Market Design for AI: Beyond the Copyright Binary

Maryam Farboodi* Negin Golrezaei† Sepehr Shahshahani‡

January 15, 2026

Abstract

How can we design a market of human-generated content for use by generative artificial intelligence (AI) that both enables technological progress and preserves individuals' incentive to create high-quality content? Existing answers to this important question take two strong forms: Some argue that the technological-progress imperative requires that the use of copyrighted content in model training qualify as a noninfringing fair use (the "free for all" model) while others argue that strong intellectual property rights are necessary to preserve human creative incentives (the "strong property rights" model). We show, using a game-theoretic model, that both answers fall short. One of our central findings is that not only the free-for-all model but also, counterintuitively, the strong property rights regime lead to underpowered creative incentives. This result inverts common wisdom on the incentive-boosting effect of intellectual property rights. We show, further, that a market characterized by strong individual property rights will privilege homogenized content over creative contributions. Dynamically, the degradation and homogenization of human content eventually undermines AI performance. We draw on the static and dynamic market failures we uncover to propose a novel design of a data market for AI, incorporating a data intermediary who can help perserve creative incentives while enabling technological progress.

1 Introduction

Recent breakthroughs in generative artificial intelligence (AI) are laden with both promise and peril. By enabling the fast, cheap delivery of useful content, AI models have improved work and life for millions of people. But AI's ability to generate this useful content depends on a rich reservoir of human-created content on which to train the AI models. To date, AI firms have acquired this content by scraping it from online sources—largely without human creators' consent, compensation, or credit. While this practice has enabled rapid AI progress, it undermines human creators' incentives to create. We are already seeing warning signs: Stack Overflow's traffic declined

*MIT Sloan School of Management. farboodi@mit.edu.
†MIT Sloan School of Management. golrezae@mit.edu.
‡Washington University School of Law. sepehrs@wustl.edu.

significantly after the rise of Large Language Models (LLMs) like ChatGPT, as users consumed AI answers rather than contributing to producing them. And various artists and other content creators have sued AI firms contending that their practices have deprived them of the rewards of their creative work.

These trends point to a clear misalignment between the incentives of AI developers and human creators, and they give rise to the central research question of this project: How can we design a market of human-generated content for AI that both enables technological progress and preserves individuals' incentive to create high-quality content?

This is among the foremost policy challenges of our times, implicating in its narrow berth important questions of innovation and technology policy and, more broadly, fundamental concerns about how to preserve human creative flourishing. But the tools we have for answering the question come from a decades-old Copyright Act that was not designed with anything like AI technology in mind. As such, the current debate is trapped in a legal binary: AI harvesting is either fair use (what we call a "free for all" system) or copyright infringement (what we call a "strict property rights" regime). This paper argues that both extremes are destined to fail and points instead to a sustainable "third way" via data intermediaries.

Our project proceeds in two thrusts. First, we develop and solve a formal model capturing the interaction of AI firms with human content creators to better understand the economic forces at play and the properties of existing proposals. Second, drawing on our findings in the first part, we design and propose an architecture for a market of human-created content for AI use that preserves human creative incentives without shortchanging technological progress.

In the first thrust, we show, consistent with some commentators' concerns, that a free-for-all system will not work because it leads to human under-investment in high-quality content creation. But we show that the opposite solution—strict property rights—is also doomed to fail. Specifically, we identify *two short-run market failures and a long-run market failure*. First, contrary to conventional wisdom about the effects of intellectual property (IP) rights, we show that *strong individualistic IP rights result in under-powered creative incentives* because of correlation among different human creators' content and the market power of AI firms. Second, we show that AI's statistical preference for "representative" data creates an *originality penalty*, incentivizing the production of homogenized content. Lastly, beyond the tradeoff between human creative incentives and technological progress in the short term, our model explores long-run dynamics, including the risk of *model collapse* whereby the degradation and homogenization of human inputs eventually undermines AI's own capabilities.

In the second thrust, we draw on our findings about the failures of the current system and existing proposals to come up with a better solution to our central policy challenge. Our tentative solution takes the form of a *data intermediary* who helps coordinate individual content owners' decisions and apportion royalties among them. The goal of this thrust is to develop design principles for a market intermediary that overcomes the market failures identified in the first thrust: halting the

race to the bottom in the price of content that undermines individual incentives and maintaining the diversity of content required for human flourishing and long-term innovation, and thereby sustaining long-run model performance.

The paper is organized as follows. The next section explains the legal-policy background and situates our contribution in the literature. Section 3 develops and solves a model that uncovers the shortcoming of the existing legal regime and commonly proposed solutions, laying bare both static and dynamic market failures. Building on these results, section 4 sketches the contours of an institutional design that can overcome these shortcomings and outperform commonly proposed solutions. The Appendix furnishes formal statements and proofs. The analysis is quite preliminary in parts—including the discussion of dynamic market failure and, especially, our proposed market design—but we have endeavored in those parts to sketch the outlines of argument and projected results.

2 Background and Literature

The question of how best to resolve AI's disruption of the market for creating and consuming informational content is hotly debated not just in academic literature but also in courts. The most consequential legal questions affecting the future of AI have been raised in several lawsuits filed by content owners against AI firms alleging copyright infringement.[1] To a layperson, the juxtaposition of copyright with the cutting edge of technology may appear unusual—after all, copyright is thought to be about artistic creativity, whereas the high-technology context might seem better suited to patent law or other regulatory regimes—but in fact copyright law has long been at the forefront of regulating new technologies in the United States. From the piano roll to cable to VCRs to peer-to-peer file sharing to online video streaming and beyond, the first steps in deciding the fate of emerging technologies have often been taken by courts in copyright cases (Shahshahani, 2018). To appreciate the reasons for this, it is essential to review the fundamental structure of American IP laws, of which copyright is a central part.

The goal of American IP laws, in the words of the U.S. Constitution, is "to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries" (U.S. Const. art. I, § 8, cl. 8). The reason such an "exclusive right"—in other words, a temporary monopoly—has been considered necessary to achieve the objective of progress in arts and science traces to the "public good" nature of intangible products of the mind: Unlike tangible goods (say, a chair), intangible products are "nonrivalrous"—one person's use does not diminish another person's use—and "nonexclusive"—the use of an idea once disclosed is hard to limit to the first recipient. As such, the theory of American IP laws is that exclusive rights are necessary to prevent freeriding on others' creativity and maintain incentives

[1] A useful list appears at https://chatgptiseatingtheworld.com/2024/08/27/master-list-of-lawsuits-v-ai-chatgpt-openai-microsoft-meta-midjourney-other-ai-cos/.

to create. But the theory also acknowledges that these exclusive rights come at a cost: Like any monopoly, IP rights raise the cost of a work protected by IP and make it more difficult for users to access. These include not only costs to end users, such as a person who wants to read a copyrighted novel or take a patented medicine, but also—of special importance in the AI context—costs to follow-on creators who want to use existing content as building blocks for creating new content (Freilich and Shahshahani, 2023). Balancing creative incentives against access costs—the *incentives-access tradeoff*—is thus at the core of American intellectual property law. This has been the dominant conceptual framework for understanding IP in economics for a long time (e.g., Plant, 1934; Machlup, 1958; Arrow, 1962; Nordhaus, 1969) and in US law for an even longer time (e.g., *Wheaton v. Peters*, 33 U.S. 591, 657–58, 661 (1834); *Kendall v. Winsor*, 62 U.S. 322, 327–29 (1858)); *Sears, Roebuck & Co. v. Stiffel Co.*, 376 U.S. 225, 229–31 (1964); *Twentieth Century Music Corp. v. Aiken*, 422 U.S. 151, 156 (1975)).

The incentives-versus-access paradigm helps explain the current array of scholarly opinion on the question of whether the unauthorized use of copyrighted content in training AI models should be permitted.[2] On one side, scholars have argued that AI firms' unrestricted use of copyrighted works in training AI models undermines individuals' incentives to create high-quality content, so robust IP rights are needed to preserve creative incentives and promote the constitutional goal of artistic and scientific progress (Opderbeck, 2024; Barnett, forthcoming 2026). We call this position the "full property rights" model. On the other side, scholars have argued that unrestricted access to data is essential to innovation, concluding that the use of copyrighted works in AI training constitutes "fair use"[3] and is not a copyright infringement (Sag, 2019; Lemley and Casey, 2021; Lee, 2025). We call this the "free for all" model.

