NBER WORKING PAPER SERIES

COPYRIGHT POLICY OPTIONS FOR GENERATIVE ARTIFICIAL INTELLIGENCE

Joshua S. Gans

Working Paper 32106
http://www.nber.org/papers/w32106

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2024

Joshua Gans has drawn on the findings of his research for both compensated speaking engagements and consulting engagements. He has written the books Prediction Machines, Power & Prediction, and Innovation + Equality on the economics of AI for which he receives royalties. He is also chief economist of the Creative Destruction Lab, a University of Toronto-based program that helps seed stage companies, from which he receives compensation. He conducts consulting on anti-trust and intellectual property matters with an association with Charles River Associates and his ownership of Core Economic Research Ltd. He also has equity and advisory relationships with a number of startup firms. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

The author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w32106

Copyright Policy Options for Generative Artificial Intelligence
Joshua S. Gans
NBER Working Paper No. 32106
February 2024
JEL No. K20,O34

ABSTRACT

New generative artificial intelligence (AI) models, including large language models and image generators, have created new challenges for copyright policy as such models may be trained on data that includes copy-protected content. This paper examines this issue from an economics perspective and analyses how different copyright regimes for generative AI will impact the quality of content generated as well as the quality of AI training. A key factor is whether generative AI models are small (with content providers capable of negotiations with AI providers) or large (where negotiations are prohibitive). For small AI models, it is found that giving original content providers copyright protection leads to superior social welfare outcomes compared to having no copyright protection. For large AI models, this comparison is ambiguous and depends on the level of potential harm to original content providers and the importance of content for AI training quality. However, it is demonstrated that an ex-post `fair use' type mechanism can lead to higher expected social welfare than traditional copyright regimes.

Joshua S. Gans
Rotman School of Management
University of Toronto
105 St. George Street
Toronto ON M5S 3E6
and NBER
joshua.gans@rotman.utoronto.ca

1 Introduction

In the last few years, a new wave of artificial intelligence (AI) applications called "generative AI" have become useful and popular. These AI models are noted for their ability to generate text, images and videos from text prompts. Generative models are machine learning models (specifically, transformer-based deep neural networks) that are trained on data to learn key patterns and relationships and generate outputs with similar characteristics. The common applications involve users inputting prompts in a natural language to generate outputs. These include text outputs from large language models (LLMs), including OpenAI's ChatGPT, Google's Bard, Microsoft's Copilot and Meta's LLaMA and also image outputs from Open AI's DALL-E, Stable Diffusion and Midjourney.

These new generative AI applications have raised copyright concerns from a number of original content providers. Specifically, these concerns regard the use of copy-protected content in training data. A key question is whether the use of such copy-protected material in training is covered by fair use provisions in copyright laws or must be licensed with permission from copyright owners. AI providers argue that such licensing would be prohibitively costly especially considering the transactions costs that may be involved. Content providers argue that without such protection, they will be inadequately rewarded for their content creation.

Another aspect of generative AI models has complicated these copyright issues. In a lawsuit in 2023, the *New York Times* alleged that OpenAI had used its copy-protected content without permission in training its GPT LLMs. It has asked the Court for measures to prevent the availability of models trained with its content and/or statutory damages for harm caused. The evidence from the *New York Times* for this was a demonstration that with certain prompting, both ChatGPT and BingChat (which licenses OpenAI's GPT) could reproduce articles almost verbatim from the *New York Times*. Similar prompting in image generation models can produce likenesses of commercially owned characters and digital assets (Marcus and Southen, 2024). OpenAI responded by arguing that they had not knowingly trained their models on *Times'* content but that, instead, the examples were evidence of "regurgitation."[1] This is a situation where, because certain text is available on public sites, the large AI model can statistically reproduce that text (Tänzer et al., 2021). In other words, the examples were something different than pure copying.[2] Nonetheless, if AI models could be used to reproduce original content, this 'leakage' could have effects on the

[1]OpenAI wrote: "Memorization is a rare failure of the learning process that we are continually making progress on, but it's more common when particular content appears more than once in training data, like if pieces of it appear on lots of different public websites." `https://openai.com/blog/openai-and-journalism`

[2]At the time of writing, it is not known with certainty whether the *Times*' examples were regurgitation or not.

commercial interests of original content holders if users of AI models choose not to purchase from original content providers and, instead, substitute towards the content reproduced by AI models. Copyright law recognises that certain otherwise fair use of copy-protected material may not be covered if there is identifiable and significant harm to original content creators.

A natural set of research questions arises from this context: how do copyright regimes impact the incentives of original content providers to invest in quality and the ability of AI providers to train their models? In addition, what are the expected social welfare outcomes from different copyright regimes? This paper develops an economic model based on approaches to developing generative AI to answer these questions.[3] Interestingly, the model developed here for generative AI has a 'human' analogue that is typically seen as not being covered by copyright regulation. Here, the examination, while based on generative AI, can be informative as to why these human activities are not seen as being subject to similar regulation.

Broadly speaking, there are two main approaches to training generative AI models that have implications as to how copyright regimes might operate. As will be argued here, they are different precisely as to how the relationship between content used for training and the output of AI models can be established. Legal scholars term this "provenance." First, there are what I will call "small" AI models. These models are clearly trained on a set of identifiable content, and the outputs generated can be seen as being derived from that content. Presently, some of these models utilise retrieval-augmented generation (RAG) that provides content for the AI model that is not necessarily in its training data or, more generally, becomes a key part of its training data. For instance, a researcher may upload a paper and develop an AI application built on a larger foundational model that is prompted to only generate output from that paper.

The issue with small AI models arises when the content utilised in, say, a RAG model, is owned by someone other than the person developing or operating the AI model. Here are two scenarios based on where, for each, an AI version and a human version are provided:

Scenario 1a Spoilers (AI Version): *An AI provider creates a chatbot that allows users to ask questions about specific television series, including plot details, characters and key quotes from the series. The chatbot is trained on transcripts and other data from the television series.*

[3] As such, it asks questions that economists have been asking regarding copyright since Landes and Posner (1989).

Scenario 1b Spoilers (Human Version): *A person answers questions about a specific television series over social media, including plot details, characters and key quotes from the series. The person has watched the television series multiple times.*

Scenario 2a TL;DR (AI Version): *A developer builds and sells access to a website that provides summaries of business books. The summaries were generated by AI and trained on scanned text from books the developer purchased.*

Scenario 2b TL;DR (Human Version): *A person*[4] *sells summaries of business books. These summaries were written by the person after the person had purchased and read a copy of each book.*

In each case, the human version is not subject to copyright regulation. The question is currently open for the AI versions. But each is isomorphic to one another. While it is possible to then draw a connection and claim "if humans are free from regulation, why should AI providers be subject to it?" more appropriately, it would be worthwhile to understand what conditions imply that both the human or AI versions as the case may be could be subject or not subject to copyright regulation.

Section 2 develops a model to examine this question. The model involves a single original content provider and a single AI provider, each of whom makes investments in content quality and training quality, respectively. After the content provider has made their investments, because the provenance of the content is clear, the original content provider and the AI provider engage in a negotiation of whether the content can be used to train the AI or not. Allowing such use lowers the cost of AI training while also potentially resulting in leakage that causes commercial harm to the content provider. Section 3 then applies that model to various copyright regimes and shows that these impact the negotiations between the two parties. Critically, however, it is demonstrated that copyright protection results in greater social welfare than no copyright. The reason is that the original content provider is investing prior to any negotiations, and therefore, their incentives matter. In addition, those negotiations, regardless of the regime, result in pricing outcomes for both the original content and the AI that are not impacted beyond the different quality investments. Thus, the copyright regime does not change the overall consumption of original content. As copyright protection creates the strongest incentives for the original content provider to invest, it leads to higher expected social welfare. Moreover, in this context, it creates incentives for the content provider beyond their commercial interests to improve quality, which may

[4] Like Cliff of the Notes fame.

additionally lower AI training costs. Critically, this implies that the reason the human versions of the above scenarios should not be subject to copyright protection lies in factors such as whether such activities actually cause commercial harm to the copyright holder or not.

The second class of AI models are "large" models, which are the type of models that motivated the research in this paper. These models are trained on such a large volume of data that it is not possible to establish provenance ahead of time because, for the general operation of these models, no individual piece of content or even collection of content is significant enough to impact the vast majority of output from these AI models. This is not to say that the provision and availability of such AI models may not end up having a commercially harmful impact on a content provider. It is just that it is not clear in advance whether that might occur, nor is it clear what precise value specific content has in reducing AI training costs, even if the content used in training *as a whole* has an identifiable benefit.

Once again, how content is used in training large models has been discussed considering human analogues. This is from OpenAI's response to the *New York Times*' lawsuit:

> Just as humans obtain a broad education to learn how to solve new problems, we want our AI models to observe the range of the world's information, including from every language, culture, and industry. Because models learn from the enormous aggregate of human knowledge, any one sector—including news—is a tiny slice of overall training data, and any single data source—including The New York Times—is not significant for the model's intended learning.[5]

Here is a scenario that captures the connection:

Scenario 3a Teaching (AI Version): *A developer builds a chatbot that teaches students economics based on a corpus of textbooks.*

Scenario 3b Teaching (Human Version): *A person teaches students economics based on their reading of a corpus of textbooks.*

Note that in teaching it is generally agreed upon that human teachers are not subject to copyright prohibitions, but it is precisely the issue with AI that remains an open copyright regulatory question. Once again, it is important to use the model to understand precisely why the human version involves a particular conclusion to see if it can assist in understanding the approach for the AI version.

