IN-DEPTH ANALYSIS

Requested by the JURI Committee

The economics of copyright and AI

Empirical evidence and optimal policy

Policy Department for Justice, Civil Liberties and Institutional Affairs
Directorate-General for Citizens' Rights, Justice and Institutional Affairs
PE 778.859 - December 2025

EN

The economics of copyright and AI

Empirical evidence and optimal policy

Abstract

This in-depth analysis, commissioned by the European Parliament's Policy Department for Justice, Civil Liberties and Institutional Affairs at the request of the Committee on Legal Affairs, examines how copyright policy should respond to artificial intelligence (AI). It combines historical lessons from digital markets, insight on the economic value of data, and a formal model to study welfare effects. It assesses economic effects of various policy options, including an exception, an exception with opt-out, licencing market ("opt-out") and statutory licencing, in a search for optimal policy.

This document was requested by the European Parliament's Committee on Legal Affairs.

AUTHOR

Christian PEUKERT, Professor of Digitization, Innovation and Intellectual Property, University of Lausanne, Faculty of Business and Economics (HEC Lausanne)

ADMINISTRATOR RESPONSIBLE

Mariusz MACIEJEWSKI

EDITORIAL ASSISTANT

Christina MARGELI

LINGUISTIC VERSIONS

Original: EN

ABOUT THE EDITOR

Policy departments provide in-house and external expertise to support EP committees and other parliamentary bodies in shaping legislation and exercising democratic scrutiny over EU internal policies.

To contact the Policy Department or to subscribe for updates, please write to:
Policy Department for Justice, Civil Liberties and Institutional Affairs
European Parliament
B-1047 Brussels
Email: poldep-iust-b@europarl.europa.eu

Manuscript completed in December 2025

This document is available on the internet at:
http://www.europarl.europa.eu/supporting-analyses

CONTENTS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
CMO	Collective Management Organization
TDM	Text-and-Data-Mining
UGC	User-Generated Content
DMA	Digital Markets Act
DSA	Digital Services Act
GenAI	Generative Artificial Intelligence
RAG	Retrieval-Augmented Generation
C2PA	Coalition for Content Provencance and Authenticity

LIST OF TABLES

EXECUTIVE SUMMARY

This in-depth analysis examines how copyright policy can balance cultural production and technological innovation in the age of generative Artificial Intelligence (AI). Drawing on historical lessons from digitisation, new evidence on the economics of data, and a calibrated welfare model, it identifies one central insight: policy should focus on sustaining the flow of new, high-quality data.

Core idea. Creators decide whether and how much to create by weighing their costs against the returns they expect in future. Copyright grants exclusive rights to generate those future returns. The economic rationale of copyright is not to provide insurance for revenue streams from existing works but to incentivise future creation. However, empirical evidence shows that when creators learn their works are used for AI training, they reduce output afterwards. Should creators therefore be compensated for new works going forward? If we want to preserve the level of creation that existed before AI—and we do, if society's gains from AI depend on a continuous flow of high-quality works—then compensation for creators can provide the necessary incentives. The central question of this briefing is which politically realistic policy option achieves this goal most effectively for society as a whole, i.e. making creators better off while minimising the costs to end-users and AI developers.

Key Findings

Consumers ultimately fund incentives. Copyright royalties are largely borne by end users, not AI firms. This makes consumer welfare—not just creators' or firms' interests—the central benchmark for policy.

Statutory licensing appears optimal. Among realistic options, statutory licensing emerges as a robustly welfare-superior framework under most calibrations. Making the licence compulsory ensures wide access to works, a royalty rate set by a regulatory authority balances interests of rightsholders, AI developers and end users, enabling royalty payments that create incentives for enough new creation to keep AI systems valuable over time.

Other regimes underperform. Relying on licensing markets alone risks suffering from high clearance costs and too little available training data due to opt-outs. Copyright exceptions without remuneration risk undercutting incentives for future creation. Copyright exceptions without remuneration and opt-outs appear as the worst policy option, because opt outs reduce access to the existing stock of data and there is no funding for a continued flow of data.

Complementary measures matter. Statutory licensing is most effective when paired with lean administration, strong competition enforcement in the AI market. To encourage further research and development, I propose a targeted carve-out for non-commercial research that links eventual commercialisation back to creator funding.

Geopolitical relevance. With most AI development outside Europe, statutory licensing in Europe helps ensure that a sustained flow of European works remains in training datasets, making non-European AI services reflective of preferences of European consumers (e.g. corresponding to local languages) and therefore valuable for European consumers.