Our first insight is that these polar ways of thinking about the problem fall short. It is easy to see why the free-for-all model will not work: If content can be taken without rewarding its creator, people will have little incentive to invest resources into creating content, so the amount and quality of human-created content will decrease. That is the very freerider problem that American copyright and patent laws are designed to ameliorate. The problem will be especially pronounced as consumers

[2]That is not the only copyright-law question implicated by AI. Grossly simplified, there are least three major legal questions: (1) Does the use of copyrighted material in a AI system's training set infringe the copyright in the material? (2) Does the output of a AI model infringe the copyright in works in its training set or other works? What, in this connection, is the content and the sphere of application of the substantial-similarity standard, and does it differ from the standard when considering non-AI cases of copyright infringement? (3) Is output produced by someone using AI copyrightable? If so, who is the author (initial owner)? It is the first of these questions that is most directly applicable to this project, though the second question also potentially comes into play.

[3]Fair use is a longstanding judge-made doctrine, now codified in the Copyright Act, 17 U.S.C. § 107, which holds that certain uses of copyrighted works, though otherwise impinging on the exclusive rights of the copyright owner, qualify as noninfringing in light of the purpose and character of the use, the nature of the copyrighted work, the amount and substantiality of the material used, and the effect of the use on the market for the original work. The two main theoretical justifications for the fair use doctrine are that it cures market failure in situations where the copyright owner is unlikely to license a valuable use (Gordon, 1982) and that it encourages "transformative" uses whereby a second-generation creator uses a copyrighted work as raw material in the creation of new information, insights, or aesthetics (Leval, 1990). While these academic justifications are useful and influential, the doctrine is highly fact-dependent and case results cannot be confidently predicted based only on academic theory.

increasingly turn to AI platforms, rather than the original source, to get the content they want.

Admittedly, this analysis does not do full justice to the concept of fair use, which is flexible and capable of adaptation to a variety of factual circumstances. The doctrine does not require that *all* uses of copyrighted work for training AI models be deemed noninfringing (though some commentators have advocated just that); depending on the circumstances, some uses may be shielded by fair use while others are held to be copyright infringement. For example, in a case involving Anthropic, the firm that developed the large language model Claude, a court recently held that the unauthorized copying and use of copyrighted books to train the model qualified as fair use with respect to books that Anthropic purchased but not with respect to books that it downloaded for free from "pirate" websites (*Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1033 (N.D. Cal. 2025)). Whether particular AI uses qualify as fair use could thus be determined case by case in light of particular factual circumstances, as the recent Copyright Office Report on artificial intelligence predicts (Register of Copyrights, 2025, 74). While such flexibility can be a virtue, it also breeds uncertainty and unpredictability, a longstanding criticism of fair use doctrine (see Shahshahani (2015) for a critical survey). For example, on facts very similar to the *Anthropic* case, a different court held that Meta's use of copyrighted books downloaded from pirate websites to train the Llama models *was* fair use, but took pains to point out that the decision was based on the record before it and could well have gone the other way if the plaintiffs had developed a better factual record (*Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1036–37 (N.D. Cal. 2025)). But if the legality of use will depend on each case's particular facts and circumstances, or on the judge presiding over the case, then both AI developers and individual creators are left without much *ex ante* guidance about whether AI firms can freely take human creators' content. So, whereas a blanket fair use privilege results in under-powered creative incentives, a highly fact-specific doctrine leads to uncertainty and potential chilling effects, particularly for risk-averse tech firms and content creators.

What about the full property rights model? Proponents of this model rely on the Coase Theorem (Coase, 1960) to argue that if property rights are clearly assigned to creators, then creators and AI firms will negotiate a price and efficient data transfers or licenses will occur. The standard counterargument in the literature is that *transaction costs* of different flavors—the logistical difficulty of an AI company negotiating with millions of individual creators, overlapping or unclear ownership of IP rights, cultural and institutional barriers to mutual understanding, divergent estimates of value, and others—may stand in the way of an efficient bargain (Ostrom, 1990; Heller, 1998; Heller and Eisenberg, 1998).

That is *not* why the full property rights model fails in our theoretical framework. We show, more pessimistically, that the model will fail even in the absence of garden-variety transaction costs—that is, even if the parties can reach an agreement to sell or license human-created content to AI firms. The reason behind this failure is *correlation* between different human creators' content. Because each individual's content is similar to other individuals' content in that domain, an individual selling her own content is also effectively selling "part of" other creators' content to the AI firm.

This dynamic pushes down the price of content to such a level that many individuals will not find it worthwhile to invest in creating content. So the property rights model breaks down not because it fails the technological-progress objective but, counterintuitively, because it fails the creative-incentives objective. This result inverts common wisdom about the effect of strong intellectual property rights.

In deriving it, we draw on recent work in information economics, particularly Acemoglu et al. (2022) and Bergemann et al. (2019), which demonstrate that when individual data points are correlated, selling one's data generates a negative externality by revealing information about others, leading to "information leakage" that depresses prices far below the social optimum. We extend this literature in several ways. Whereas Acemoglu et al. (2022) focus on consumer privacy, we shift attention to the production of creative works, which changes the action space of agents and complicates the analysis. This richer analysis allows us to investigate the differential effects of AI on human creative works with different levels of novelty or originality, demonstrating the "originality penalty" discussed above. In addition, we move beyond the static context of existing models to explore the dynamic effects of AI interaction with potential content creators, theoretically deriving the "model collapse" phenomenon predicted by computer scientists (Shumailov et al., 2024) whereby the impoverishment of human input causes AI model quality to degrade.

Our analysis of market failure in the presence of IP draws on and contributes to an extensive economics literature on property rights. Building on the seminal contributions of Coase (1960) and Williamson (1979), this literature has largely focused on transaction costs, and options-to-own have been proposed as potential solutions (Che, 2006; Segal and Whinston, 2016). As we are interested in efficient allocation of property rights in the context of data, we specialize the model in thinking about data as a productive asset that can be traded in a market (Farboodi et al., 2019; Farboodi and Veldkamp, 2022, 2023).

For our data market design solution, we draw on an extensive literature in economics that analyzes the role of intermediaries in different markets (Rubinstein and Wolinsky, 1987; Farboodi et al., 2023). And we build on the literature on auction mechanisms to design payment structures to achieve desired market outcomes (Golrezaei et al., 2021a,b). Finally, there is an active literature in economics and operations research on market design, starting with the foundational work of Gale and Shapley (1962). We use insights from this literature on dynamic incentive-aware learning to study how creators strategically alter their content production in response to AI pricing schemes (Golrezaei et al., 2019, 2023).

3 Model

We present a game-theoretic model to uncover the equilibrium properties of the free-for-all and property-rights models, then we extend that model to account for dynamic realities like model collapse, and finally we propose a market-design solution based on our theoretical findings. Full

mathematical formulation and arguments are provided in the Appendix.

We begin by analyzing a static market with multiple content creators and one AI developer. Under the classic copyright binary, either the unauthorized use of copyrighted works to train AI models is copyright infringement—in which case the AI firm is required to obtain authorization from each copyright owner—or the use is fully shielded from copyright liability under fair use. We first show that, consistent with common intuition, a free-for-all system fails to provide incentives for human content creation. Less intuitively, we show that the strict property rights regime leads to two distinct market failures in the short run. First, as long as human-created contents are not perfectly uncorrelated, assigning full property rights to content creators leads to under-investment in content creation, and the inefficiency is amplified as content correlation increases. Second, the role of AI as a content aggregator effectively disincentivizes the production of highly original human content, leading to a proliferation of homogenized content. We demonstrate these short-run market failures in the next section and explore their long-run consequences in the section that follows.

3.1 Short-Run Market Failures

Baseline Model. We formalize the interaction between human creators and a AI developer as a market for information with costly production of correlated data. The economy centers on an unobserved state variable $X \in \mathbb{R}$, representing the best or most useful response to a given user query. N representative agents ("creators"), indexed by $i = 1, 2, \ldots, N$, create content. We model the content created by agent i as a noisy signal s_i of X, namely

$$s_i = X + \varepsilon_i.$$

As such, $\varepsilon_i \sim N(0, k_i^{-1})$ denotes the deviation of an individual creator's content from X. To produce this content, agent i chooses a *precision* level k_i, incurring a convex cost $C(k_i)$. In our framework, precision captures the quality of the content. This setup captures a fundamental tradeoff for each human creator: high-quality content requires costly effort, but low-quality content has less value. In what follows, we will refer to $C(\cdot)$ as the "cost of precision," which is the same as the cost of an individual creator's effort to produce high-quality content.

A defining feature of our framework is the correlation structure of signals. We assume the noise terms in the agents' signals are jointly normally distributed with correlation coefficients $\rho_{i,j} \in [0, 1)$. The correlation coefficient captures the degree of substitutability between the content generated by different creators: as $\rho_{i,j} \to 1$, creators produce redundant information; as $\rho_{i,j} \to 0$, they provide distinct, idiosyncratic insights. For ease of exposition, in the main text we assume the same correlation coefficient for all pairs of creators, namely $\rho_{i,j} = \rho$ for all i and j.