[5] `https://openai.com/blog/openai-and-journalism`

Analysing large AI models requires some adjustment to the baseline model of Section 2. First, the lack of provenance means that there is no opportunity for negotiation for the use of particular content, nor is it possible to know what the precise commercial harm (if any) to a content provider might be. Thus, the baseline model is amended to include uncertainty over potential commercial harm prior to setting prices for original content and AI use and to remove the ability to negotiate. Nonetheless, a copyright regime could prohibit (say, subject to large statutory infringement damages) the use of such content or permit its use. The model of Section 4 examines the expected social welfare outcomes from each of these traditional rights regimes and demonstrates that having no copyright protection is superior to copyright protection if the value of the content as a whole in lowering AI training costs is proportionately high relative to the expected commercial harm across all content providers.

The large AI model context also creates an opportunity for a different type of rights regime. Section 4 proposes and analyses a mechanism that is an ex post 'fair use' regime. Rather than intended to operate prior to potential infringement taking place, this mechanism operates ex post. This mechanism allows an AI model to be trained on all available content, and then, once original content owners observe leakage and realise commercial harm, exercising an option to sue the AI provider for lost profits. The regime sets a threshold on the magnitude of such harm, and those content providers who experience harm above that threshold exercise the option for lost profits while others do not. It is demonstrated that this mechanism offers stronger incentives for content providers than no copyright protection but also stronger incentives for AI training than traditional rights regimes. Moreover, subject to the financial feasibility of the AI provider, the threshold can be set low, implying that the mechanism will generate the highest expected social welfare amongst rights regimes.

There is no similar analysis in the literature of the impact of copyright regimes in a generative AI context. The closest model is that of Gans (2015) that involves some of the elements of the small AI model, notably a negotiation stage, but that involves all parties making investments prior to negotiation. Moreover, in that model, users cannot generate new outputs except by making use of original content. The model here allows AI models to be generated without such content for training. Gans (2015) also considers 'remix rights', which is a copyright regime proposed when users remix original content and produce transformed outputs. That mechanism operates ex-post like the one considered in Section 4 here, but it is analysed in a negotiation context, whereas here, the mechanism is proposed to deal with potential transaction costs arising in a large model context where it is potentially the optimal mechanism.

2 Model setup

Suppose that an original content creator (OC) can generate content of quality, x, at a cost of $c_{OC}(x)$ where $c_{OC}(.)$ is an increasing and convex function with $C(0) = 0$. There is a $[0, 1]$ mass of consumers with willingness to pay of θ, that are independently and identically distributed according to a uniform distribution. Willingness to pay types are private. If the consumer consumes the content, their utility is $x\theta$.

There is also an AI provider (AI) who can generate general-purpose AI products of quality, y, at a cost of $c_{AI}(y, sx)$ where $c_{AI}(.)$ is decreasing and convex in y and increasing in sx where sx is the use of a sample, $s \in [0, 1]$ of original content and x is the quality of that content. It is assumed that $c_{AI}(y, sx)$ is submodular in (y, xs).[6] Consumers have a common willingness to pay for the AI products of $u(y)$. In other words, their intrinsic demand for AI products is independent of their willingness to pay for original content.

Supplying AI products generates a by-product in the form of 'leaked' (or imitative) original content whenever $s > 0$. It is assumed that a consumer who has purchased an AI product can produce and consume original content with probability, $\rho(s)$, without purchasing it from the content provider. $\rho(s)$ is non-decreasing in s. Therefore, a consumer of type θ's expected utility if they purchase both the OC and AI products is $x\theta - p_{OC} + u(y) - p_{AI}$ where p_{OC} and p_{AI} are the prices set by OC and AI, respectively. However, if a consumer only purchases AI, then their expected utility is $u(y) + \rho(s)x\theta - p_{AI}$. Therefore, a consumer who has purchased AI has a willingness to pay for OC of $(1-\rho(s))x\theta$. Thus, leakage reduces demand for original content from OC.

Below, in order to understand the trade-offs and outcomes more precisely, we will, where appropriate, rely on the following functional forms: $c_{OG} = \frac{1}{2}x^2$, $c_{AI} = \frac{1}{2(1+\gamma sx)}y^2$, $u(y) = y$ and

$$\rho(s) = \begin{cases} s & s \leq \bar{s} \\ \bar{s} & s > \bar{s} \end{cases}$$

where $\bar{s} < 1$. The purpose of this functional form on $\rho(s)$ is to create a bound on the amount of competitive harm that can be brought on OC if the content shared is beyond a certain level. It also allows the use of $\bar{s}$ as a parameter of interest.

[6]That is, $\frac{\partial c_{AI}(y,sx)}{\partial y}$ is non-increasing in sx for all (y, sx).

2.1 Benchmark Outcome

Given this setup, it is useful to define the benchmark outcome that will maximise social welfare. The planner's problem is:

$$\max_{x,y,s} \int_0^1 \theta x \, d\theta - c_{OC}(x) + u(y) - c_{AI}(y, sx)$$

which gives first-order conditions:

$$\frac{1}{2} - s^* \frac{\partial c_{AI}(y^*, s^* x^*)}{\partial x} \leq \frac{\partial c_{OC}(x^*)}{\partial x}$$

$$\frac{\partial u(y^*)}{\partial y} \leq \frac{\partial c_{AI}(y^*, s^* x^*)}{\partial y}$$

$$x^* \frac{\partial c_{AI}(y^*, s^* x^*)}{\partial s} \geq 0$$

Note that the final first-order condition implies that $s^* = 1$; that is, it is socially optimal for all original content to be available for use in training the AI. Moreover, given this, alongside the consumption benefit of original content, the returns on investment in x include their impact in reducing the cost of generating AI products. For future reference, not that, using the specific functional forms, $x^* = \frac{1}{4}(2 + \gamma)$ and $y^* = \frac{1}{8}(4 + \gamma(2 + \gamma))$.

2.2 Timing

The timing of the game (depicted in Figure 1) is as follows:

1. (*Rights Regime*) A rights regime (r) is chosen by the planner.
2. (*Content Investment*) OC chooses x.
3. (*Negotiations*) s and a lump-sum payment, τ, is negotiated by OC and AI according to the Nash bargaining solution.
4. (*AI Training*) AI chooses y.
5. (*Pricing*) OC and AI simultaneously choose p_{OC} and p_{AI}, respectively.
6. (*Payoffs*) Consumer purchases are made, and profits are realised.

Below, several rights regimes, r, will be considered, including no copyright (NC), full copyright (CP) and forms of fair use (FU). In each case, the rights regime defines the rights of each party to use or prohibit the use of content should negotiations break down.

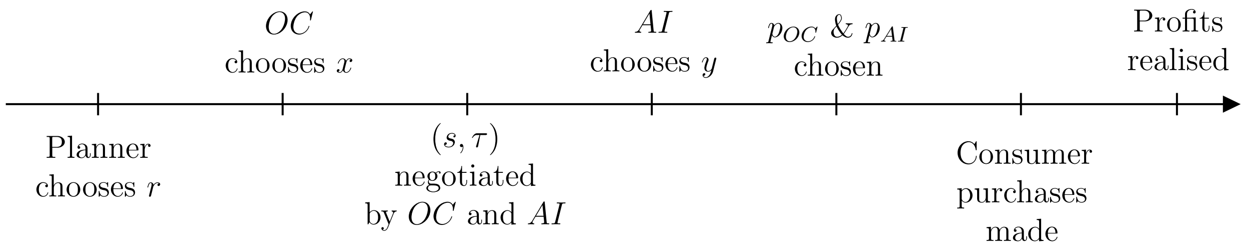

Figure 1: **Model Timeline (with Negotiations)**

The negotiation stage is critical for the types of realised social welfare outcomes. If a negotiated level of s is not achieved, then the choice of s occurs according to the rights regime (e.g., if the original content is copy-protected, OC can choose s unilaterally). One difference between the present model and that paper is that the other party (in that paper a user) makes an investment prior to negotiations whereas here the other party (an AI provider) makes its investment after negotiations. The idea is that copyright regimes may prevent the use of the material in AI training, and thus if there is a negotiation, it makes sense for it to occur prior to that training. This means that in the present paper, at least for the small AI model, the only agent subject to hold-up is OC.[7] In Section 4, a large AI model case is presented where such negotiations are not possible and, thus, a hold-up problem exists for both parties.

Here, this game is analysed backwards, starting with the pricing subgame.

2.3 Pricing subgame

Suppose that OC has generated content of quality x. If $s > 1$, then the pricing choices of OC and AI interact. There are two relevant cases: (1) where $p_{AI} = u(y)$ and all consumers purchase the AI product and (2) where $p_{AI} > u(y)$ and only some consumers purchase the AI product. The following proposition characterises the equilibrium outcome in the pricing subgame, holding x and y as fixed as these are chosen before the subgame commences.