Policy Roadmap

1. Consider statutory licensing with royalties set by an independent authority that balances interests of rightsholders, AI firms, and—crucially—consumers/end users.
2. Prioritise lean, transparent administration—automated tools where possible, CMOs as second-best.
3. Integrate copyright with competition and research policy, including the DMA and DSA.
4. Close evidence gaps through systematic disclosure on training data, retraining frequency, and royalty flows.

This framework secures broad access, sustains incentives for new creation, and preserves the substantial societal benefits of AI while keeping costs as low as possible for European consumers.

At a Glance: Copyright Policy Options and Economic Effects

Option	Pros	Cons
Exception	Broad access to stock of data; no need to identify rightsholders (no orphan-works problem).	No funding for new works; risks undercutting future creative supply (flow of data).
Exception with Opt-Out	Access to stock of data; no need to identify rightsholders (no orphan-works problem).	Accessible stock of data shrinks when creators opt out. No funding for new works; risks undercutting future creative supply (flow of data).
Licensing Market ("Opt-In")	Access to stock of data, as well as access to flow of data because compensation gives incentives for continued creation; preserves rightsholder choice.	Transaction costs (inefficient bargaining with large number of individual rightsholders) leading to fragmented coverage of data (stock and flow) leading to selective (less representative) data. Especially when licensing deals are selective (e.g. few large rightsholders) to reduce total transaction costs
Statutory Licensing	Broad access to stock of data; no need to identify rightsholders (no orphan-works problem). Royalty stream restores creation incentives and ensures flow of data; low transaction costs because royalty rate is set centrally.	Requires independent authority to set royalty rate; administration of royalty payout must be lean not to undercut effectiveness.

All simulations presented in this briefing are generated from a public, interactive dashboard: https://copyright-policy.streamlit.app. Users can adjust parameters such as pass-through, administrative overhead, and the importance of fresh data, and immediately observe how welfare rankings and optimal royalties' change. This ensures that results are not contingent on any single calibration and allows stakeholders to stress-test conclusions under their own assumptions.

1. STATE OF THE ART: THE ECONOMICS OF COPYRIGHT

KEY FINDINGS

Lessons from history: Despite copyright-related challenges, creative industries mostly benefited from digitisation through licensing and business model innovation. Digital technology reduced the cost of copying existing works and made it cheaper to create new ones. Attempts to enforce copyright against large-scale infringement (online piracy) were short-lived and ineffective. Consumers enjoyed a huge increase in the supply of traditional new content. An even larger explosion of user-generated content was enabled by a shift from enforcement (notice-and-takedown) to platform-mediated, algorithmic licensing, which reduced per-transaction friction and opened new revenue streams.

What is different this time around? Artificial intelligence further lowers the cost of copying and creation. However, it has effects that extend far beyond the creative industries. In recent estimates, aggregate consumer/user surplus is on the order of 10 times larger than the revenues of the AI industry. The output of the creative industries is an important data input. The sustained value of AI hinges on fresh and informative data, but access to it is often governed by copyright. Private incentives to enforce these rights are greater than is optimal for society.

Do we need to regulate? The policy target is not retroactive payment for the stock of existing works for its own sake, but to maintain incentives for the flow of future high-quality creation that feeds AI. Regulation is needed if there is a market failure: a situation where neither party fully works in favour of consumers/end users. Creators will demand too much for their data, AI firms will offer too little to get full access to it and uncover the full potential for society, and, on top of it all, bilateral clearance at internet scale is infeasible. The role of policy is to maximise total welfare: keep access broad, protect dynamic incentives for creators, and minimise administrative overhead.

1.1. Copyright and the First Wave of Digital Technology

Copyright has long been the institutional framework through which societies balance incentives for creation with the dissemination of knowledge and culture. Traditionally, copyright law granted creators exclusive rights to reproduce, distribute, and adapt their works, thereby allowing them to recoup investments in time, talent, and capital. Yet these rights also generate deadweight loss and restrict access to works for users (Landes and Posner, 1989).

The advent of digital technology had pronounced implications for copyright (see Peukert and Windisch, 2025 for a detailed review of the literature). Beginning in the late 1990s, digitisation enabled the lossless copying and distribution of music, films, and books at near-zero marginal cost. Peer-to-peer file-sharing services such as Napster brought mass-scale demand-side copyright infringement, but enforcement through litigation or stricter laws proved largely ineffective. Evidence shows that announcements of anti-piracy laws in France and Sweden led only to short-lived increases in sales, and the effects typically vanished within six months (Adermon and Liang, 2014; Danaher et al., 2014b). While legal efforts succeeded in shutting down or blocking particular unlicensed services (Danaher and Smith, 2014; Poort et al., 2014; Peukert et al., 2017; Danaher et al., 2020), alternatives were quick to emerge (Lauinger et al., 2013; Aguiar et al., 2018). Similarly, lawsuits against individual users failed to curb piracy, and self-regulation by the involved actors failed to reduce the supply of unlicensed copies (Batikas et al., 2019).