The AI developer acts as a monopolist "aggregator". It purchases the human creators' signals to construct a Best Linear Unbiased Estimator (BLUE) of X. The firm's production technology is defined by the *effective precision* of this estimator, denoted K, with the interpretation that (a

monotonic transformation of) the effective precision functions as the "total factor productivity" of the AI model.[4] Practically speaking, a higher K means that the content generated by the AI model in response to a user query is more useful to the user. The effective precision of the AI model's estimator is increasing in human content precision (k_i) but decreasing in correlation of human content (ρ), reflecting the diminishing marginal returns of redundant data.

To incentivize the production of high-quality data while balancing the purchase cost, the AI aggregator offers a price p_i per unit of precision to creator i. The creators react by deciding their individual precision level k_i at cost $C(k_i)$. The creators are identical and fully symmetric in the simplified model of this draft—same cost function and same pairwise correlation coefficient. As such, we focus on the symmetric equilibrium (common price per unit of precision), namely $p_i = p$ for all i.[5] We plan to extend the framework and incorporate heterogeneous creators in order to study price discrimination by AI.

In this framework, we contrast the *equilibrium* outcome with the *social planner* (or *first best*) solution. The equilibrium represents the outcome of a strategic interaction between the AI developer and the human creators each pursuing their own interests in light of others' strategies, under either the free-for-all or strict-IP regimes; the social optimum represents the outcome that would result if a benevolent social planner dictated the choices of all economic agents. We refer to any variable x that corresponds to equilibrium and social planner solutions as x^* and x^{sp}, respectively, as customary in economics literature.

When a benevolent social planner maximizes total welfare (the quality of the AI output minus the cost of producing content), our model shows that the resulting individual and aggregate precision is high. That is because the social planner values the distinct contribution of every creator. Moreover, it is straightforward to see that a free-for-all regime leads to under-investment by content creators. The free-for-all regime is equivalent to a market where a zero per-unit-precision price, $p = 0$, is enforced in equilibrium. This implies that creators will not be compensated for any effort they would invest in content creation. It follows from the break-even constraint of the creators that there will be no content creation in equilibrium.

The more interesting insight is the failure of individual property rights, which are traditionally thought to work well in the absence of transaction costs. Comparing the equilibrium and social optimum enables us to identify two different forms of market failure in the short run, illustrated below through two observations.

Observation 1 (**Intellectual property rights do not provide effective creative incentives).** *Assume creators and the AI firm can bargain effortlessly, i.e., there are no transaction costs. Even so, the market cannot implement the optimal level of investment in content creation, which in turn leads to a degradation in the quality of AI's aggregated output. Specifically, in a symmetric*

[4] As signal errors are correlated, the Generalized Least Squares (GLS) estimator is the BLUE. See Appendix A for detailed derivation of K.

[5] In Appendix A, we solve the model with general correlation coefficients $\{\rho_{i,j}\}_{i\neq j}$ and heterogeneous prices $\{p_i\}_{i=1,\cdots,N}$. Observation 1 continues to hold.

equilibrium, the quality (precision) of the output produced in the market is exactly half the socially optimal level:

$$K^* = \frac{1}{2}K^{sp}$$

The reason for under-investment in content creation is tewofold: the AI market is not fully competitive and the creators' content is correlated. In this market, the AI firm is the price setter while the creators are price takers. As such, the correlation between different creators' content implies information leakage, which is not internalized by each individual creator. In other words, as different creators' contents are correlated, each creator by selling her informational content is also effectively selling part of other creators' information, so the AI firm is unwilling to highly compensate each creator. This, in turn, reduces the incentive of creators to invest. The social planner, by contrast, internalizes this negative externality, which leads to a higher level of investment by creators. We show that equilibrium aggregate precision is exactly half of that in the first best allocation.

This leads to a further insight about the role of correlation.[6] We show that as the correlation between creators (ρ) increases, the equilibrium price paid to content creators falls. This formalizes a *substitution effect*: as AI learns the general patterns of human creativity (high correlation), it devalues the very humans it relies on. The market equilibrium is an inefficient trap where humans under-invest in creation because they cannot capture the value of their marginal contribution due to information leakage.

Together, the halved innovation and the substitution effect summarize the first market failure we uncover: Contrary to conventional wisdom, robust individual IP rights fail to provide sufficient creative incentives when the market includes an AI firm who buys and aggregates correlated human-created content.

We also uncover a second market failure—an *originality penalty.* Bringing out this point requires adjusting the model to make the signal structure a bit more complicated.

Model Adjustment. We now generalize the deviation of the individual creator's content (s_i) from the state (X). Recall that in the baseline model, this deviation consists of solely the individual creator's error, the precision of which can be reduced by the creator through costly effort. To incorporate a diverse array of human creators with varying degrees of conformity to conventional consensus or crowd opinion, we allow the above deviation, $s_i - X$, to have two different components: first, a *common component* that is shared among all creators but weighted differently by different creators; second, an *idiosyncratic component* whose precision is an increasing function of the individual agent's effort to produce high-quality content. The common component represents a common bias, shared to varying degrees among all content creators, while the idiosyncratic component reflects content creators' individual effort to produce high-quality content. Specifically, let

[6]For a general variance-covariance matrix, we refer to an increase in "correlation" as a uniform increase in pairwise correlations $\rho_{i,j}$ across creators. Thus, our comparative statics can be interpreted as applying to the symmetric environment in which $\rho_{i,j} = \rho$ for all $i \neq j$, or as an increase in the average level of correlation across creators. Our results extend naturally to heterogeneous correlation structures.

the signal of agent i be

$$s_i = X + \beta_i \eta + \varepsilon_i$$

where η is the common factor (which may or may not be zero-mean), β_i represents how strongly agent i loads on the "common factor," and ε_i stands for the idiosyncratic noise of each agent. High β_i means that individual i's content highly conforms to the crowd opinion.[7]

We suggest two (mutually compatible) interpretations for the $\beta_i \eta$ term. In the first interpretation, η represents a general *trend*, such as a stylistic convention, so a large β_i denotes agent i's conformity to prevailing conventions whereas a low (near-zero) β_i signifies agent i's willingness to buck the trend and strike out on her own, to march to the beat of a different drummer. That is the sense in which we characterize high-β creators as "rank and file" or "run of the mill" and low-β creators as highly "original" or "creative." And the "originality penalty" is the greater quality (effort) decline suffered by highly original creators, as compared to run-of-the-mill creators, when we move from the social optimum to the AI market equilibrium.

In the second interpretation, η represents a common *bias*, such as socially prevalent prejudice against a group or practice or thought, and β_i measures the extent to which agent i shares that bias. Then, the originality penalty can be interpreted as *bias amplification* by AI—not only in the general sense that AI-generated content is more biased (less accurate) than is socially optimal, but in the specific sense that the equilibrium quality reduction (compared to the social optimum) is *greater* for content with low common-bias than for content with high common-bias.

Depending on the context, one or the other interpretation may be more natural. As noted, the two interpretations are entirely compatible because, in the current model specification, the trend is biased (the addition of η can never decrease the deviation of the creators' signals from the true state, $s_i - X$), so higher creativity implies higher quality—all else (i.e., effort) being equal.[8]

In this modified setting, the information aggregator's objective is no longer just to incentivize greater effort to produce high-quality content, but also to avoid the common bias present (to varying degrees) in all creators' content (subject to minimizing the price paid). Intuitively, it would appear that this richer setting would prompt the AI firm to offer a better (or less bad) price to highly original content. More precisely, even allowing for the fact that, due to information leakage, the AI firm "lowballs" human creators by offering a price that does not incentivize sufficient investment in quality, one would expect that the AI firm would lowball original content creators less than it lowballs run-of-the-mill creators in order to get the benefit of the former's individuated (non-redundant) signals. But it turns out that the equilibrium outcome is precisely the opposite: In the AI market equilibrium, relative to the social optimum, the reduction in quality of content is

[7] As explained more fully in Appendix B, the baseline model is a special case of this general model.

[8] In principle, it's possible to conceptualize a notion of originality or creativity (or a residual of it) that is independent of accuracy or quality. But, without getting into the details of how such a notion could be mathematically operationalized, our conjecture is that if there is an equilibrium penalty for the kind of originality that contributes to quality (as we have shown there is), then *a fortiori* there must also be an equilibrium penalty for the kind of originality that does not contribute to quality.

greater for original creators than for run-of-the-mill creators. Observation 2 states this result more precisely.