Proposition 1 *If $\frac{u(y)}{x} \geq \frac{1}{2}\left(3 - 2\sqrt{2}\right)(1 - \rho(s))\rho(s)$, a Nash equilibrium of the pricing subgame exists with $\hat{p}_{OC} = \frac{1}{2}(1 - \rho(s))x$ and $\hat{p}_{AI} = u(y)$. This equilibrium is unique if $\frac{u(y)}{x} \geq \frac{1}{2}(1 - \rho(s))\rho(s)$. OC profits are $\frac{1}{4}(1 - \rho(s))x - c_{AI}(x)$ and AI profits are $u(y) - c_{AI}(y, sx)$.*

The proofs of all propositions are in the appendix. In the case of Proposition 1, which characterises one type of equilibrium, the proof in the appendix characterises addition equilibria

[7] This assumption may not be reasonable if the AI provider cannot tell in advance which training data they are using that may be infringing. If this is the case, a model more like that presented in Section 4 would be applicable and some of the broader intellectual property issues such as those explored by Green and Scotchmer (1995) will arise.

in the pricing subgame. Also, as shown in the appendix, when the bargaining stage is overlayed prior to the pricing subgame, the other equilibrium outcomes become plausibly less relevant, and that is why their characterisation and discussion are placed in the appendix.

The equilibrium outcome highlighted in the proposition involves the AI provider targeting the direct utility of AI products rather than any additional consumer benefit that might arise from the possibility of reproducing original content without payment. The original content provider is, however, constrained in its pricing by the possibility of leakage, which is open to all consumers since they all purchase AI productions in equilibrium. Nonetheless, the original content provider sells to those consumers with the highest willingness to pay for their content.

3 "Small" Model AI

The first environment we examine is where there is an identifiable source of original content that the AI provider wants to train on. It is assumed, as is done throughout this paper, that the original content is readily available, meaning that AI does not have to deal with OG in order to access the content. The ready availability of content does not mean it cannot be protected by copyright. The question explored here is whether and why it ought to be when training certain types of AI models. In this section, "small" AI models are considered. They are small in the sense that the original content owner can be identified, allowing negotiations to take place prior to AI training. The next section will look at "large" AI models where that is impossible, precluding the model timeline's negotiation stage.

Recall that the equilibrium of the pricing subgame involves AI setting $\hat{p}_{AI} = u(y)$ and OC setting $\hat{p}_{OC} = \frac{1}{2}(1 - \rho(s))x$. In this case, for a given s, the profits of each party are:

$$\pi_{OC}(x, s) = \tfrac{1}{4}(1 - \rho(s))x + \tau(s, x, y) - c_{OC}(x)$$

$$\pi_{AI}(y, x, s) = u(y) - \tau(s, x, y) - c_{AI}(y, sx)$$

where τ is the net lump-sum payment (if any) from AI to OC. Note that τ can be negative.

3.1 Nash bargaining outcome

When OC and AI negotiate over (s, τ), they do so anticipating AI's choice of quality, $\hat{y}(s, x)$, and the outcome of the pricing subgame. Below, the y chosen if there is a negotiated outcome is written $\hat{y}_n$, leaving out the arguments (s, x) for notational convenience. It is only if this negotiation breaks down that the rights regime plays a role. Specifically, let s_r be the chosen level of s in rights regime $r \in \{NC, CP\}$, then OC's expected payoff in the absence of a negotiated outcome if $\pi_{OC}(x, s_r)$ while AI's expected payoff is $\pi_{AI}(\hat{y}_r, s_r, x)$. Here $\hat{y}_r$ is the chosen level of y that maximises AI's expected payoff if negotiations break down under the rights regime, r.

In the absence of frictions such as information asymmetries, negotiations are generally expected not to break down, and the parties will agree to a mutually beneficial use of content and payments. As no bargaining frictions are assumed here, the content used for AI training realised through negotiations will be the same regardless of the rights regime. That regime will only change the price (τ) associated with that content use.

To see this negotiated outcome, recall that it is assumed to follow the Nash bargaining solution with transferable utility – in this case, AI pays OC a sum payment of τ (which may be negative) if $s \neq s_r$. Having chosen s, AI will train its model at a quality of $\hat{y}_n$ given by the first-order condition:

$$\frac{\partial u(\hat{y}_n)}{\partial y} = \frac{\partial c_{AI}(\hat{y}_n, x)}{\partial y}$$

Note that $\hat{y}_n$ is non-decreasing in s and x. In Nash bargaining, s is chosen to maximise the sum of expected payoffs of OC and AI:

$$\max_s \tfrac{1}{4}(1 - \rho(s))x + u(\hat{y}_n) - c_{AI}(\hat{y}_n, sx)$$

Let $\hat{s}_n$ be the optimal s that solves this problem (strictly speaking, it is $\hat{s}_n(\hat{y}, x)$ but is shortened for notational convenience). If $\hat{s}_n$ is an interior solution, $\hat{s}_n$ is characterised by the first-order condition of $-\frac{1}{4}\frac{\partial \rho(s)}{\partial s}x + \frac{\partial u(\hat{y}_n)}{\partial y}\frac{\partial y}{\partial s} = \frac{dc_{AI}(\hat{y}_n, sx)}{ds}$ or, applying the envelope theorem,

$$-\tfrac{1}{4}\frac{\partial \rho(\hat{s}_n)}{\partial s} = x\frac{\partial c_{AI}(\hat{y}_n, \hat{s}_n x)}{\partial s}$$

It is instructive to note when $\hat{s}_n$ is a corner solution. This will occur when either the impact on OG profits from leakage is very high (i.e., $\frac{\partial \rho(s)}{\partial s}$ is large) that pushes s to 0 or when the impact on AI's training costs from using original content is very high (i.e., $\frac{\partial c_{AI}(y, sx)}{\partial s}$ is large) that pushes s to 1.

It is useful to examine this using our specific functional forms. First note that $\hat{y}_n = 1 + \gamma s x$. Second, given this, the marginal net benefit of s in terms of joint profit is $\frac{2\gamma - \bar{s}}{4}x$. Thus,

$$\hat{s}_n = \begin{cases} 1 & 2\gamma \geq \bar{s}\alpha \\ 0 & 2\gamma < \bar{s}\alpha \end{cases}$$

Note that when $s = 1$, joint profits are $\frac{1}{4}(2 + (1 + 2\gamma - \bar{s})x)$ and when $s = 0$ joint profits are

$\frac{x+2}{4}$. Thus, $\hat{s}_n = 1$ if $\bar{s} \leq 2\gamma$ and $\hat{s}_n = 0$ if $\bar{s} > 2\gamma$. Put simply, it is worthwhile negotiating AI full use of original content when marginal training costs are impacted greatly by sx and/or $\bar{s}$ is low.

Given the negotiated level of s and the resulting choice of y, the parties' expected payoffs will be $\pi_{AI}(\hat{y}_n, \hat{s}_n, x) - \tau$ and $\pi_{OC}(\hat{s}_n, x) + \tau$. The negotiated payment $\hat{\tau}$ is found by equating the net surpluses of each party; that is:

$$\pi_{OC}(\hat{s}_n, x) + \hat{\tau} - \pi_{OC}(x, s_r) = \pi_{AI}(\hat{y}_n, \hat{s}_n, x) - \hat{\tau} - \pi_{AI}(\hat{y}_r, s_r, x)$$

$$\Leftrightarrow \hat{\tau} = \tfrac{1}{2}\Big(\pi_{AI}(\hat{y}_n, \hat{s}_n, x) - \pi_{OC}(\hat{s}_n, x) - \pi_{AI}(s_r, \hat{y}_r, x) + \pi_{OC}(s_r, x)\Big)$$

OC's expected payoff, $\pi_{OC}(\hat{s}_n, x) + \hat{\tau}$, becomes:

$$\tfrac{1}{2}\left(\tfrac{1}{4}\left(2 - \rho(\hat{s}_n) - \rho(s_r)\right)x + u(\hat{y}_n) - c_{AI}(\hat{y}_n, \hat{s}_n x) - (u(\hat{y}_r) - c_{AI}(\hat{y}_r, s_r x))\right)\Big)$$

Therefore, anticipating this OC's choice of x, $\hat{x}_r$ is determined by:

$$\frac{2 - \rho(\hat{s}_n) - \rho(s_r)}{8} + \frac{1}{2}\left(s_r \frac{\partial c_{AI}(\hat{y}_r, s_r \hat{x}_r)}{\partial x} - \hat{s}_n \frac{\partial c_{AI}(\hat{y}_n, \hat{s}_n \hat{x}_r)}{\partial x}\right) = \frac{\partial c_{OC}(\hat{x}_r)}{\partial x}$$

where the envelope theorem is applied on s and y. This demonstrates that how the rights regime impacts on OG's choice of content is through $\rho(s_r)$ and $\frac{\partial c_{AI}(\hat{y}_r, s_r \hat{x}_r)}{\partial x}$.