This all happened against the background that was not clear whether massive-scale infringement actually caused material harm to rightsholders. Some evidence suggests that negative sales effects are negligible (Oberholzer-Gee and Strumpf, 2007; Aguiar and Martens, 2016) or even positive because of word-of-mouth effects (Givon et al., 1995; Peukert et al., 2017; Lu et al., 2020). In contrast, other studies suggest displacement effects (i.e., reduced sales) in music, movies, and software between 3.5% and 64% (Rob and Waldfogel, 2006, 2007; Hong, 2013; Yue, 2020). In sum, displacement effects are heterogeneous and context-dependent; while losses exist in some settings, they are far from total and vary widely across media and time.

The turning point came when the market started to offer licensed digital services. Platforms such as iTunes pioneered legal digital distribution in the early 2000s, unbundling music albums into single tracks at low prices (Danaher et al., 2014a). Later, streaming services like Spotify and Netflix offered vast catalogues under a variety of subscription models, which reduced the demand for piracy (Wlömert and Papies, 2016; Aguiar and Waldfogel, 2018b). These innovations successfully shifted consumer behaviour towards legal consumption (Danaher et al., 2010; Lu et al., 2021). Aggregate data indicate that unlicensed music consumption has been continuously decreasing since 2010, while consumption of unlicensed movie and TV show content has been stagnant and slower to decrease (UK Intellectual Property Office, 2016; EU Intellectual Property Office, 2021).

At the same time, digital technology enabled unprecedented supply-side dynamics. The costs of creation, promotion, and distribution fell drastically, leading to a massive increase in the number of new works across books, music, film, and other media (Waldfogel, 2017). Platforms hosting user-generated content further democratised creative expression, giving rise to billions of creative expressions in the form of text, image and video on platforms such as Facebook, Twitter, Instagram, YouTube and TikTok. There has been much public debate about the revenue distribution models of music streaming platforms, and different academic papers arrive at different conclusions regarding the welfare effects (e.g. fairness) of different revenue sharing models (e.g. pro-rata vs. user-centric model) (Lei, 2023; Moreau et al., 2024; Bergantiños and Moreno-Ternero, 2025)

Yet this surge in User-Generated Content (UGC) also created new challenges for copyright enforcement, since much of it involves reuse of existing works. The institutional response was the shift from pure enforcement (notice-and-takedown systems) to algorithmic licensing. YouTube's ContentID system exemplifies this transformation: rightsholders are automatically notified when content matches their works and can choose either to block or monetise it. By 2021, ContentID had processed hundreds of millions of claims, with more than 90% resulting in monetisation, generating billions of dollars in revenue for rightsholders (YouTube, 2021).

A growing evidence base shows that licensing (automated or traditional) leads to new income streams for original rightsholders: UGC on YouTube boosts music sales and streams for most rightsholders (Kretschmer and Peukert, 2020; Wlömert et al., 2024), TikTok increases discovery of new songs (Cheng et al., 2025; Bairathi et al., 2024), and re-use in films and new songs (digital sampling) promotes sales of original works (Watson, 2017; Aguiar and Chen, 2024). In the context of books, Nagaraj and Reimers (2023) find that the availability of digital versions (via the Google Books project, licensed through settlement with rightsholders) can increase the sales of physical books. Sales increases occur more strongly for less popular books, with demand spilling over into non-digitised works by authors.

From a welfare perspective, licensed access expanded consumption and reduced deadweight loss relative to the enforcement-only era. At the same time, platform-mediated and algorithmic licensing lowered per-transaction frictions compared to bilateral licensing. These systems approximate collective licensing and reduce administrative overhead relative to one-off clearance, though they raise concerns about privatised enforcement and potential bias toward large rightsholders (Erickson and Kretschmer, 2018).

1.2. The Economics of AI

1.2.1. Demand-Side and Overall Effects

AI, and Generative AI (GenAI) in particular, has further lowered production costs and enabled an explosion of content. The scope and potential spillovers, however, are far broader. Models can generate text, images, music, and code that are, at least for specific tasks, on par with or even exceeding the quality of human creations. This has implications not only for creative industries but also for knowledge production and the broader economy.