Observation 2 (**AI Penalizes Originality**). *Comparing the equilibrium outcome to the social optimum, we obtain two distinct outcomes for highly creative agents (those with low β_i) versus rank-and-file agents (those with high β_i):*

- ***"Rank-and-file" creators*** *($\beta_i \gg 0$). These creators have almost unreduced precision under the decentralized equilibrium relative to the social optimum, i.e., $k_i^* \approx k_i^{sp}$.*
- ***"Original" creators*** *($\beta_i \approx 0$). These creators face a severe reduction in precision—and thus price—when moving from the social optimum to the decentralized equilibrium. Specifically, unless their signal precision k_i is exponentially higher than the crowd, they will almost halve their precision, i.e., $k_i^* \approx \frac{1}{2} k_i^{sp}$.*

Note that Observation 2 is *not* saying that the AI firm pays a lower per-unit-of-precision price for original creators' content than for run-of-the-mill creators' content. Rather, the meaning of the "originality penalty" is that when we move from the social optimum to the AI market equilibrium, original creators face a greater drop in incentives to invest in effort compared to run-of-the-mill creators. In other words, AI penalizes originality more than it penalizes run-of-the-mill content by disincentivizing the original creators more. Given the redundancy (high correlation) of rank-and-file creators' output, eliciting a great deal of effort from them would not be worthwhile even in the social optimum: relative to the low optimal level of effort, the AI firm's strategic underpricing does not result in a great deal of reduced effort and reduced precision. Original creators, by contrast, should optimally exert great effort, but the AI firm's underpricing greatly reduces their incentive to do so and, thus, the precision of their content.

Observation 2 provides a theoretical microfoundation for the observation that RLHF (Reinforcement Learning from Human Feedback) tends to steer models toward safe, homogenized outputs. It also provides a microfoundation for—and a more precise statement of—complaints about the prevalence of "AI slop" (Tullis, 2025). What is more, the exclusion of original creators' potential contributions from AI training sets not only deprives the world of the value of this content in the short run but also degrades AI models in the long run: The AI firm sows the seeds of its own decay by stifling human creativity. The next section explains this long-term dynamic.

3.2 Long-Run Market Failure

The foregoing analysis shows that standard IP rights cannot solve the AI data crisis in a simple, static world. We plan to generalize these insights by incorporating dynamics into the economy to study content creators' repeated investments and their market interactions with the AI developer.

Dynamics and the Economics of "Model Collapse". The models in 3.1 are static, but the AI training pipeline is recursive: modern LLMs are trained on web-scraped data, which is increasingly

populated by the output of previous LLMs. This creates a "pollution of the commons" problem leading to model collapse—a degenerative process where learning from synthetic data causes the AI model to lose variance and converge to the mean (Shumailov et al., 2024).

We challenge the standard economic intuition that the market will self-correct. One might assume that as high-quality data becomes scarce, its price will rise, inducing unique creators to return. However, based on the bias identified in Observations 1-2, the AI developer's short-term demand is driven by representativeness (high ρ). Therefore, the market-clearing price at time t (or p_t) will be set by the marginal "average" creator. We hypothesize that this price will be insufficient to cover the costs of "unique" creators, causing them to exit permanently.[9] Creators' decision in turn impacts the quality of AI output and through that the price p_t, which then feeds back to creators' decision and requires solving a fixed point problem.

Model Adjustment. We introduce a time dimension t. We model the evolution of the AI model's effective precision K_t (e.g., the inverse of uncertainty). The knowledge stock depreciates over time and is replenished by new inputs:

$$K_t = (1-\delta)K_{t-1} + \underbrace{\mathcal{I}_{Human}(p_t)}_{\text{Fresh Signal}} + \underbrace{\mathcal{I}_{Syn}(K_{t-1})}_{\text{Synthetic Signal}}$$

Here, $\mathcal{I}$ represents the informational value added by new data. Crucially, synthetic data $\mathcal{I}_{Syn}$ is derived from the previous model, so it adds no new variance (it is collinear with K_{t-1}). Therefore, the system relies entirely on $\mathcal{I}_{Human}$ to counteract depreciation δ.

The Mechanism of Failure. We model the human input $\mathcal{I}_{Human}$ not as a homogeneous block but as a sum over heterogeneous creators who only participate if the price exceeds their cost:

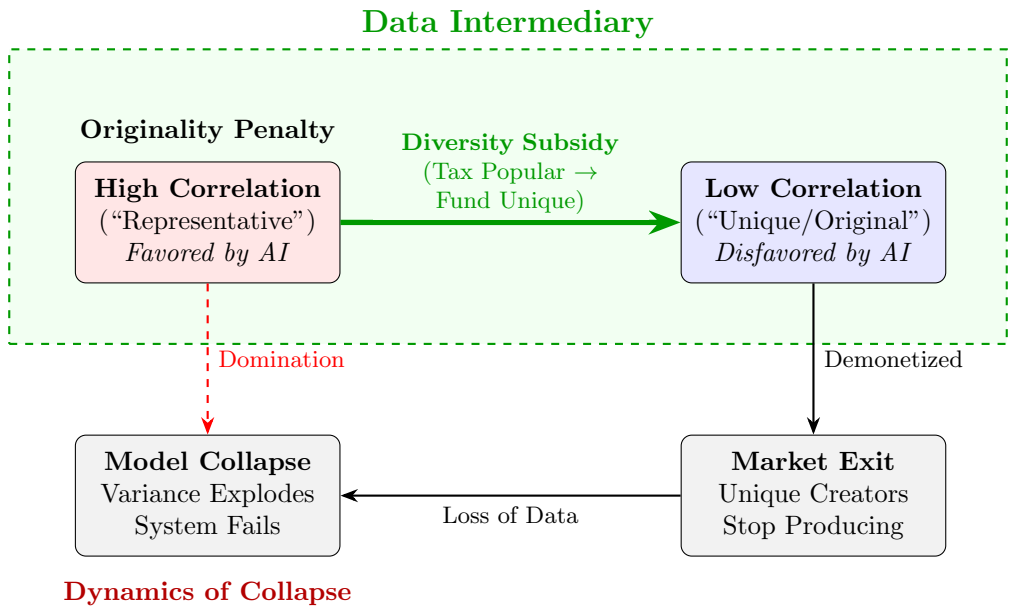
$$\mathcal{I}_{Human}(p_t) = \sum_{i=1}^{N} \mathbf{1}_{\{p_t \geq C_i\}} \cdot \mathcal{I}_i(\rho_i)$$

Our conjecture is that the equilibrium price p_t satisfies the participation constraint for low-cost/high-redundancy creators (C_{low}), but fails for unique creators ($p_t < C_{unique}$). Consequently, the "fresh signal" term loses its variance-expanding components, and K_t decays over time.

Analysis Plan. Figure 1 illustrates our research plan, organized into two interdependent thrusts. Thrust I dissects the market failures of the classic intellectual property paradigm into two static observations and their dynamic consequence. Thrust II proposes a solution.

[9] In the current version of the model, human individuals' action space does not include an option to sell their content directly to consumers, as opposed to the AI firm. Going forward, we would like to augment the model by adding the possibility that content providers can access the market directly, albeit at a "search/signaling cost." In that extension, individual creators have heterogeneous incentives with respect to selling their content to the AI firm, selling it directly to consumers, or permanently exiting the market (not producing content). The heterogeneity is driven by the creativity and quality of their content, as well as their search cost of direct market access and signaling cost.

[11] "Idea pool" in Thrust II is an adaptation to our context of the concept of "gene pool" from genetic algorithms in computer science.



The Model Logic:

1. **Thrust I:** failure of binary intellectual property paradigm
   - **Static:** AI market undervalues investment in quality and penalizes originality.
   - **Dynamic:** Without original data creation, AI model deteriorates.
2. **Thrust II:** Intermediary must transfer value from run-of-the-mill creators to creative creators to preserve the "idea pool."

Figure 1: **Framework of Research Thrusts.** Thrust I identifies the sources of static market failure in a regime of strict IP rights and demonstrates long-run model collapse. Thrust II proposes an intermediate institutional solution—a cross-subsidy mechanism to preserve data diversity.[11]

The analysis in Thrust I is not complete, and we plan to refine it. We aim to prove that the market price will remain too low. We will simulate the dynamic system to identify the *collapse threshold*: the critical proportion of unique human creators that must be retained in the market to prevent K_t from collapsing. We expect to show that a purely decentralized price mechanism consistently misses this threshold, necessitating the intervention designed in Thrust II.

4 Mechanism Design

Thrust I reveals a problem: IP rights are meant to incentivize creativity, but they underpower creative incentives, especially for highly creative content. The high correlation between human works leads to a "race to the bottom" in content pricing and a systemic "originality penalty," which leads to under-investment in content creation, especially for highly original content. These

failures create a trajectory toward "model collapse" where the AI model eventually starves itself of the novel human input it needs to remain robust.

In the second thrust, we move from diagnosis to treatment. We propose and analyze the architecture of a *data intermediary* to protect the diversity of human ideas. We aim to design a system that encourages original work and prevents content homogenization, keeping the AI ecosystem healthy for years to come.

