the issue is greatly exacerbated. This is because digitization of earlier works reflects a series of decisions about how to direct resources. For example, projects where cultural bodies of work have been digitized tend to reflect the type of works that exhibit traditionally valued artistic characteristics.

A large number of works that are not in English have not been digitized (Nicholas and Bhatia, 2023). Similarly, a large number of works that are not from the United States have not been digitized (Kizhner, 2021). Generally, for works that are not natively digital, the decision to make a digital copy depends on several factors that include the characteristics of the potential audience. We expect works that appeal to large or well-resourced audiences will be digitized disproportionately more often than works that appeal to a smaller or under-resourced audiences.

There have, however, been attempts to correct digitization biases. For example, the Digitization Center at Munich has more than 10,000 digitized incunabula and nearly 300,000 digitized printed books from the sixteenth and seventeenth centuries. These are written in German, Latin, and the Romance languages. However, this effort still reflects the choices of those running the State Library about which books should be part of its collection. As a result, the collection itself reflects the social norms and tastes of those who were arbiters of taste in the past. Those tastes may not now look particularly inclusive. Also, such efforts do not change the fact that many languages are either newer (such as post-1910s Turkish or Greek) or come from countries that use non-Roman alphabets. This means there are fewer digitized corpora from these perspectives that models can be trained on.

With respect to variations in the scale of available data, developers are likely to be more attracted to large bodies of data with highly concentrated rights ownership. When developers are subject to infringement liability for training activities, they will need to gain licenses to use works outside of the public domain. License negotiations come with transaction costs that increase with the number of distinct rightsholders involved. For example, a developer may prefer to negotiate with the owner of a large online forum rather than negotiate separately with a diverse set of literary authors and their publishers.

Indeed, Reddit, as an example, has announced a deal to license its users' postings to developers. However, such forums are poor representations of society at large. Certain types of privilege give some people more time to generate content on forums like Reddit, and there are systematic socioeconomic differences between the people with such privilege and those without. Reddit is only used by 11 percent of U.S. adults (Pew Research Center, 2019). Of those, around two-thirds are men. Ultimately, Reddit is not an accurate representation of society. However, such data sets, because they are easy to access and license, are likely to become frequent sources of training data. Developers' dependence on them will increase with tighter restrictions on access to training materials.

References

House, Christian. "Jane Austen's Life-Changing Legacy." *Sotheby's*, November 21, 2023. https://www.sothebys.com/en/articles/jane-austens-life-changing-legacy

Kizhner, Inna, Melissa Terras, Maxim Rumyantsev, Valentina Khokhlova, Elisaveta Demeshkova, Ivan Rudov, and Julia Afanasieva. "Digital Cultural Colonialism: Measuring Bias in Aggregated Digitized Content Held in Google Arts and Culture." *Digital Scholarship in the Humanities* 36, no. 3 (September 2021): 607–640. https://doi.org/10.1093/llc/fqaa055

Nicholas, G., and A. Bhatia. 2023. "Lost in Translation: Large Language Models in Non-English Content Analysis." *Center for Democracy & Technology*. https://arxiv.org/pdf/2306.07377

Pew Research Center. "Share of U.S. Adults Using Social Media, Including Facebook, Is Mostly Unchanged Since 2018." *Pew Research Center*, April 10, 2019. https://www.pewresearch.org/short-reads/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

U.S. Copyright Office. "Lifecycle of Copyright: 1928 Works in the Public Domain." *Copyright: Creativity at Work*, January 8, 2024. https://blogs.loc.gov/copyright/2024/01/lifecycle-of-copyright-1928-works-in-the-public-domain/

UNITED STATES COPYRIGHT OFFICE
101 INDEPENDENCE AVE. S.E.
WASHINGTON, D.C. 20559-6000

COPYRIGHT.GOV

PUBLISHED FEBRUARY 2025 / M-363