Using our specific functional forms, note that:

$$\hat{x}_r = \frac{2 - \rho(s_r) - \rho(s) + 2\gamma(s - s_r)}{8} \qquad \hat{y}_r = 1 + \gamma s \frac{2 - \rho(s_r) - \rho(s) + 2\gamma(s - s_r)}{8}$$

The important finding here is that AI training quality is higher, the lower is s_r due to the feedback effect through OC's choice of x.

3.2 No IP protection (NC)

If there is no IP protection, then AI can unilaterally set $s = 1$.[8] If there is no negotiated outcome then AI will choose $\hat{y}_{NC}$ to maximise its own profits (with $\tau = 0$) giving the first-order condition:

$$\frac{\partial u(\hat{y}_{NC})}{\partial y} = \frac{\partial c_{AI}(\hat{y}_{NC}, x)}{\partial y}$$

If this occurs, AI's profits are $\pi_{AI}(1, \hat{y}_{NC}, x)$ and OC's profits are $\pi_{OC}(1, x)$.

[8] AI may face costs in acquiring the original content, but for simplicity, these are set aside for the analysis here.

Applying the Nash bargaining solution as outlined above with $s_{NC} = 1$ implies that OC's choice of x will be $\hat{x}_{NC}$ as determined by:

$$\tfrac{1}{8}\left(2 - \rho(\hat{s}_n) - \rho(1)\right) + \tfrac{1}{2}\left(\frac{\partial c_{AI}(\hat{y}_{NC}, \hat{x}_{NC})}{\partial x} - \hat{s}_n \frac{\partial c_{AI}(\hat{y}_n, \hat{s}_n \hat{x}_{NC})}{\partial x}\right) = \frac{\partial c_{OC}(\hat{x}_{NC})}{\partial x}$$

Note that when $\hat{s}_n = 1$, then OC does not take into account its contribution to reducing AI training cost at all in choosing its content quality, while when $\hat{s}_n = 0$ that contribution *diminishes* its incentives to raise content quality as that only strengthens the bargaining position of AI. Finally, note that, under (NC), $\tau_{NC} \leq 0$ with equality if $\hat{s}_n = 1$.

For the purposes of comparison going forward, it is useful to state these outcomes using our specific functional forms:

$$\hat{x}_{NC} = \begin{cases} \frac{1-\bar{s}}{4} & 2\gamma \geq \bar{s} \\ \frac{1}{8}(2 - \bar{s} - 2\gamma) & 2\gamma < \bar{s} \end{cases} \qquad \hat{y}_{NC} = \begin{cases} 1 + \gamma\frac{1-\bar{s}}{4} & 2\gamma \geq \bar{s} \\ 1 & 2\gamma < \bar{s} \end{cases}$$

This shows starkly how, to the extent that AI causes potential substitution for consumers in purchasing original content (i.e., the level of $\bar{s}$), this diminishes the quality chosen by both OC and AI when AI is pricing for general purpose AI. Note that when $\bar{s}$ is low, OC's choice of content becomes less distorted by the presence of AI. Given this, social welfare under (NC) is:

$$\begin{cases} (3+\bar{s})\frac{1-\bar{s}}{32} + \frac{1}{2}\left(1 + \gamma\frac{1-\bar{s}}{4}\right) - \frac{1}{2}\left(\frac{1-\bar{s}}{4}\right)^2 & 2\gamma \geq \bar{s} \\ \frac{3(2-\bar{s}-2\gamma)}{64} + \frac{1}{2} - \frac{1}{2}\left(\frac{2-\bar{s}-2\gamma}{8}\right)^2 & 2\gamma < \bar{s} \end{cases}$$

Here, however, there are social welfare benefits to consumer substitution away from original content purchases as this increases the consumption of original content at least when AI is trained on that content (i.e., when $\hat{s}_n = 1$).

3.3 Copyright Protection (CP)

If there is complete IP protection, then OC can unilaterally set $s = 0$. If there is no negotiated outcome, AI will choose $\hat{y}_{CP}$ to maximise its own profits with the corresponding first-order condition:

$$\frac{\partial u(\hat{y}_{CP})}{\partial y} = \frac{\partial c_{AI}(\hat{y}_{CP}, 0)}{\partial y}$$

If this occurs, AI's profits are $\pi_{AI}(0, \hat{y}_{CP}, 0)$ and OC's profits are $\pi_{OC}(0, x)$.

Applying the Nash bargaining solution as outlined with $s_{CP} = 0$ OC's choice of x be $\hat{x}_{CP}$

as determined by:

$$\tfrac{1}{8}(2-\rho(\hat{s}_n))-\tfrac{1}{2}\hat{s}_n\frac{\partial c_{AI}(\hat{y}_n,\hat{s}_n\hat{x}_{CP})}{\partial x}=\frac{\partial c_{OC}(\hat{x}_{CP})}{\partial x}$$

Comparing this with the equivalent first-order condition under (NC), note that $\hat{x}_{CP}\geq\hat{x}_{NC}$ as

$$\tfrac{1}{4}\rho(1)>\frac{\partial c_{AI}(\hat{y}_{NC},\hat{x}_{NC})}{\partial x}$$

Intuitively, OC increases its level of quality under (CP) because, when it can unilaterally prevent AI from using its content, it does not have to be concerned about a revenue reduction in competition with AI and it can appropriate a higher share of the marginal contribution of its content quality's role in reducing AI training costs both of which increase the negotiated τ under (CP) relative to (NC).

Using our specific functional forms, this gives:

$$\hat{x}_{CP}=\begin{cases}\frac{2-\bar{s}+2\gamma}{8} & 2\gamma\geq\bar{s}\\ \frac{1}{4} & 2\gamma<\bar{s}\end{cases}\qquad\qquad \hat{y}_{CP}=\begin{cases}1+\gamma\frac{2-\bar{s}+2\gamma}{8} & 2\gamma\geq\bar{s}\\ 1 & 2\gamma<\bar{s}\end{cases}$$

As noted earlier, when $\hat{s}_n=1$, $\hat{x}_{CP}>\hat{x}_{NC}$. When $\hat{s}_n=0$, then OC effectively is shielded from the existence of general-purpose AI, and so it and AI choose their qualities as if the other did not exist. Nonetheless, while the under-supply of content to consumers still exists, that distortion has not changed as we move from an (NC) to a (CP) rights regime. Social welfare under (CP) is:

$$\begin{cases}(3+\bar{s})\frac{2-\bar{s}+2\gamma}{64}+\frac{1}{2}\left(1+\gamma\frac{2-\bar{s}+2\gamma}{8}\right)-\frac{1}{2}\left(\frac{2-\bar{s}+2\gamma}{8}\right)^2 & 2\gamma\geq\bar{s}\\ \frac{9}{16} & 2\gamma<\bar{s}\end{cases}$$

This is higher than the corresponding levels in (NC) because x and y are higher. In fact, the improvement in social welfare is a general outcome in this model, as summarised by the following proposition.

Proposition 2 *Comparing (CP) and (NC), $\hat{x}_{CP}>\hat{x}_{OC}$, $\hat{y}_{CP}>\hat{y}_{NC}$ and expected social welfare is higher under (CP) than (NC).*

The proof (omitted) follows from the inspection of OC's first order conditions for x and noting that $\hat{y}_n$ is non-decreasing in x and $\hat{s}_n$ is the same under either rights regime.

In summary, when negotiations are possible, then original content quality and, by implication, AI training quality are both higher under full copyright protection than under no copyright protection. This is because such protection provides the strongest incentives for OC while AI's choices take place after a negotiated outcome, and so are not made to influence those negotiations. If AI's training occurred in whole or in part before those negotiations, its incentives would change as it would also be subject to hold-up.[9] That said, even under (CP), the quality levels are lower than the levels that would maximise social welfare.

3.4 Fair Use (FU)

Using our specific function forms, recall that maximising social welfare involves $s = 1$ and solving for the optimal qualities gives $x^* = \frac{2+\gamma}{4}$ and $y^* = \frac{4+\gamma(2+\gamma)}{8}$. These optimal outcomes arise, under (CP), as $\bar{s} \to 0$. But when $\bar{s}$ becomes high, the chosen qualities fall, and the parties may negotiate $s = 0$, but more consumers consume original content.

One of the consequences of a policy of fair use is that it promotes further use of the content that otherwise might be restricted by a copyright holder. One of the criteria for fair use corresponds to the degree of harm such use may cause the copyright holder. In this model, this is captured by $\bar{s}$ or $\rho(1)$. Thus, one way of conceiving of a fair use policy is whether in situations where $\bar{s}$ is below some threshold, Γ, the rights regime should be (NC) rather than (CP).

Note that, for our specific functional forms, when $\bar{s} > 2\gamma$, the parties would negotiate $\hat{s}_n = 0$, and so no content sharing or additional consumption would arise if the fair use threshold was in this range. Thus, if there is an optimal threshold, then $\Gamma \in [0, 2\gamma)$. However, implementing fair use in this range, compared with (CP), would only lower quality choices and not change the level of realised consumption. Thus, if there is a rationale for fair use, it lies beyond the small AI model setup here.