A growing body of literature shows that AI can raise productivity and improve the quality of human work across a variety of settings. For example, AI reduces completion times in professional writing by about 40% while improving quality by nearly 20% (Noy and Zhang, 2023). In customer support, productivity increases by around 15% thanks to AI, though gains differ substantially across workers (Brynjolfsson et al., 2025). In software development, AI assistants shift effort toward coding and away from management tasks (Hoffmann et al., 2025). A field experiment in management consulting shows that tasks are completed more quickly and with higher-rated outputs (Dell'Acqua et al., 2023). In creative work, studies suggest that incorporating GenAI in the human workflow can make creators more productive, helping some creators to create works that are further away from existing works ("increase in peak novelty") while the average novelty of works decreases (Zhou and Lee, 2024; Doshi and Hauser, 2024). Overall, this strand of literature suggests that within individual tasks, AI often complements human effort. The implication is clear: if user benefits are large, preserving the conditions for a continued flow of quality training data becomes a central welfare objective.

At the same time, several studies document the displacement of certain types of work. User engagement on Q&A platforms such as StackOverflow declined markedly following the release of ChatGPT (Quinn and Gutt, 2025; Burtch et al., 2024; del Rio-Chanona et al., 2023; Yilmaz et al., 2023). A possible implication is that GenAI may deplete its own future training data: if fewer questions are asked by humans, fewer human-generated answers will be available. Similar concerns are raised in a study that looks at equilibrium effects in an art platform. Goldberg and Lam (2025) find that GenAI substitutes for non-GenAI content, crowding out lower-quality non-GenAI creators, while entry of GenAI producers increases variety, quality, and sales. Consumers and platforms benefit, but non-GenAI creators are disadvantaged, raising long-term concerns about sustaining original content for future training. More broadly, there is now quite a lot of high-quality evidence on labour effects of AI. On freelance platforms, demand for translation, writing, and graphic design services has fallen, with corresponding reductions in employment and earnings (Demirci et al., 2025; Hui et al., 2024; Yilmaz et al., 2023). The advertising sector provides a particularly illustrative case. Hartmann et al. (2025) show that AI-generated advertising imagery is often rated as more aesthetically appealing than comparable human-created content, though quality, creativity, purchase intention, and brand engagement remain statistically indistinguishable. In real-world campaigns, however, about half of AI-generated stock photos outperform human-made ones in click-through rates by up to 50%, while the other half underperform by up to 25%. Importantly, however, displacement is not uniform across tasks. Evidence from a freelancing platform shows that the demand for AI-related coding has grown by 24%, with AI chatbot development nearly tripling (Teutloff et al., 2025). Overall, the evidence points to heterogeneity: tasks that are easily automated tend to be substituted, whereas others give rise to complementarities between human and machine skills.

The public discussion also points to potential effects of GenAI on content industries, in particular online news.[1]The worry is that AI-generated summaries of articles displace traffic to original sources, which reduces revenues that are essential to producing content in the first place. Systematic evidence on how AI summarisation, e.g. through technologies like Retrieval-Augmented Generation (RAG)[2], affects publishers is still limited. An early study suggests mixed effects (Duch-Brown et al., 2025): overall traffic to news outlets remained stable after Google introduced "AI Overviews" in the US, compared to traffic in the EU where the feature was not yet rolled out. When ChatGPT introduced links in answers, visits and pageviews increased, while pageviews per visit decreased. These findings resonate with a well-studied analogue: news aggregation platforms. When Google News introduced a localisation feature, traffic to local news outlets increased (George and Hogendorn, 2020). After Google News shut down in Spain, visits to news outlets decreased between 8% and 14% (Calzada and Gil, 2020; Athey et al., 2017). Similarly, after Google News introduced an opt-in policy in Germany, visits to opted-in outlets increased by 8% (Calzada and Gil, 2020). It is too early to come to strong conclusions, but the available evidence suggests that, rather than diverting traffic away from outlets, GenAI has the potential to expand the news market, at least on average. The related evidence, however, also suggests that the effects might be heterogeneous across publisher size and type and more empirical work is needed to better inform the policy debate.

With evidence pointing to both complementary and substitutive effects, and with distributional consequences that remain complex and largely unknown, the crucial question is what these effects

[1] See, e.g. "AI is killing the web. Can anything save it?" and "To survive the AI age, the web needs a new business model" in The Economist, https://www.economist.com/business/2025/07/14/ai-is-killing-the-web-can-anything-save-it and https://www.economist.com/leaders/2025/07/17/to-survive-the-ai-age-the-web-needs-a-new-business-model.

[2] RAG is an AI approach that combines information retrieval from external sources with text generation (e.g., when asked about a recent event, the system retrieves relevant news articles and then generates a grounded summary).

amount to in the aggregate. It turns out that, when measured at the aggregate level, the benefits of AI are strikingly large. Economists capture these gains as *consumer surplus* — the difference between what people would be willing to pay for a product and what they actually spend. Recent methodological advances make it possible to measure consumer surplus at scale, in ways that are representative of an entire population. Massive online choice experiments directly elicit willingness to accept compensation for giving up access to digital services: respondents are offered monetary rewards (actually paid out), and by varying the reward, researchers trace out demand and calculate the value of access (Brynjolfsson et al., 2018). Applied to AI, this approach estimates annual consumer surplus in the United States at roughly $97 billion compared to $7 billion in industry revenues (Brynjolfsson and Collis, 2025).