4.1 Correcting Coordination Failure: The Collective Bargaining Unit

The first role of the intermediary is to correct the collective action problem among the creators. Because individual creators are atomic and their content is often substitutable, they suffer from coordination failure and tend to impose a negative externality on each other when facing the AI firm. We model the intermediary as a collective licensing body (analogous to ASCAP or BMI in the music industry) that aggregates the rights of a critical mass of creators. As such, the intermediary internalizes the negative externalities of data sales on behalf of the content creators.[12] We envision the functionality of the intermediary to be:

- **Internalizing Information Leakage:** Unlike an individual creator who cannot prevent their data from "leaking" information about others, the intermediary manages a group of creators. It can set a "price floor" that reflects the total social value of the training set rather than the marginal cost of a single signal.
- **The Two-Part Tariff:** To permit technological progress while retaining incentives, we propose a nonlinear pricing model. AI firms pay a low marginal fee for data *volume* (ensuring they have the scale needed for training) but a high fixed *ecosystem access fee* that is redistributed to creators to preserve the incentive to create.

In what follows, we propose a tentative design for the intermediary. We will build upon and expand this preliminary proposal as we make headway on Thrust I and characterize the static and dynamic frictions in the AI market more precisely.

We start the analysis by assuming that the intermediary is fully informed about the creators, an assumption that is unlikely to be satisfied in the real world. As such, the following market design should be viewed as a benchmark. The full-information benchmark is useful in that any inefficiencies identified here will persist in a more complex model with information asymmetries.

[12] If the intermediary is a private party, there can be delegation problems between the content creators and the intermediary. Currently, we focus on the legal design of a regulated intermediary and abstract from that second layer of frictional delegation.

4.2 The Diversity Subsidy: A Formula for Originality

To prevent model collapse, the intermediary must ensure that original creators stay in the market even when the AI firm's short-term statistical preferences would otherwise devalue them. We propose a "diversity-weighted redistribution" rule that decouples what the AI firm pays from what an individual creator receives.

Cross-Subsidization Mechanism. We define a dual-payment redistribution formula:

$$\text{Payment}_i = \underbrace{(1-\tau) \cdot R_i}_{\text{Market Share}} + \underbrace{\gamma \cdot \sigma_i \cdot C(k_i)}_{\text{Diversity Bonus}},$$

where the parameters are defined as follows:

- R_i (Usage Revenue): The market value of creator i's content, based on how frequently it is used by the AI firm.
- τ (Ecosystem Tax): A small percentage deducted from all usage revenue to fund the preservation of the creative ecosystem.
- σ_i (Originality Score): A measure of how original creator i's work is relative to the existing aggregate pool.
- $C(k_i)$ (Production Cost): The investment of effort required by the creator, ensuring the subsidy is sufficient to prevent market exit.
- ξ (Sustainability Weight): A policy variable set by the intermediary. As the risk of model collapse increases, ξ is raised to funnel more resources toward highly original creators.

This system effectively taxes "representative" content (which is high-volume but redundant) to fund "original" content (which is essential for the AI model's long-term intelligence).

4.3 Governance: Protecting the Minority from the Majority

A central challenge for the intermediary is deciding who controls the "diversity bonus." If the intermediary is governed by the creators themselves, a conflict of interest arises between the "average" majority and the "original" minority:

- **The Risk of Popularity Bias:** In a standard democratic or majoritarian system ("one person one vote"), the majority of rank-and-file creators might vote to eliminate the diversity bonus to maximize their own immediate market-share payments.
- **The Stewardship Model:** We use game theory to analyze different governance structures that can protect original creators from being outvoted. We compare *democratic voting* to

weighted stewardship where voting power is tied to the uniqueness or long-term value of content.

Our goal is to identify a governance framework that prevents shortsighted members from "eating the seed corn"—defunding the unique content the AI needs to prevent long-term model collapse.

4.4 Legal Implementation: Beyond the Binary

As discussed above, the need for our solution emerges out of the failure of the two options usually available to courts. Neither a system of blanket immunity via fair use nor a legal regime of full control over access to content by copyright owners will be sufficient to preserve creative incentives, so a workable solution must transcend the traditional copyright binary. Although our diagnosis of market failures in this context is novel, our proposal for going beyond the common copyright binary is by no means without precedent. The US Copyright Act (to say nothing of other countries' laws) contains a number of such "intermediate" regimes.[13] We plan to draw lessons from these regimes by understanding their precise legal workings, the political and economic forces that shaped them, and their policy successes and failures. Although this research is only just beginning, it's useful preliminarily to point out some comparisons between these precedents and the AI context we target.

Broadly speaking, existing intermediate regimes share two features with the present AI context. First, like the intermediate regime we propose, existing intermediate regimes were often responses to a disruptive technology that upended settled ways of doing things (Shahshahani, 2018). Second, consistent with our intuitions and preliminary results about the role of data intermediaries, these legislative compromises often involve organizations to coordinate collective action among copyright holders (or, occasionally, among other stakeholders). These organizations include the Mechanical Licensing Collective, SoundExchange, the Copyright Clearance Center, AP Images, Artists Rights Society, reprographic rights organizations, ASIP, and performance rights organizations such as ASCAP, BMI, SESAC, and Global Music Rights. Some of these organizations are created or sanctioned by legislation (or by regulations enacted pursuant to legislation), and some of them are outgrowths of "organic" coordination among copyright holders or other stakeholders.

Notwithstanding these similarities, existing intermediate regimes differ from our proposed regime in one crucial respect: Under existing regimes, the primary function of intermediaries is often to reduce transaction costs and smoothen frictions in collective action—exactly the sort of problems which one would intuit to be at the root of bargaining breakdown, but which were *not* the main mechanism driving the failure of the property-rights model in our theoretical framework (see Observation 1

[13]Examples include: 17 U.S.C. § 111 (governing secondary transmissions of broadcast programming by cable); 17 U.S.C. § 112 (governing ephemeral recordings); 17 U.S.C. § 114 (governing digital audio transmission of sound recordings); 17 U.S.C. § 115 (governing compulsory licenses to make and distribute phonorecords of nondramatic musical works); 17 U.S.C. § 118 (governing noncommercial broadcasts); 17 U.S.C. § 119 (governing secondary transmissions of distant TV station signals by satellite carriers); 17 U.S.C. § 122 (governing secondary transmissions of local TV station signals by satellite carriers).

and surrounding discussion). Given that, in our model, market failures happen even in the absence of garden-variety transaction costs, the function of intermediaries in our market design must go beyond reducing transaction costs and enabling bargaining. Hence the discussion of mechanisms to prevent information leakage and subsidize originality in sections 4.1–4.2.

Beyond highlighting these similarities and differences, researching existing intermediate regimes points up two important questions of institutional design. The first question relates to the *compulsory or voluntary* character of participation in an intermediary organization: Once the intermediary is constructed, can its benefits be realized by content creators' uncoordinated, self-interested equilibrium behavior or do collective-action problems require centralized enforcement and a legal obligation to join? This question implicates issues of collective action that have been of perennial interest in political economy at least since Olson (1965).[14]

Another important question is that of *price setting*. It is widely understood that, in many contexts, a centralized system is not as good at determining the value of assets as the decentralized information-aggregating mechanism of the market (Hayek, 1945). The way an intermediate regime, such as a compulsory-licensing scheme, sets the price at which an exchange must take place is therefore an important ingredient of the regime's success or failure. Pricing content may be easy if that content has a market outside the use for which the compulsory license is sought, in which case one can piggyback on the market price in setting the license fee. But if the primary market for the content is the use for which the compulsory license is sought, then there is no ready recourse to the market price signal. In the context of a data market for AI, pricing may not be difficult in the beginning because the data in question have had a long history of consumer demand predating the advent of AI, which history can be relied upon for pricing; however, going forward, if we anticipate the primary (direct) market for content to be as fodder for AI models, then pricing becomes more difficult. The question has at least two aspects: (1) determining the price at which the intermediary sells the content to the AI firm, (2) determining how to apportion the price received by the intermediary among different content creators. Some of our discussion of the "diversity subsidy" in section 4.2 effectively assumes that the originality and long-term value of content is *known* to the intermediary, but the forcing out of such information (assuming it is even *privately* available) is a central difficulty of centralized mechanisms with which our solution must contend.

4.5 Synthesis and Validation via Simulation

To validate the efficacy of our proposal, we will conclude with a computational simulation that pits our intermediary against the free-for-all and strict-IP regimes identified in the first thrust of the project. Our hypothesis is that only the intermediary model can simultaneously maintain a high effective precision (K_t) for the AI model and a stable creative frontier for human creators over

[14] From a legal perspective, it is important to note that the facilitation of collective action by means of intermediaries probably requires a statutory exemption from antitrust liability, and the aforementioned intermediated regimes under the Copyright Act generally contain such exemptions.

a 100-generation training cycle. This simulation will provide a proof of concept for a sustainable third way that prevents the hollowing out of human creativity in the age of generative artificial intelligence.