4 Large AI Models

The core assumption for the analysis of small AI models was that the use of original content could be identified and that sufficient information regarding its likely impact on the content creator's interests could be determined and negotiated over. This core assumption cannot be presumed to hold for large AI models that are trained on a large amount of content where the provenance of any particular piece of content cannot be readily determined. As noted in the introduction, it is this situation that mirrors the use of content in "training" humans that is not typically considered to infringe copyright laws.

[9] This occurred in the analysis of Gans (2015) although even there full copyright often generated higher investments from both parties due to the complementarity between original content investments and those of subsequent users in that context.

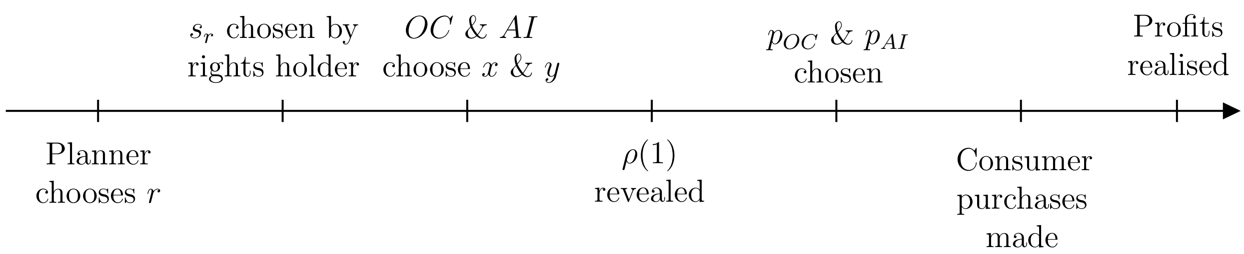

Figure 2: **Large AI Model Timeline**

To model this context, three changes are made to the baseline model. First, it is presumed that ex-ante negotiations between an original content provider and an AI provider are not feasible, and so there is no negotiation stage.

Second, the precise impact of an AI model on an OC's profits cannot be determined ex ante. Specifically, it is assumed that $\rho(1)$ (or $\bar{s}$) is independently and identically distributed amongst content providers according to a cumulative distribution function, $F(.)$ with corresponding density function, $f(.)$ distributed on $[0, 1]$. After an AI model is launched but prior to the pricing subgame, $\rho(1)$ for a content provider is revealed to all and thus, its pricing takes that into account. The timeline for the Large AI model is shown in Figure 2.

Finally, original content is valued by N distinct subgroups of a unit measure in the economy with willingness to pay uniformly distributed as in the baseline model. To capture the notion that there is a "large" amount of content, N is the measure of a continuum of content. These content markets are independent. The OC in each of these markets is indexed by $i \in [0, N]$. For notational simplicity and comparison with the earlier model, N is set equal to 1. Each consumer still has a common direct and identical value for the AI product of $u(y)$. Thus, the changes made to the baseline model here allow that model's pricing and other investment stages to carry over without change.

4.1 Outcomes Under Traditional Rights Regimes

We begin by examining the traditional rights regimes of (NC) and (CP). Given no negotiations take place over the use of original content in training AI systems, the rights regime dictates what will take place. Under (NC), the AI provider can simply make use of any content that is available. Under (CP), each OC can, using the threat of, say, significant statutory damages, restrict content use in AI, and the AI provider will not use that content. Specifically, an OC can sue the AI provider incurring a cost of C and, if successful, receive a damages payout of $D > C$. D is assumed to be high enough that, under full copyright protection (CP), the AI provider does not want to use the content and $s = 0$.[10]

[10] That is $D > \max_y\{u(y) - c_{AI}(y, x)\}$.

The following proposition characterises the outcomes under traditional rights regimes.

Proposition 3 *In the large AI model, for each OC, i, $\hat{x}_{i,CP} \geq \hat{x}_{i,NC}$ and for AI, $\hat{y}_{CP} \leq \hat{y}_{NC}$. Using the specific function forms, expected social welfare under (NC) will exceed that under (CP) if and only if:*

$$\gamma > \frac{\mathbb{E}[\rho_i(1)]^2}{2(1 - \mathbb{E}[\rho_i(1)])}$$

The proof (in the appendix) compares the first-order conditions for each type of investment and follows on by calculating social welfare under (NC) and (CP), respectively.

The condition in the proposition has a strong intuition. It is better to permit large AI models to train models without any copyright liability if the value of the content in reducing training costs exceeds a measure of expected harm from leakage of content on content providers. This, of course, rationalises why human training is not subject to copyright liability, as the potential harm from this on content provider profits is unlikely to be large on average.

4.2 Ex post 'Fair Use' Regime

Both of the traditional rights regimes involve inefficiencies: (CP) potentially restricting the use of original content and both regimes trading off the incentives of the OCs and AI. The question is whether a different mechanism can create a more favourable balance of incentives for both original content investment and AI training quality while preserving the use of original content in consumption and AI training.

Recall that, in the small AI model context, the AI provider would agree not to use content if the potential harm from leakage of that content were high. Of course, here, the extent to which leakage might be harmful to individual content providers is not known ex ante. It is, however, known ex post. This suggests a potential mechanism that involves liability assessed ex post is feasible. That mechanism would proceed as follows:

1. If $\rho_i(1) < \Gamma$, the AI provider is not liable for any damages from the use of i's content.
2. If $\rho_i(1) \geq \Gamma$, the AI provider must pay full compensation to i of an amount $D(\rho_i(1)) = \frac{1}{4}\rho_i(1)\hat{x}_{FU}$

Here (FU) stands for 'fair use', and $\hat{x}_{FU}$ is the chosen content quality under that regime. The level of damages, if triggered, allows the content provider to earn $\frac{1}{4}\hat{x}_{FU}$ as if the leakage never occurred.[11]

[11] This mechanism is similar to the Remix Rights with Full Compensation mechanism explored in Gans (2015).

The following proposition characterises the outcomes under (FU) and compares it to traditional rights regimes.

Proposition 4 *Suppose that AI earns non-negative profits under the (FU) mechanism. In the large AI model, for each OC, i,* $\hat{x}_{i,CP} \geq \hat{x}_{i,FU} \geq \hat{x}_{i,NC}$ *and for AI,* $\hat{y}_{FU} \geq \hat{y}_{CP} \leq \hat{y}_{NC}$*. Using the specific function forms, expected social welfare under (NC) will exceed that under (CP) if and only if:*

$$\gamma > \frac{\mathbb{E}[\rho(1)|\Gamma]^2 + \mathbb{E}[\rho(1)|\Gamma] - \mathbb{E}[\rho(1)](1 - \mathbb{E}[\rho(1)|\Gamma])}{4(1 - \mathbb{E}[\rho(1)|\Gamma])}$$

where $\mathbb{E}[\rho(1)|\Gamma] = \int_0^\Gamma \rho_i(1) dF(\rho_i(1))$*.*

The proof (in the appendix) compares the first-order conditions for each type of investment and follows on by calculating social welfare under (FU) and compares it to those under traditional rights regimes. The condition on *AI* profits arises because it is possible that Γ is such that:

$$u(\hat{y}_{FU}) - c_{AI}(\hat{y}_{FU}, \hat{x}_{i,FU}) < (1 - F(\Gamma)) \int_\Gamma^1 \tfrac{1}{4}\rho(1)\hat{x}_{i,FU} dF(\rho(1))$$

in which case *AI* would not be finally viable because expected damages are above its profits from AI. Below, when we examine what Γ is chosen by the planner, this feasibility constraint needs to be satisfied.

The (FU) mechanism has two advantages over traditional rights regimes. First, under the (FU) mechanism, *AI* chooses to use all content in training the AI (i.e., $s = 1$). Thus, the prices for content and the AI have the same structure as under (NC) but, importantly, in terms of the consumption of that content, all consumers will do so (potentially) with half of the market purchasing from *OC* and the other half relying on the AI product. Thus, the (FU) mechanism allows content to be used at its socially desirable level for AI training and also to be consumed relatively widely.

Second, this has an impact on incentives. As all content is used in AI training, AI training quality under (FU) is the same as that under (NC) for fixed content quality and, therefore, highest among the three regimes compared as *OC* content has a higher quality under (FU) than (NC). *OC* content is higher than under (NC) because if potential harm from AI training exceeds a certain threshold, each *OC* expects to receive a damages payout. Thus, from their perspective, while under (NC) their potential appropriation is $\frac{1}{4}(1 - \mathbb{E}[\rho_i(1)])$ for each unit of x_i, under (FC) it is $F(\Gamma)\frac{1}{4}\big(1 - \mathbb{E}[\rho_i(1)|\Gamma]\big) + (1 - F(\Gamma))\frac{1}{4}$, a higher marginal return as $\mathbb{E}[\rho_i(1)|\Gamma] \leq \mathbb{E}[\rho_i(1)]$ (with an equality if $\Gamma = 0$).

These advantages hold for any $\Gamma < 1$. But what level of Γ would be optimal? Note that expected social welfare is higher as Γ is reduced, and there is a lower threshold of harm that can trigger compensation. This has the effect of raising both content provider incentives and AI training incentives. The latter effect arises because, regardless of Γ, all content is available for AI training. Moreover, a lower Γ does not change the total number of consumers who consume original content (that is, $\frac{1}{2} + \frac{1}{2}\mathbb{E}[\rho(1)]$).