Nevertheless, the implication is striking: although AI produces nuanced and sometimes disruptive effects in particular labour markets, the aggregate welfare gains are an order of magnitude larger than industry revenues. Most of the value accrues not to firms but to users, both consumers and businesses, highlighting the substantial societal benefits of AI. Because most value accrues to users, any AI policy, including those regarding copyright, needs to consider total welfare effects, weighing consumer surplus alongside creator and firm surplus.

1.2.2. The Production Side

The tremendous societal value of AI raises a central policy question: how is this value created, and is it distributed in a fair way? Addressing this question requires a closer examination of the underlying production process.

The economics of data

The value chain of AI rests on three inputs: engineering, compute and data. Recent estimates in Cottier et al. (2024) suggest that the amortised cost of training top foundation models exceeds tens of millions of dollars. These costs have been increasing at a rate of approximately 240% per year since 2016, and have grown so large that, with a few exceptions of recent government-backed investments, only commercial players can afford to produce frontier models (Ahmed et al., 2023).

If this trend continues, the most expensive training runs could exceed $1B by 2027.[3] Through detailed case studies (GPT-3/4, OPT-175B, Gemini Ultra), the authors calculate that engineering (i.e. staff compensation) accounts for about 29–49% of total amortised development cost, and the remainder is the cost for compute (hardware 47–65% and energy 2–6%). Perhaps the most important reason for the sharp rise in costs, however, is that the third key input factor — *data* — has been scaled up dramatically. Data inputs have expanded to the point that commentators describe firms as being close to "exhausting most of the internet's data."[4] Data enters the cost calculation in Cottier et al. (2024) only indirectly through the cost for compute. The price to access data, however, is not always zero and often governed

[3] A reason to be skeptical about this projection is that it assumes linear growth in logs. Historically, diffusion and performance improvements in general-purpose technologies typically follow S-curves that plateau as adoption saturates, bottlenecks bite, and incremental improvements become harder (Geroski, 2000).

[4] See, e.g., "AI firms will soon exhaust most of the internet's data," in The Economist, https://www.economist.com/schools-brief/2024/07/23/ai-firms-will-soon-exhaust-most-of-the-internets-data.

by copyright law. As of September 2025, AI developers have closed deals with content providers such as news publishers, scientific publishers, and book authors worth at least $300mn.[5]

How is data turned into economic value? Does progress in AI inevitably depend on ever-bigger datasets and ever-higher budgets? The answer appears to be no. Large-scale evidence from a variety of empirical settings clearly shows that there are decreasing returns to data: adding more improves performance, but at a diminishing rate (Peukert et al., 2024; Schaefer and Sapi, 2023; Kaplan et al., 2020). Importantly, however, the value of data can depreciate over time (Valavi et al., 2022). An example is automated news recommendation: in a large-scale field experiment, Peukert et al. (2024) show that the effectiveness of personalised news recommendation, in terms of click-through rates, drops to zero after just 1.5 days without adding fresh personal data on a user's click history. While not every domain is as fast-paced as the news of the day, the idea is general. As the world evolves, yesterday's data loses informational value, making it necessary to update.

Hence, improving AI performance cannot be achieved by adding ever more data alone, but by specialising in the type of data that is most useful, e.g. in terms of domain-specificity and timeliness. Fine-tuning has proven particularly effective in this regard: by adapting foundation models on smaller, carefully curated datasets, performance on specific tasks can exceed that of large general-purpose models, often at a fraction of the cost. For example, Zhao et al. (2024) show that across 310 fine-tuned models, most not only surpassed their base models but also outperformed GPT-4 on a wide range of tasks. This highlights a key shift: rather than relying solely on scaling, AI developers and sophisticated end-users can achieve better results by deploying smaller, specialised models tailored to their domain. This, in turn, raises an important policy question: how can we create incentives for the kinds of high-quality data that appear central to maintaining the current value of AI or drive further progress? One possible answer is synthetic data, letting AI itself generate additional training material (Lee, 2025). Yet this approach has inherent limits, as models can only recombine what they have already seen, and evidence suggests that performance based on synthetic data can be fragile, leading to so-called "model collapse" (Shumailov et al., 2024). For many applications, new data will therefore need to come either from human expertise (e.g., annotation), from domain-specific measurement (e.g., sensors), or as a side-product of human creative expression. While annotation and sensor data can, in principle, be purchased or built, human creativity raises more complex questions because its production is incentivised by copyright. As a result, further advances in AI may hinge not only on investments in compute or engineering, but on markets for data, as well as the design of copyright law and related incentive structures.