5 Conclusion

To come.

Appendix

A Observation 1: Under-Powered Creative Incentives

A.1 Model Primitives

We consider a market with two types of participants. There are $N \geq 2$ information producers ("creators")—indexed by $i = 1, 2, \ldots, N$—and a final producer ("GenAI aggregator"). We assume there is an unknown scalar state variable $X \in \mathbb{R}$ that all players care to know about. The creators are able to produce noisy signals about X; the aggregator buys these signals, combines them into a best estimate of X, and earns revenue from the precision of this estimate.

Assumption 1 (Signals and Precisions)**.** *Creator i chooses a* precision $k_i > 0$ *and produces a signal*

$$s_i = X + \varepsilon_i,$$

where the noise vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)'$ is jointly normal and zero-mean with covariance matrix:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \frac{1}{k_1} & \frac{\rho_{1,2}}{\sqrt{k_1 k_2}} & \cdots & \frac{\rho_{1,N}}{\sqrt{k_1 k_N}} \\ \frac{\rho_{2,1}}{\sqrt{k_2 k_1}} & \frac{1}{k_2} & \cdots & \frac{\rho_{2,N}}{\sqrt{k_2 k_N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\rho_{N,1}}{\sqrt{k_N k_1}} & \frac{\rho_{N,2}}{\sqrt{k_N k_2}} & \cdots & \frac{1}{k_N} \end{pmatrix},$$

where for each $i \neq j$, parameter $\rho_{i,j} = \rho_{j,i} \in [0, 1)$ represents the correlation between the generated content from creators i and j. Therefore, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

Assumption 2 (Cost)**.** *Each creator i—when choosing precision k_i—incurs a convex production cost $C_i(k_i) = \frac{c}{2} k_i^2$.*

The GenAI aggregates the creators' data, i.e. the noisy signals, to an estimate of X that it provides to final users. The objective of the GenAI aggregator is to minimize the variance of the estimator, subject to the estimator being unbiased, represented by

$$\begin{aligned} \max_{\hat{X}} \ & \mathbb{E}[(\hat{X} - X)^2] \\ \text{s.t.} \quad & \mathbb{E}[\hat{X}] = X \end{aligned} \tag{1}$$

As such, the optimal choice of the GenAI is the BLUE estimator.

Pricing. To acquire creators' data, the the aggregator pays creator i a per-unit-of-precision price $p_i > 0$. The aggregator has full power on deciding p_i, and all creators simultaneously decide their own k_i's after observing p_i's.

Each creator $i = 1, 2, \ldots, N$ maximizes their utility $U_i(k_i) = p_i k_i - C_i(k_i)$ (income minus production cost) while the aggregator maximizes $U_a(\boldsymbol{k}) = K(\boldsymbol{k}) - \sum_i p_i k_i$. Here, $K(\boldsymbol{k})$ is the *effective precision* of the Best Linear Unbiased Estimator (BLUE) of X given precisions $\boldsymbol{k} = (k_1, k_2, \ldots, k_N)'$.

Derivation of Effective Precision. We assume that the parameters $\{\rho_{i,j}\}_{i \neq j}$ and c are known to the aggregator before the game. Given signals $\boldsymbol{s} = (s_1, s_2, \ldots, s_N)'$, the BLUE estimator is given by the weighted sum $\hat{X} = (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{s}$ where $\mathbf{1} = (1, 1, \ldots, 1)'$. Let $\boldsymbol{D} = \mathrm{diag}(k_1, k_2, \ldots, k_N)$, and let $\boldsymbol{R}$ be the $N \times N$ matrix with diagonal entries equal to 1 (i.e., $R_{i,i} = 1$ for all i) and off-diagonal entries equal to correlations (i.e., $R_{i,j} = \rho_{i,j}$ for $i \neq j$). Then by definition,

$$\boldsymbol{\Sigma} = \boldsymbol{D}^{-1/2}\boldsymbol{R}\boldsymbol{D}^{-1/2}.$$

The precision of $\hat{X}$ is therefore given as

$$K(\boldsymbol{k}) = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1} = \mathbf{1}'(\boldsymbol{D}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{D}^{1/2})\mathbf{1}.$$

A key property that we shall exploit is the *homogeneity* of K: For any constant $\alpha > 0$,

$$K(\alpha\boldsymbol{k}) = \mathbf{1}'(\sqrt{\alpha}\boldsymbol{D}^{1/2}\boldsymbol{R}^{-1}\sqrt{\alpha}\boldsymbol{D}^{1/2})\mathbf{1} = \alpha \cdot \mathbf{1}'(\boldsymbol{D}^{1/2}\boldsymbol{R}^{-1}\boldsymbol{D}^{1/2})\mathbf{1} = \alpha \cdot K(\boldsymbol{k}).$$

This further reveals that for any $\alpha > 0$ we have $\nabla K(\alpha\boldsymbol{k}) = \nabla K(\boldsymbol{k})$.

A.2 Social Optimum vs. Decentralized Equilibrium

We now compare the Social Optimum against the decentralized Equilibrium.

Social Optimum. The social planner maximizes social welfare

$$W(\boldsymbol{k}) = \sum_{i=1}^{N} U_i(k_i) + U_a(\boldsymbol{k}) = K(\boldsymbol{k}) - \sum_{i=1}^{N} \frac{c}{2} k_i^2.$$

The First-Order Condition (FOC) then gives the following condition for the Social Optimum:

$$\nabla K(\boldsymbol{k}^{sp}) = c \cdot \boldsymbol{k}^{sp}.$$

Decentralized Equilibrium. The monopolist aggregator sets prices $\boldsymbol{p} = (p_1, p_2, \ldots, p_N)'$ to maximize profit. For each creator i, their FOC reveals the best precision $k_i = p_i/c$. The aggregator's

utility is then $U_a(\boldsymbol{k}) = K(\boldsymbol{k}) - \sum_{i=1}^{N} ck_i^2$. The FOC reveals:

$$\nabla K(\boldsymbol{k}^*) = 2c \cdot \boldsymbol{k}^*.$$

Observation 1 (Halved Innovation). Utilizing the homogeneity of K, we have:

$$\nabla K(2\boldsymbol{k}^*) = \nabla K(\boldsymbol{k}^*) = 2c \cdot \boldsymbol{k}^* = c \cdot (2\boldsymbol{k}^*)\,,$$

thus satisfying the Social Optimum FOC. Therefore, the first observation is that

$$k_i^* = \frac{1}{2} k_i^{sp}, \quad \forall i = 1, 2, \ldots, N.$$

Observation 1' (Substitution Effect). Specifically, under the special case where all correlations are the same, i.e., $\rho_{i,j} = \rho$ for all $i \neq j$, there must exists a symmetric equilibrium such that $k_i = k$ for all $i = 1, 2, \ldots, N$. Moreover, for such $\boldsymbol{k} = k\mathbf{1}$, we have

$$K(k\mathbf{1}) = \frac{Nk}{1 + (N-1)\rho}.$$

The equilibrium precision is $k^* = \frac{1}{2c(1+(N-1)\rho)}$ and the price is $p^* = \frac{1}{2(1+(N-1)\rho)}$. Differentiating the equilibrium price with respect to correlation ρ:

$$\frac{\mathrm{d}p^*}{\mathrm{d}\rho} = -\frac{N-1}{2(1+(N-1)\rho)^2} < 0.$$

Hence, as human content becomes more correlated, the price paid to creators falls. This confirms that data correlation—in addition to the decentralized market structure—further erodes incentives.

B Observation 2: Originality Penalty

While the baseline model in Appendix A establishes the fundamental inefficiency that $\boldsymbol{k}^* = \frac{1}{2}\boldsymbol{k}^{sp}$, the real-world market is characterized by heterogeneity. This appendix analyzes a more general model featuring N agents with varying degrees of correlation with the general (crowd) opinion.

B.1 Model Primitives: The Common Factor Model

To capture the distinction between "representative" and "unique" data, we extend the model to $N \geq 2$ creators and decompose the noise structure into a common (shared) noise and an idiosyncratic (individual) noise. To avoid confusion with the baseline model, we denote the decision variable (idiosyncratic precision) of creator i as h_i.

Assumption 3 (Common Factor Model)**.** *The signal s_i produced by creator i is generated as*

$$s_i = X + \beta_i \eta + \tilde{\varepsilon}_i,$$

where $X \in \mathbb{R}$ is the state variable, $\eta \sim \mathcal{N}(\mu_\eta, \sigma_\eta^2)$ is a hidden common noise, shared by all creators, and $\tilde{\varepsilon}_i \sim \mathcal{N}(0, 1/h_i)$ is an idiosyncratic noise, independent across creators.