Note that as $\Gamma \to 0$, $\hat{x}_{FU} \to \hat{x}_{CP}$ and $\hat{y}_{FU} \to 1 + \gamma\frac{1}{4}$. This is feasible if AI profits $(u(\hat{y}_{FU}) - c_{AI}(\hat{y}_{FU}, \hat{x}_{FU}) - \int_0^1 \frac{1}{4}\rho(1)\hat{x}_{FU} dF(\rho(1)))$ remain positive. Compared with (CP), $\Gamma = 0$ behaves like a regime where all content is copy-protected. However, it is permissive of infringement with compensation, which generates the benefits of original content for AI training without diminishing the incentives of original content providers to invest in quality. In effect, the difference is akin to a prohibition versus a Pigouvian-like tax on the use of original content where external effects are internalised through ex-post prices.

In practice, an (FU) mechanism is respectful of transaction costs associated with asserting copyright protection when limited harm occurs. For this reason, it also rationalises the human scenarios explored earlier in that those scenarios are very unlikely to cause commercial harm to original content providers, even in the aggregate.

5 Conclusion

There is a sense in which the copyright challenges associated with generative AI are not that different to those that arose with digitisation and the rise of the Internet. In each case, a new technology lowered the cost of utilising original content while enhancing its potential scope or value. For generative AI, this arises because such AI can be trained on original content but also, precisely because that content is widely available, reproduce that content within AI products themselves. This threatens existing business models for original content providers that, not surprisingly, they are protective of.

Copyright law played a role in the evolution of digital technologies. But one thing that was preserved was long-standing copyright protections, including the ability to prevent use except where that use was regarded as fair. This paper has analysed generative AI, taking into account the particular role of such technologies as well as how generative AI products threaten original content creators' commercial activities. It is demonstrated that different rights regimes can lead to distinct social welfare outcomes but that these rest firmly on the ability of original content providers and AI providers to negotiate over the use of copy-protected material in AI training. When they can negotiate – something that requires knowledge of what content is proposed to be used in AI training as well as what potential harm to original content providers' commercial activities it might cause – then traditional full copyright protection leads to stronger incentives to produce original content and that this determined the social welfare properties of such protections. The conditions that allow for this, however, require that the AI models be small in the sense of using and relying on a relatively small corpus of content. Such ”fine-tuned” or RAG models are currently being developed. The value of copyright protection, however, rests on the potential for commercial harm. If that potential is low, as it arguably is for similar human uses of content in their own “training,” then a strong protective regime may not be of much social value.

Consequently, it is when the AI models are large in the sense of using an almost unimaginable amount of content in training that the type of rights regime imposed can have important social welfare consequences. Because negotiations are prohibitive as it is hard to identify copyright use, let alone determine clearly whether copy-protected content has been used in training, copyright protection, while potentially strengthening incentives to create original content, serves to limit its use both for consumption and training, which involves adverse social consequences. No copyright protection does not involve those costs, but if the threat to original content providers' commercial activities is high, it could undermine the production of original content and, in the process, its uses. However, if the threat to those commercial activities is low, as it arguably is for equivalent human uses of content, then a permissive regime for the use of content in AI training is socially desirable.

For that reason, this paper identifies an alternative mechanism based on fair use. That mechanism permits the use of copy-protected content in AI training but subjects AI providers to damages should the realised commercial harm of content providers be large. Thus, the purpose here is to provide some insurance. The presumption is that content used in AI training is, for the most part, not likely to damage original content providers' commercial interests. However, should it turn out to do so, they will receive compensation. Insured against such risks, the original content providers' incentives to create content are undiminished. This represents a practical way of respecting copyright ownership while also allowing its large-scale use in AI training in a way that minimises transaction costs that might arise from injunctive legal action or license negotiations. It is also conceivable that copyright holders could opt out of this type of regime, although the technical requirements to do so would likely require further development as it would require content used to be *traceable*.

When a new technology arises that creates copyright challenges, a clarification of the rights regime can often lead to institutional and technological developments to “make it work.” This happened for music rights with radio and public broadcasting in the form of collecting societies. For AI, Besen (2023) argues that the large volume of data lends itself towards collecting copyright payments by collective societies as seen in other domains such as music. He believes that this would have to be established by government regulation. But in other areas, the developments were technological. For instance, Google developed a rights management system for YouTube that allowed people to post content with copy-protected materials but for rights holders to be notified and then to utilise a revenue-sharing arrangement (Gans (2015)). By settling on a policy approach to generative AI, even if some inefficiencies remain at present, these may incentivise new ways of minimising those inefficiencies.

In summary, this paper represents a first, admittedly high-level, pass at the economics of copyright issues related to generative AI. The presumption here is that the AI providers might be liable for an infringement rather than the users of AI. It remains an open question whether, even under the rights regimes explored here, those are the correct focal parties. Nonetheless, policy approaches in this area will likely be informed by empirical research that allows a clearer picture of the parameters that determine which rights regime may be preferable.

6 Appendices

6.1 Proof of Proposition 1

Proposition 1 states the conditions for the existence of one type of equilibrium. Rather than proving that proposition only here, the following proposition that characterises all of the pure strategy Nash equilibrium outcomes of the pricing subgame is proved.

Proposition 5 *Let* $\Theta_1 \equiv \frac{1}{2}\left(3-2\sqrt{2}\right)(1-\rho(s))\rho(s)$ *and* $\Theta_2 = \frac{1}{2}(1-\rho(s))\rho(s)$. *The Nash equilibrium prices and payoffs of the pricing subgame are as follows:*

1. *If* $\frac{u(y)}{x} \geq \Theta_1$, $\hat{p}_{OC} = \frac{1}{2}(1-\rho(s))x$ *and* $\hat{p}_{AI} = u(y)$. *OC profits are* $\frac{1}{4}(1-\rho(s))x$ *and AI profits are* $u(y)$.

2. *If* $\frac{u(y)}{x} < \Theta_2$, $\hat{p}_{OC} = \frac{2(1-\rho(s))x-u(y)}{4-\rho(s)}$ *and* $\hat{p}_{AI} = \frac{(2-\rho(s))u(y)+(1-\rho(s))\rho(s)x}{4-\rho(s)}$. *OC profits are* $\frac{(2(1-\rho(s))x-u(y))^2}{(4-\rho(s))^2(1-\rho(s))x}$ *and AI profits are* $\frac{((2-\rho(s))u(y)+(1-\rho(s))\rho(s)x)^2}{(4-\rho(s))^2(1-\rho(s))\rho(s)x}$.

For $\frac{u(y)}{x} \in [\Theta_1, \Theta_2]$, *there are two pure strategy Nash equilibrium akin to (1) and (2). If* $\frac{u(y)}{x} > \Theta_2$, *the unique pure strategy Nash equilibrium is (1) while if* $\frac{u(y)}{x} < \Theta_1$, *the unique pure strategy Nash equilibrium is (2).*

The proof proceeds by characterising the conditions that support equilibrium (1) – the equilibrium of Proposition 1 – before characterising the conditions that support equilibrium (2). Then, the conditions are examined to demonstrate the second half of the proposition.

Note that for AI, the inverse demand curve has a kink at $p_{AI} = u(y)$ where it is downward sloping for $p_{AI} > u(y)$ and then flat up to the size of the market. Thus, evaluating any Nash equilibrium outcome will involve examining AI's payoff at $p_{AI} = u(y)$ and $p_{AI} > u(y)$, respectively, holding p_{OC} constant.

First, suppose that $p_{AI} = u(y)$. In this case, the purchasers of original content are high θ types greater than a marginal type, θ_{OC}, which is the type where $(1-\rho(s))\theta_{OC}x = p_{OC}$ or $\theta_{OC} = \frac{p_{OC}}{(1-\rho(s))x}$. Given this, OC chooses p_{OC} to maximise $p_{OC}(1-\frac{p_{OC}}{(1-\rho(s))x})$. This gives $\theta_{OC} = \frac{1}{2}$ and $\hat{p}_{OC} = \frac{1}{2}(1-\rho(s))x$ as in the statement of equilibrium (1).

The consumer indifferent between purchasing AI or not will be θ_{AI} so that $u(y) + \rho(s)\theta_{AI}x = p_{AI}$ or $\theta_{AI} = \frac{p_{AI}-u(y)}{\rho(s)x}$. Note that $\theta_{AI} < \theta_{OC}$ implies that $\rho(s)x > 2(p_{AI}-u(y))$. This constrains p_{AI}. As soon as $p_{AI} > u(y) + \frac{1}{2}\rho(s)x$, demand for the AI product falls to 0.