Copyright and the Economics of Data

Copyright plays a central role in access to high-quality data, particularly when such data originates from human creative expression. Copyright protection applies automatically, typically lasting for the author's life plus 70 years. As a result, copyrighted material is pervasively part of training datasets.[6] Avoiding

---

[5] See e.g. the licence deal tracker here: https://petebrown.quarto.pub/pnp-ai-partnerships/, as well as reports about copyright dispute settlements, e.g. here: https://www.theguardian.com/technology/2025/sep/05/anthropic-settlement-ai-book-lawsuit.

[6] We lack precise evidence on what share of the data that underpins AI applications is subject to copyright protection. Some of the most economically valuable data (for instance, behavioural data used in advertising such as clickstreams) may fall largely outside copyright. Nevertheless, given the automatic protection of creative expression, it is plausible that a large share of AI training data falls within the scope

copyright is not only difficult but can be costly in terms of model performance: recent evidence shows that large language models provide systematically better responses related to copyrighted books that were included in training than when asked about copyrighted books that were not included in training (Jia and Nagaraj, 2025). This is in line with evidence showing that exclusive rights for creative works can have substantial societal costs. Consumers and downstream innovators are often better off under weaker copyright regimes. Empirical evidence shows that the exclusive nature of copyright is mirrored in higher prices and lower availability (Biasi and Moser, 2021; Reimers, 2019; Li et al., 2018; Flynn et al., 2019; Heald, 2014), which decreases consumer surplus more than it increases profits of copyright holders (Reimers, 2019). For example, copyright restrictions reduced reuse on Wikipedia by up to 135%, cutting traffic by 20% (Nagaraj, 2018). Similarly, works produced before 1923, which are generally in the public domain, are far more accessible today than those produced later (Heald, 2008; Buccafusco and Heald, 2013; Reimers, 2019). When works enter the public domain, they are more widely performed and distributed, increasing consumer surplus more than any incremental revenue for rightsholders (Watson et al., 2023).

Legal uncertainty compounds these challenges. Even in jurisdictions that provide statutory pathways for data analysis, such as the US through fair use or the EU through Text-and-Data-Mining (TDM) exceptions, it remains contested whether these provisions extend to AI training (e.g., Lemley and Casey, 2020; Lucchi, 2025) with a total of 50 filed lawsuits against AI companies in the US alone.[7] Put differently, while AI firms believe that copyright exceptions apply, authors believe that AI models violate their rights. To add to the complexity, in many other countries, copyright exceptions do not exist at all, effectively requiring licences for any use of copyrighted material in training (Fiil-Flynn et al., 2022). Licence costs can be prohibitively high for two reasons. First, identifying individual rightsholders at web-scale is close to impossible. Second, rightsholders' private incentives are misaligned with societal welfare. Empirically, the absence of copyright rules that favour data access is associated with macro-economic costs. There is a correlation: countries with relevant copyright exceptions tend to generate more research output, and investment in AI innovation (Handke et al., 2021; Peukert, 2025b). It is difficult to disentangle all other factors from the impact of copyright law in a cross-country study and these studies can only control for factors such as the rule of law, population size, economic prosperity and the R&D intensity of a country. Overall, countries where copyright law allows for text-and-data-mining or has similar exceptions, have about 0.5% quicker growth in research output, patents and new venture foundations than countries without such exceptions.

Copyright can Incentivise the Flow of New Data

If copyright likely hinders more innovation (e.g. in terms of AI value) than it helps to create (e.g. in terms of cultural value), should we simply abolish it in this context and allow AI firms to use whatever data they can access without compensating rightsholders? The answer is most likely no. Economically, it's a question of dynamic efficiency. To decide whether copyright helps or hinders AI innovation, we need to distinguish between the *stock* and the *flow* of creative works. The existing stock of copyrighted material is already fixed; compensating for its use does not alter the amount of data currently available for training.

of copyright law. In addition, given the long duration of copyright (varying across jurisdictions, but minimum 50 years after the death), the number of works in the public domain is likely tiny. For example, a community effort to make an image dataset comprised only of public domain images yielded 12.4 million image-text pairs (Spawning PD12M), whereas the web-scale alternative has 5 billion image-text pairs (LAION-5B).

[7] See https://chatgptiseatingtheworld.com/2025/09/08/copyright-suits-v-ai-hits-50-book-authors-hendrix-v-apple/.