The objective function of the GenAI agregator is still represented by (1). We assume a Bayesian setting where the aggregator has a prior distribution on μ_η—the mean of the common bias η—as $\mu_\eta \sim \mathcal{N}(0, \gamma)$ where γ is known to the aggregator. All other parameters $\{\beta_i, \sigma_\eta^2, h_i\}$ are directly observable to the aggregator.

Here, β_i represents the **representativeness** (loading on the common factor) of creator i.

- "Rank-and-file" creators (high β_i) produce content highly correlated with the consensus.
- "Unique" creators (low β_i) produce novel content dominated by idiosyncratic information.

Production Costs. In this model, since only h_i captures the idiosyncratic effort of each creator, we shall assume the production cost of creator i is $C(h_i) = \frac{c}{2} h_i^2$. Similarly, the price paid by the aggregator to each creator is now $p_i h_i$. Therefore, when being offered a price of p_i, creator i's FOC gives an effort level of $h_i = p_i/c$.

Consistency with Baseline Model. The noise model presented in Assumption 1 in Appendix A is structurally equivalent to the special case of Assumption 3 where $\gamma = 0$ (i.e., the common factor η is always zero-mean). Specifically, for any set of parameters $\{h_i, \beta_i, \sigma_\eta^2\}$ in Assumption 3, the baseline parameters $\{k_i, \rho_{ij}\}$ in Assumption 1 are uniquely determined by matching the first two moments of $\varepsilon_i = \beta_i \eta + \tilde{\varepsilon}_i$:

$$k_i = \frac{1}{\mathrm{Var}(\varepsilon_i)} = \frac{1}{\beta_i^2 \sigma_\eta^2 + h_i^{-1}},$$
$$\rho_{ij} = \frac{\mathrm{Cov}(\varepsilon_i, \varepsilon_j)}{\sqrt{\mathrm{Var}(\varepsilon_i)\,\mathrm{Var}(\varepsilon_j)}} = \frac{\beta_i \beta_j \sigma_\eta^2}{\sqrt{(\beta_i^2 \sigma_\eta^2 + h_i^{-1})(\beta_j^2 \sigma_\eta^2 + h_j^{-1})}}.$$

We shall remark that, however, due to the change in production costs and payments, the equilibrium in this section is slightly different from that in Appendix A.

Deriving the BLUE Precision. In our game, the following events happen sequentially: *(i)* The aggregator observes $\gamma, \sigma_\eta^2, \{\beta_i\}_i$ and decides individual prices $\{p_i\}_i$. *(ii)* Each creator decides their effort level, which in equilibrium will be $h_i = p_i/c$. *(iii)* The common bias is realized as $\mu_\eta \sim \mathcal{N}(0, \gamma)$ and $\eta \sim \mathcal{N}(\mu_\eta, \sigma_\eta^2)$ (both realizations unobservable to the aggregator), and signals $\boldsymbol{s} = (s_1, s_2, \ldots, s_N)'$ are generated according to Assumption 3. *(iv)* The aggregator finds an estimator $\hat{X}$ that uses $\boldsymbol{s}$ to estimate X.

We solve this game backwards. In the last step *(iv)*, the aggregator knows $\tilde{\varepsilon}_i \sim \mathcal{N}(0, 1/h_i)$, $\mu_\eta \sim \mathcal{N}(0, \gamma)$, and $\eta \sim \mathcal{N}(\mu_\eta, \sigma_\eta^2)$. Therefore, the signals are generated as

$$\boldsymbol{s} = \mathbf{1}X + \boldsymbol{\xi}, \quad \text{where } \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{H}^{-1} + (\sigma_\eta^2 + \gamma)\boldsymbol{\beta}\boldsymbol{\beta}'),$$

where $\boldsymbol{H} = \text{diag}(h_1, \ldots, h_N)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_N)'$. Given this Gaussian model, the BLUE estimator for X is

$$\hat{X} = (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{s}, \quad \text{where } \boldsymbol{\Sigma} = \boldsymbol{H}^{-1} + (\sigma_\eta^2 + \gamma)\boldsymbol{\beta}\boldsymbol{\beta}'. \tag{2}$$

Using the Sherman-Morrison-Woodbury identity, we invert $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{H} - \frac{(\sigma_\eta^2 + \gamma)\boldsymbol{H}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{H}}{1 + (\sigma_\eta^2 + \gamma)\boldsymbol{\beta}'\boldsymbol{H}\boldsymbol{\beta}}.$$

Substituting this into the precision formula, we derive

$$K(\boldsymbol{h}, \boldsymbol{\beta}) = \mathbf{1}'\left(\boldsymbol{H} - \frac{(\sigma_\eta^2 + \gamma)\boldsymbol{H}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{H}}{1 + (\sigma_\eta^2 + \gamma)\boldsymbol{\beta}'\boldsymbol{H}\boldsymbol{\beta}}\right)\mathbf{1} = \sum_{i=1}^N h_i - \frac{(\sum_{i=1}^N h_i\beta_i)^2}{(\sigma_\eta^2 + \gamma)^{-1} + \sum_{i=1}^N h_i\beta_i^2}. \tag{3}$$

Alternative Interpretation. Instead of the Bayesian perspective, an alternative interpretation is via the minimax robust optimization perspective. Suppose that the aggregator doesn't know exactly μ_η but only knows $\mu_\eta^2 \le \gamma$. The aggregator wants to find a linear estimator $\hat{X} = \boldsymbol{w}'\boldsymbol{s}$ with $\sum_{i=1}^N w_i = 1$ to minimize the worst-case Mean-Square Error (MSE):

$$\min_{\boldsymbol{w}} \max_{\mu_\eta^2 \le \gamma} \mathbb{E}[(\boldsymbol{w}'\boldsymbol{s} - X)^2], \quad \text{s.t. } \boldsymbol{w}'\mathbf{1} = 1.$$

One can decompose the MSE into the variance and the squared bias:

$$\text{MSE}(\hat{X}) = \mathbb{E}[(\hat{X} - X)^2] = \underbrace{\text{Var}(\hat{X})}_{\text{Variance}} + \underbrace{(\mathbb{E}[\hat{X}] - X)^2}_{\text{Squared Bias}}.$$

When $\hat{X} = \boldsymbol{w}'\boldsymbol{s}$ with $\boldsymbol{w}'\mathbf{1} = 1$, we have $\text{Var}(\hat{X}) = \boldsymbol{w}'\tilde{\boldsymbol{\Sigma}}\boldsymbol{w}$ where $\tilde{\boldsymbol{\Sigma}} = \boldsymbol{H}^{-1} + \sigma_\eta^2\boldsymbol{\beta}\boldsymbol{\beta}'$, and $(\mathbb{E}[\hat{X}] - X)^2 = \mu_\eta^2(\boldsymbol{w}'\boldsymbol{\beta})^2$. Therefore, the aggregator solves

$$\min_{\boldsymbol{w}} \left(\boldsymbol{w}'\tilde{\boldsymbol{\Sigma}}\boldsymbol{w} + \gamma(\boldsymbol{w}'\boldsymbol{\beta})^2\right), \quad \text{s.t. } \boldsymbol{w}'\mathbf{1} = 1.$$

The Lagrangian gives $\mathcal{L}(\boldsymbol{w}, \lambda) = \boldsymbol{w}'(\tilde{\boldsymbol{\Sigma}} + \gamma\boldsymbol{\beta}\boldsymbol{\beta}')\boldsymbol{w} - \lambda(\boldsymbol{w}'\mathbf{1} - 1)$. Hence we have $\boldsymbol{w}^* \propto (\tilde{\boldsymbol{\Sigma}} + \gamma\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\mathbf{1}$. This exactly recovers Eq. (2) because the $\boldsymbol{\Sigma}$ there equals $\tilde{\boldsymbol{\Sigma}} + \gamma\boldsymbol{\beta}\boldsymbol{\beta}'$.