At $\hat{p}_{OC}$, AI earns $u(y)$ if $p_{AI} = u(y)$. To examine whether a deviation is profitable, note that if AI chooses to set $p_{AI} > u(y)$, then those consumers who purchase original content (that is, consumers with $\theta \geq \theta_{OC}$) will not find it optimal to purchase AI. There are two thresholds to consider for the purchase of AI. First, those who purchase OC at the new p_{AI}

will be indifferent between doing so and purchasing AI if $u(y)+\rho(s)\theta x - p_{AI} = \theta x - \hat{p}_{OC} = \theta x - \frac{1}{2}(1-\rho(s))x$ or $\theta \leq \frac{1}{2} - \frac{p_{AI}-u(y)}{(1-\rho(s))x}$ will purchase AI. Second, those who do not purchase OC at the new p_{AI} will be indifferent between purchasing the AI or not if $u(y)+\rho(s)\theta x = p_{AI}$ or $\theta \geq \frac{p_{AI}-u(y)}{\rho(s)x}$. Given this, assuming that it sets $p_{AI} \in (u(y), u(y)+\frac{1}{2}\rho(s)x)$, if it deviates AI will deviate to a p_{AI} that maximises:

$$p_{AI}\left(\frac{1}{2} - \frac{p_{AI}-u(y)}{(1-\rho(s))x} - \frac{p_{AI}-u(y)}{\rho(s)x}\right) = p_{AI}\left(\frac{1}{2} - \frac{p_{AI}-u(y)}{(1-\rho(s))\rho(s)x}\right)$$

This gives $\hat{p}_{AI} = \frac{1}{4}(2u(y)+(1-\rho(s))\rho(s)x)$ (which is less than $u(y)+\frac{1}{2}\rho(s)x$) and profits of $\frac{(2u(y)+(1-\rho(s))\rho(s)x)^2}{16(1-\rho(s))\rho(s)x}$. Note that $\hat{p}_{AI} > u(y)$ if and only if $\frac{u(y)}{x} < \frac{1}{2}(1-\rho(s))\rho(s)$ while profits are higher than $u(y)$ only if $\frac{u(y)}{x} > \frac{1}{2}\left(3-2\sqrt{2}\right)(1-\rho(s))\rho(s) = \Theta_1$. Thus, for $\frac{u(y)}{x} \in [\Theta_1, \frac{1}{2}(1-\rho(s))\rho(s))$ a deviation to $p_{AI} > u(y)$ is feasible but not profitable. Hence, equilibrium (1) exists when $\frac{u(y)}{x} \geq \Theta_1$.

Second, suppose that $p_{AI} > u(y)$. In this case, if consumers who purchase from OC, would not purchase the AI. Again, the purchasers of original content are high θ types. In this case, however, if they did not purchase from OC, those consumers would purchase from AI. Thus, the willingness to pay of type θ for original content is $\theta x - (\rho(s)\theta x + u(y) - p_{AI})$. This implies that the marginal purchaser of original content is θ_{OC} such that $\theta_{OC}x - (\rho(s)\theta_{OC}x + u(y) - p_{AI}) = p_{OC}$ or $\theta_{OC} = \frac{p_{OC}-p_{AI}+u(y)}{(1-\rho(s))x}$. Thus, p_{OC} is chosen to maximise $p_{OC}(1-\theta_{OC})$. For AI, while θ_{AI} is the same as above, it chooses p_{AI} to maximise $p_{AI}(\theta_{OC}-\theta_{AI})$ so long as $\theta_{OC} > \theta_{AI}$. For the moment, this inequality will be taken as an assumption. Given this, the first order conditions for OC and AI are:

$$p_{OC} = \tfrac{1}{2}(p_{AI} - u(y) - (1-\rho(s))x)$$

$$p_{AI} = \tfrac{1}{2}(\rho(s)p_{OC} + u(y))$$

This gives, as stated in the proposition for equilibrium (2):

$$\hat{p}_{OC} = \frac{2(1-\rho(s))x - u(y)}{4-\rho(s)}$$

$$\hat{p}_{AI} = \frac{(2-\rho(s))u(y) + (1-\rho(s))\rho(s)x}{4-\rho(s)}$$

Note that $p_{AI} > u(y)$ if and only if $\frac{u(y)}{x} < \frac{1}{2}(1-\rho(s))\rho(s) = \Theta_2$ and that $\theta_{OC} > \theta_{AI}$ if and only if $(2-\rho(s))u(y) + (1-\rho(s)\rho(s)x > 0$ which always holds. Finally, OC profits are $\frac{(2(1-\rho(s))x-u(y))^2}{(4-\rho(s))^2(1-\rho(s))x}$ and AI profits are $\frac{((2-\rho(s))u(y)+(1-\rho(s))\rho(s)x)^2}{(4-\rho(s))^2(1-\rho(s))\rho(s)x}$ as stated in the proposition for equilibrium (2).

Given $\hat{p}_{OC} = \frac{2(1-\rho(s))x-u(y)}{4-\rho(s)}$ would AI's profits be higher if it set $p_{AI} = u(y)$? AI profits are greater than $u(y)$ if and only if:

$$\frac{u(y)}{x} < (1-\rho(s))\rho(s)\frac{12-(6-\rho(s))\rho(s)-(4-\rho(s))\sqrt{8-(4-\rho(s))\rho(s)}}{2(2-\rho(s))^2}$$

However, the right-hand side of this inequality is greater than Θ_2, so equilibrium (2) only holds for $\frac{u(y)}{x} \leq \Theta_2$.

Finally, a simple comparison of Θ_1 and Θ_2 demonstrates that $\Theta_1 < \Theta_2$. Given this, then it is possible that $\Theta_1 \geq \frac{u(y)}{x} < \Theta_2$ and both equilibrium outcomes (1) and (2) co-exist. If $\frac{u(y)}{x} \geq \Theta_2$, the unique equilibrium is (1) and if $\frac{u(y)}{x} < \Theta_1$, the unique equilibrium is (2).

6.2 Pricing AI for Imitation

Proposition 3, earlier in the appendix, shows that two relevant cases exist as equilibria in the pricing subgame. There are several interesting features to note. First, it is possible that the two equilibrium types co-exist. When AI sets a lower price, this creates additional competition for OC whose prices are low accordingly. A deviation from either party to a higher price reduces profits. Interestingly, when both set a high price, even though AI could obtain the entire market by lowering its price to $u(y)$, this is not profitable because, given OC's higher price, the set of consumers who purchase that AI product in equilibrium is high enough that serving the entire market is not profitable for AI.[12]

Second, OC prices and profits are higher in the equilibrium with $\hat{p}_{AI} > u(y)$ than where $\hat{p}_{AI} = u(y)$. When the AI price is high, the marginal consumer for OC is choosing between purchasing original content or the AI product. By contrast, when the AI price is low, the marginal consumer for OC is choosing between purchasing original content or accessing it via the AI. The second condition reflects more intense competition between OC and AI, whose price is effectively 0 as consumers purchase AI regardless.

Here, the impact of various rights regimes is characterised when equilibrium (2) in Proposition 3 is expected by the parties. As will be shown, once the full game is analysed, there are good reasons to believe that equilibrium (2) is not a very relevant case of interest. This is why the analysis is confined here to the appendix.

Consider the Nash bargaining outcome when equilibrium (2) is expected.

[12]Technically, AI's reaction function has a discontinuity and an upward jump, allowing it to intersect with OC's reaction function at two places.

Joint profits are:

$$\frac{(4-(3-\rho(s))\rho(s))u(y)^2 - 2(1-\rho(s))\rho(s)^2u(y)x + (1-\rho(s))^2\rho(s)(4+\rho(s))x^2}{(4-\rho(s))^2(1-\rho(s))\rho(s)x}$$

It can be shown that, over the range where equilibrium (2) exists (i.e., $\frac{u(y)}{x} < \Theta_2$), these profits are decreasing in ρ but are increasing in $u(y)$. Taking into account AI's choice of y, it can be demonstrated that the s that maximises these joint profits is $s = 0$.

However, at this point, as $\rho(s)$ becomes 0, equilibrium (2) no longer exists. Nonetheless, it can be demonstrated that over the entire range, i.e., $\frac{u(y)}{x} < \frac{1}{2}(1-\rho(s))\rho(s)$ where equilibrium (2) exists, joint profits are lower than $\frac{1}{4}x + u(y)$, the profits realised when $s = 0$. Thus, the parties will choose $s_n = 0$ in this case leading to the outcomes analysed in Section 3 that focussed on equilibrium (1).

6.3 Proof of Proposition 3

Under no copyright protection, the AI provider can use all of the content without liability. Thus, for each content i, $s_i = 1$ and, thus, the harm to an OC i's profits is determined by their individual $\rho_i(1)$. For each content provider, the pricing subgame reflects their individual draw of $\rho_i(1)$ and will be based on Proposition 1. Thus, each individual content, i, sells for $\hat{p}_i = \frac{1}{4}(1-\rho_i(1))x_i$. For AI, $\hat{p}_{AI} = u(y)$ and AI's profits are $u(y) - c_{AI}(y, \int s_i x_i di)$. Recall that, c_{AI} is not impacted by any individual s_i but is impacted by the sum of content used in training as there is a continuum of original content providers.

Given this, it is easy to see that $\hat{x}_{i,NC}$ and $\hat{y}_{NC}$ are determined by the following first order conditions:

$$\tfrac{1}{4}(1-\mathbb{E}[\rho_i(1)]) = \frac{\partial c_{OC}(x)}{\partial x}$$

$$\frac{\partial u(y)}{\partial y} = \frac{\partial c_{AI}(y, \int \hat{x}_i di)}{\partial y}$$

Note that the content quality is determined taking into account $\mathbb{E}[\rho_i(1)]$ as this is not known to each OC when it is investing in content.