The real policy challenge is to secure the *flow* of new works in future. If creators anticipate that their output will be used without adequate returns, they may scale back production, constraining the supply of fresh, high-quality data on which the societal value of AI depends. If it is true that creative production is financed intertemporally, with revenues from yesterday's works funding today's effort, reduced expected returns today risk shrinking tomorrow's supply. Copyright compensation, from the perspective of overall society, is therefore *not* about retroactively offsetting creators for changed revenue streams. However, the key policy instrument is to compensate creators for future works to preserve the dynamic incentives that ensure a continued flow of creative data for future AI development.

The Role of Regulation

The design of institutions for data access is crucial for the benefit of society (i.e. total welfare). We want to achieve two goals: (i) ensure attribution and compensation commensurate with the value created, and (ii) minimise distortions to innovation and markets. Striking this balance effectively, especially in the context of international legal asymmetries, can itself become a source of competitive advantage.

Having established that the central concern is to sustain the flow of new, high-quality data, the next question is whether the market on its own will provide sufficient incentives. Put differently: *do we face a risk of too little creative expression to support further progress in AI?*

If AI systems output material that is identical or close substitutes toexisting works, creators' revenues could fall. Technically, models are capable of memorising and reproducing copyrighted works: Cooper et al. (2025) show that large language models can regurgitate books verbatim, while Liu et al. (2024) demonstrate that copyrighted text can frequently be elicited, especially under "jailbreaking" attacks. Yet the practical relevance of such outputs appears limited. For example, reconstructing Harry Potter and the Sorcerer's Stone from an LLM would take at least 17 minutes, cost around $30, and require significant technical expertise, while a used version of the book retails for less than $2 on Amazon. Similarly, overly strict model safeguards sometimes block public-domain material, highlighting that current protections are imperfect in both directions (Liu et al., 2024).

All this suggests that while technically feasible, largescale revenue losses from direct substitution are unlikely. While aggregate evidence suggests substitution effects are limited relative to overall consumer gains, distributional impacts on specific sectors or rightsholder groups can still be significant and warrant policy attention. More research, e.g. to understand how users actually interact with AI systems, is necessary to understand the true economic significance. Even if actual displacement is limited, rights holders' perceptions of loss may be greater, given the difficulty of measuring the true causal impact of AI on revenues. If these perceptions feed into business decisions, investment in new creative works may decline more than they should. Four recent studies, all in the context of stock photo platforms, provide early evidence that generative AI can depress creative supply. Strikingly, while the studies differ in settings, methods, data and the calendar time they cover, they all find similar effect sizes. Lin (2024) finds a 21% reduction in digital content production on a platform that used material for training, compared to stable output on a control platform. Similarly, Peukert et al. (2025) document a roughly 20% decline in content uploads, as well as a decrease in novelty and variety of new contributions after the stock photo platform Unsplash released images for commercial AI use. Huang et al. (2023) report that introducing generative AI on one platform reduced user content, while a competing platform that banned AI saw increased contributions. Together, these findings suggest that expectations of uncompensated use can reduce both the quantity and quality of creative output. Goldberg and Lam (2025) show that the number of images not created with AI decreases by about 20% after a stock photo platform introduces a policy that allows the uploads of AI-generated content. While none of the studies can pinpoint whether the observed effects are driven by economic rationale, e.g. creators perceived harm from AI, they all seem to suggest that AI can change the amount of creative production.

Practical Concerns

Paying creators to offset AI-related reductions in output can work only insofar as creative supply is elastic with respect to remuneration. A first possibility is that remuneration "crowds out" intrinsic motivation. Evidence in favour of such a mechanism is mostly situated in the context of volunteer work, civic duty, blood donation, etc. (Frey and Jegen, 2001). In the context of art, however, evidence shows that artists do increase time spent on artistic work when expected income rises, implying that labour supply is not perfectly inelastic (Towse, 1996). Historical evidence shows that introducing copyright can have positive incentive effects: when parts of Italy received copyright protection via Napoleon's export of French law, the quantity and quality of newly composed operas rose substantially (Giorcelli and Moser, 2020).[8] Remuneration is not the only driver of creative expression. Evidence shows that intrinsic motives, self-realisation, expression, status, and participation, matter as well (Tushnet, 2009; Zimmerman, 2011; Liu et al., 2014).