B.2 Social Optimum vs. Decentralized Equilibrium

Similar to Appendix A.2, we now solve this market for the social optimum and the decentralized equilibrium. Differentiating the effective precision $K(\boldsymbol{h}, \boldsymbol{\beta})$ in Eq. (3) yields

$$\frac{\partial}{\partial h_i} K(\boldsymbol{h}, \boldsymbol{\beta}) = \left(1 - \beta_i \frac{\sum_{j=1}^N h_j \beta_j}{(\sigma_\eta^2 + \gamma)^{-1} + \sum_{j=1}^N h_j \beta_j^2}\right)^2, \quad \forall i = 1, 2, \ldots, N.$$

Social Optimum. For the social planner who maximizes $K(\boldsymbol{h}, \boldsymbol{\beta}) - \sum_{i=1}^N C_i(h_i)$, the FOC gives

$$1 - \beta_i \frac{\sum_{j=1}^N h_j \beta_j}{(\sigma_\eta^2 + \gamma)^{-1} + \sum_{j=1}^N h_j \beta_j^2} = \sqrt{c \cdot h_i}, \quad \forall i = 1, 2, \ldots, N.$$

Decentralized Equilibrium. Since creators' objectives are unchanged, we still have $h_i = p_i/c$. Solving the FOC for the aggregator's objective of maximizing $K(\boldsymbol{h}, \boldsymbol{\beta}) - \sum_{i=1}^N p_i h_i$ gives

$$1 - \beta_i \frac{\sum_{j=1}^N h_j \beta_j}{(\sigma_\eta^2 + \gamma)^{-1} + \sum_{j=1}^N h_j \beta_j^2} = \sqrt{2c \cdot h_i}, \quad \forall i = 1, 2, \ldots, N.$$

Observation 2 (Originality Penalty). For a "original" creator with $\beta_i \approx 0$, two FOCs simplify to $1 \approx ch_i$ (in Social Optimum) and to $1 \approx 2ch_i$ (in Equilibrium). Therefore,

$$h_i^* \approx \frac{1}{2} h_i^{sp}, \quad \text{for an "original" creator with } \beta_i \approx 0.$$

On the other hand, for a "rank-and-file" creator with $\beta_i \gg 0$, for notational simplicity let $\kappa^{sp} = 1$ and $\kappa^* = 2$. Consider the LHS in either FOC, namely $1 - \beta_i \frac{\sum_{j=1}^N h_j \beta_j}{(\sigma_\eta^2+\gamma)^{-1}+\sum_{j=1}^N h_j \beta_j^2}$. Since $\beta_i \gg 0$, we have $\beta_i(\sum_{j=1}^N h_j \beta_j) \gtrsim \sum_{j=1}^N h_j \beta_j^2$ and thus

$$1 - \beta_i \frac{\sum_{j=1}^N h_j \beta_j}{(\sigma_\eta^2 + \gamma)^{-1} + \sum_{j=1}^N h_j \beta_j^2} \lesssim \frac{1}{1 + \Lambda h_i}, \quad \text{where } \Lambda := \frac{(\sigma_\eta^2 + \gamma) \sum_{j=1}^N h_j \beta_j^2}{h_i} \geq \gamma \beta_i^2 \gg 0.$$

Therefore, when fixing $\boldsymbol{\beta}$ and $\boldsymbol{h}_{-i}$, if we view the LHS as a function of h_i, since $\gamma \gg 0$ and $\beta_i \gg 0$, it starts at 1 but drops extremely fast to 0. On the other hand, the two RHS's $\sqrt{\kappa c \cdot h_i}$ with $\kappa \in \{1, 2\}$ only differ by a constant factor. Since $\frac{1}{1+\Lambda h_i} \approx \sqrt{\kappa c \cdot h_i}$ has a solution of $h_i \approx (\Lambda^2 c \kappa)^{-1/3}$ We shall have

$$\frac{h_i^*}{h_i^{sp}} \gtrsim \frac{(\Lambda^2 c \cdot 2)^{-1/3}}{(\Lambda^2 c \cdot 1)^{-1/3}} = \frac{1}{\sqrt[3]{2}} \approx 0.79.$$

References

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar, "Too much data: Prices and inefficiencies in data markets," *American Economic Journal: Microeconomics*, 2022, *14* (4), 218–256.

Arrow, Kenneth, "Economic welfare and the allocation of resources for invention," *The rate and direction of inventive activity: Economic and social factors*, 1962, pp. 609–626.

Barnett, Jonathan M., "The Free Content Illusion," *Journal of Intellectual Property Law*, forthcoming 2026, *33* (1).

Bergemann, Dirk, Alessandro Bonatti, and Tan Gan, "The economics of data," Technical Report, Cowles Foundation Discussion Paper No. 2196 2019. Available at SSRN: https://ssrn.com/abstract=3446973.

Che, Yeon-Koo, "Beyond the Coasian irrelevance: Asymmetric information," *Unpublished notes, Columbia University.[110, 114]*, 2006.

Coase, Ronald H, "The problem of social cost," *Journal of Law and Economics*, 1960, *3*, 1–44.

Farboodi, Maryam, Adrien Matray, Laura Veldkamp, and Venky Venkateswaran, "Long-run growth of financial data technology," *American Economic Review*, 2019, *109* (1), 448–474.

_ **and Laura Veldkamp**, "A model of the data economy," *Review of Economic Studies*, 2022, *89* (2), 645–692.

_ **and** _ , "Data and markets," *Annual Review of Economics*, 2023, *15* (1), 23–40.

_ **, Gregor Jarosch, and Robert Shimer**, "The emergence of market structure," *The Review of Economic Studies*, 2023, *90* (1), 261–292.

Freilich, Janet and Sepehr Shahshahani, "Measuring Follow-On Innovation," *Research Policy*, 2023, *52*, 104854.

Gale, David and Lloyd S Shapley, "College admissions and the stability of marriage," *The American mathematical monthly*, 1962, *69* (1), 9–15.

Golrezaei, Negin, Adel Javanmard, and Vahab Mirrokni, "Dynamic incentive-aware learning: Robust pricing in contextual auctions," *Advances in Neural Information Processing Systems*, 2019, *32*.

_ **, Ilan Lobel, and Renato Paes Leme**, "Auction design for roi-constrained buyers," in "Proceedings of the Web Conference 2021" 2021, pp. 3941–3952.

_ , **Max Lin, Vahab Mirrokni, and Hamid Nazerzadeh**, "Boosted second price auctions: Revenue optimization for heterogeneous bidders," in "Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining" 2021, pp. 447–457.

_ , **Patrick Jaillet, and Jason Cheuk Nam Liang**, "Incentive-aware contextual pricing with non-parametric market noise," in "International Conference on Artificial Intelligence and Statistics" PMLR 2023, pp. 9331–9361.

Gordon, Wendy J., "Fair Use as Market Failure: A Structural and Economic Analysis of the Betamax Case and its Predecessors," *Columbia Law Review*, 1982, *82*, 1600–1657.

Hayek, F. A., "The Use of Knowledge in Society," *American Economic Review*, 1945, *35* (4), 519–530.

Heller, Michael A, "The tragedy of the anticommons: Property in the transition from Marx to markets," *Harvard Law Review*, 1998, *111* (3), 621–688.

Heller, Michael A. and Rebecca S. Eisenberg, "Can Patents Deter Innovation? The Anticommons in Biomedical Research," *Science*, 1998, *280* (5364), 698–701.

Lee, Edward, "Fair Use and the Origin of AI Training," *Houston Law Review*, 2025, *63*, 105–229.

Lemley, Mark A. and Bryan Casey, "Fair Learning," *Texas Law Review*, 2021, *99*, 743–785.

Leval, Pierre N., "Toward a Fair Use Standard," *Harvard Law Review*, 1990, *103* (5), 1105–1136.

Machlup, Fritz, "An Economic Review of the Patent System," Committee Print Study No. 15, U.S. Senate Subcommittee on Patents, Trademarks, and Copyrights, Washington 1958. 85th Congress, 2d session.

Nordhaus, William, *Invention, Growth, and Welfare: A Thoeretical Treatment of Technological Change*, MIT Press, 1969.

Olson, Mancur, *The Logic of Collective Action: Public Goods and the Theory of Groups*, Harvard University Press, 1965.

Opderbeck, David W., "Copyright in AI Training Data," *Oklahoma Law Review*, 2024, *76*, 951–1023.

Ostrom, Elinor, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, 1990.

Plant, Arnold, "The Economic Theory Concerning Patents for Inventions," *Economica*, 1934, *1* (1), 30–51.

Register of Copyrights, *Copyright and Artificial Intelligence – Part 3: Generative AI Training*, U.S. Copyright Office, 2025. Pre-Publication Version.

Rubinstein, Ariel and Asher Wolinsky, "Middlemen," *The Quarterly Journal of Economics*, 1987, *102* (3), 581–593.

Sag, Matthew, "The new legal landscape for text mining and machine learning," *Journal of the Copyright Society of the USA*, 2019, *66*, 291.

Segal, Ilya and Michael D Whinston, "Property rights and the efficiency of bargaining," *Journal of the European Economic Association*, 2016, *14* (6), 1287–1328.

Shahshahani, Sepehr, "The Nirvana Fallacy in Fair Use Reform," *Minnesota Journal of Law, Science & Technology*, 2015, *16* (1), 273–341.

_ , "The Role of Courts in Technology Policy," *Journal of Law and Economics*, 2018, *61* (1), 37–64.

Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson, "The curse of recursion: Training on generated data makes models forget," *Nature*, 2024, *631*, 755–763. Originally published on arXiv in 2023.

Tullis, Jane, "Sifting Through the Slop: How Generative AI Created a Market for Lemons for Text-Based Works," 2025.

Williamson, Oliver E, "Transaction-cost economics: the governance of contractual relations," *Journal of Law and Economics*, 1979, *22* (2), 233–261.