Under full copyright protection, the AI can no longer use the content, and it is assumed that D is so high that it deters such use, so $s = 0$.[13] In this case, no content provider incurs any competitive harm from the AI provider, and so $\hat{p}_i = \frac{1}{4}x_i$ while $\hat{p}_{AI} = u(y)$ as in the (NC) case.

Given this, it is easy to see that $\hat{x}_{i,CP}$ and $\hat{y}_{CP}$ are determined by the following first order

[13] That is $D > \max_y\{u(y) - c_{AI}(y, x)\}$.

conditions:

$$\tfrac{1}{4} = \frac{\partial c_{OC}(x)}{\partial x}$$

$$\frac{\partial u(y)}{\partial y} = \frac{\partial c_{AI}(y,0)}{\partial y}$$

Comparing the first-order conditions to the (NC) case, note that $\hat{x}_{i,CP} > \hat{x}_{i,NC}$ while $\hat{y}_{CP} < \hat{y}_{NC}$.

To calculate the expected social welfare under each regime using our functional forms, we begin with (NC) before turning to (CP). For (NC), note that $\hat{x}_{NC} = \frac{1}{4}(1 - \mathbb{E}[\rho_i(1)])$ for all i and $\hat{y}_{NC} = 1 + \gamma \int \hat{x}_i di = 1 + \gamma \frac{1}{4}(1 - \mathbb{E}[\rho_i(1)])$. Given this, expected social welfare under (NC) is:

$$\begin{aligned}
&\int_0^1 \left(\int_{\frac{1}{2}}^1 \theta \tfrac{1}{4}(1 - \mathbb{E}[\rho_i(1)]) d\theta + \tfrac{1}{4}(1 - \mathbb{E}[\rho_i(1)]) \int_0^{\frac{1}{2}} \int_0^1 \rho_i(1)\theta dF(\rho_i(1)) d\theta \right) di \\
&+ \tfrac{1}{2}(1 + \gamma \tfrac{1}{4}(1 - \mathbb{E}[\rho_i(1)])) - \tfrac{1}{2}(\tfrac{1}{4}(1 - \mathbb{E}[\rho_i(1)]))^2 \\
&= \tfrac{1}{32}\big(3 + \mathbb{E}[\rho_i(1)]\big)(1 - \mathbb{E}[\rho_i(1)]) + \tfrac{1}{2}(1 + \gamma \tfrac{1}{4}(1 - \mathbb{E}[\rho_i(1)])) - \tfrac{1}{2}(\tfrac{1}{4}(1 - \mathbb{E}[\rho_i(1)]))^2 \\
&= \tfrac{1}{16}(9 + 2\gamma - \mathbb{E}[\rho_i(1)](2\gamma + \mathbb{E}[\rho_i(1)]))
\end{aligned}$$

Note this mirrors social welfare under (NC) in the small AI model case but for the substitution of the expected $\mathbb{E}[\rho_i(1)]$ for a known $\rho(1)$.

For (CP), $\hat{x}_{CP} = \frac{1}{4}$ for all i and $\hat{y}_{CP} = 1$. Given this, expected social welfare is:

$$\int_0^1 \int_{\frac{1}{2}}^1 \theta \tfrac{1}{4} d\theta di + \tfrac{1}{2} - \tfrac{1}{32} = \tfrac{3}{32} + + \tfrac{1}{2} - \tfrac{1}{32} = \tfrac{9}{16}$$

Note this mirrors social welfare under (CP) in the small AI model case.

We can now compare expected social welfare under the two traditional rights regimes. Social welfare under (NC) will exceed that under (CP) if:

$$\tfrac{1}{16}(9 + 2\gamma - \mathbb{E}[\rho_i(1)](2\gamma + \mathbb{E}[\rho_i(1)])) > \tfrac{9}{16} \Leftrightarrow 2\gamma > \frac{\mathbb{E}[\rho_i(1)]^2}{1 - \mathbb{E}[\rho_i(1)]}$$

6.4 Proof of Proposition 4

Under the (FU) mechanism, note that $\hat{p}_{OC}$ and $\hat{p}_{AI}$ will have the same pricing structure as under (NC) adjusted for their respective quality changes. This is because each OC knows its realised $\rho_i(1)$ when setting its price. Importantly, this implies that half of each content market will purchase from their respective OC while the other half will potentially consume content by purchasing AI products. This implies that $\hat{x}_{i,FU}$ and $\hat{y}_{FU}$ are determined by the following first-order conditions:

$$F(\Gamma)\tfrac{1}{4}\big(1-\mathbb{E}[\rho(1)|\Gamma]\big)+(1-F(\Gamma))\tfrac{1}{4}=\frac{\partial c_{OC}(x)}{\partial x}$$

$$\frac{\partial u(y)}{\partial y}=\frac{\partial c_{AI}(y,\int \hat{x}_i di)}{\partial y}$$

Comparing first-order conditions, note that each $\hat{x}_{i,CP}>\hat{x}_{i,FU}>\hat{x}_{i,NC}$ while $\hat{y}_{FU}>\hat{y}_{NC}>\hat{y}_{CP}$. These conditions assume that AI finds it feasible to remain in operation. As assumed in the proposition, AI's feasibility condition is:

$$u(\hat{y}_{FU})-c_{AI}(\hat{y}_{FU},\hat{x}_{i,FU})\geq(1-F(\Gamma))\int_{\Gamma}^{1}\tfrac{1}{4}\rho(1)\hat{x}_{i,FU}dF(\rho(1))$$

Using our earlier specific functional forms $\hat{x}_{NC}=\frac{1}{4}(1-\mathbb{E}[\rho(1)|\Gamma])$ for all i and $\hat{y}_{NC}=1+\gamma\int\hat{x}_i di=1+\gamma\frac{1}{4}(1-\mathbb{E}[\rho(1)|\Gamma])$. Given this, expected social welfare is:

$$\begin{aligned}
&\int_0^1\left(\int_{\frac{1}{2}}^1\theta\tfrac{1}{4}(1-\mathbb{E}[\rho(1)|\Gamma])d\theta+\tfrac{1}{4}(1-\mathbb{E}[\rho(1)|\Gamma])\int_0^{\frac{1}{2}}\int_0^1\rho_i(1)\theta dF(\rho_i(1))d\theta\right)di\\
&+\tfrac{1}{2}(1+\gamma\tfrac{1}{4}(1-\mathbb{E}[\rho(1)|\Gamma]))-\tfrac{1}{2}(\tfrac{1}{4}(1-\mathbb{E}[\rho(1)|\Gamma]))^2\\
&=\tfrac{1}{32}\big(3+\mathbb{E}[\rho(1)]\big)(1-\mathbb{E}[\rho(1)|\Gamma])+\tfrac{1}{2}(1+\gamma\tfrac{1}{4}(1-\mathbb{E}[\rho(1)|\Gamma]))-\tfrac{1}{2}(\tfrac{1}{4}(1-\mathbb{E}[\rho(1)|\Gamma]))^2\\
&=\tfrac{1}{32}(18+(4\gamma+\mathbb{E}[\rho(1)])(1-\mathbb{E}[\rho(1)|\Gamma])-\mathbb{E}[\rho(1)|\Gamma](1+\mathbb{E}[\rho(1)|\Gamma]))
\end{aligned}$$

It is clear that this exceeds expected social welfare under (NC) as the consumer consumption of original content is the same as under (NC) while both content and AI training quality are higher. Expected social welfare under (FU) exceeds expected social welfare under (CP) if:

$$\gamma\geq\frac{\mathbb{E}[\rho(1)|\Gamma]^2+\mathbb{E}[\rho(1)|\Gamma]-\mathbb{E}[\rho(1)](1-\mathbb{E}[\rho(1)|\Gamma])}{4(1-\mathbb{E}[\rho(1)|\Gamma])}$$

Thus, the domain under which (CP) is optimal is reduced when (FU) is a possible option as $\mathbb{E}[\rho(1)]>\mathbb{E}[\rho(1)|\Gamma]$.

References

Besen, Stanley M, "An Economic Analysis of the Artificial Intelligence-Copyright Nexus," *TechREG Chronicle*, 2023, pp. 3–9.

Gans, Joshua S, "Remix rights and negotiations over the use of copy-protected works," *International Journal of Industrial Organization*, 2015, *41*, 76–83.

Green, Jerry R and Suzanne Scotchmer, "On the division of profit in sequential innovation," *Rand Journal of Economics*, 1995, pp. 20–33.

Landes, William M and Richard A Posner, "An economic analysis of copyright law," *The Journal of Legal Studies*, 1989, *18* (2), 325–363.

Marcus, Gary and Reid Southen, "Generative AI Has a Visual Plagiarism Problem," *IEEE Spectrum*, 2024.

Tänzer, Michael, Sebastian Ruder, and Marek Rei, "Memorisation versus generalisation in pre-trained language models," *arXiv preprint arXiv:2105.00828*, 2021.