Even if one accepts that compensation is needed to sustain incentives, implementation is non-trivial. Licensing training data faces three frictions. *First, rent capture by intermediaries:* a large share of copyright-enabled revenue accrues to publishers, diluting incentives for creators (Towse, 2000). Incentives of authors and rightsholders can diverge (e.g., authors valuing wide dissemination while publishers restrict quantities). This distributional concern is echoed in studies arguing that copyright often benefits large organisations over individual creators (Lunney Jr, 2001; Bently, 1994; Minassian, 1997) and in academic publishing (McCabe and Snyder, 2018).[9] *Second, transaction costs:* ownership is increasingly fractional—songwriters and publishers per track have tripled over 60 years— which substantially raises clearance costs (Caoui and Galasso, 2024). *Third, information failures:* large datasets lack reliable licence metadata (Shcherbakov et al., 2025; Longpre et al., 2024); orphan works are widespread (about 13% of in-copyright EU books; up to 95% for early 20th-century newspapers), and clearance can cost 20–50 times more than digitisation (Vuopala, 2010). A potential solution are opt-out regimes that permit use unless rights-holders object, which can ease the orphan-works problem. Evidence shows that existing opt-out mechanisms (e.g. DoNotTrain), while covering huge amounts of

[8] By contrast, on the intensive margin (strengthening rights for existing works, e.g., longer terms), the evidence points to higher prices with little or no additional creation. Extensions such as the UK Copyright Act of 1814 and the US Copyright Term Extension Act of 1998 raised prices (Li et al., 2018; Reimers, 2019), reduced availability, and lowered consumer surplus by more than any profit gains to rightsholders (Reimers, 2019), with negligible incentive effects from yet-longer terms (Pollock, 2009).

[9] The available evidence leads Towse (2006) to conclude that "there is more than a strong suspicion to be found in these economic studies that the main benefits of copyright are enjoyed by the 'humdrum' side of the cultural industries rather than the creators and, while it may well be the case that collecting societies have considerable revenues, the distributions of royalties to artists other than the top few stars show how relatively little they get through the copyright system." (p. 578).

works, are predominantly used by a small number of rightsholders (e.g. stock photography websites; Shcherbakov et al., 2025). Because opt-outs are concentrated among certain right-holders, the types of works that are opted out of web-scale training datasets are systematically different from the population of works, which can introduce bias in training and lead to bias in AI outputs. Algorithmic systems (e.g., ContentID) partially automate licensing/enforcement yet privatise it and may tilt outcomes toward large rightsholders (Erickson and Kretschmer, 2018).

Taken together, abolishing copyright or leaving incentives entirely to the market risks undercutting the *flow* of new creative data. Hence a structured evaluation of alternative policy instruments is essential. The next section develops an economic model to compare such options and their welfare implications.

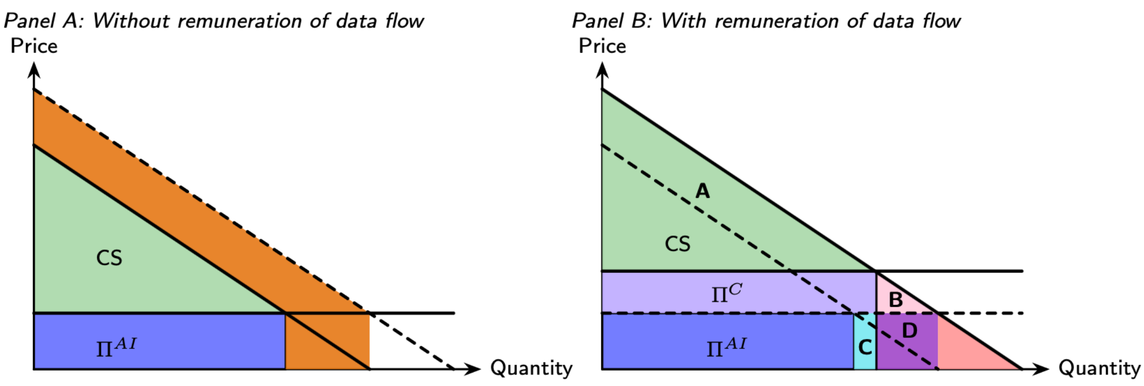
2. ECONOMIC MODEL OF COPYRIGHT AND AI

KEY FINDINGS

We develop a theoretical model that can be calibrated with empirical data to simulate different policy options. This box presents key insights using a simplified illustration: **compensating creators for new data benefits all actors, but the costs are borne mainly by consumers**. We then summarise results from the full model: **from the perspective of total welfare, statutory licensing and a licensing market ("opt-in") strictly dominate a copyright exception with opt-out**. The remainder of the section describes the full model and its results and caveats in detail.

SIMPLIFIED MODEL SETUP: Users demand AI according to how much they value it. The shape of the demand curve depends on the amount and freshness of the data AI can learn from. A firm offers AI in the market at a positive price.

Figure 1: Illustration of simplified model



Note: Without compensation of rightsholders (for the stock of data), consumers extract the entire area under the demand curve until the initial price, i.e. $CS^* = CS + A + \Pi^C + B$ in Panel B. The AI firm extracts $\Pi^{AI*} = \Pi^{AI} + C + D$. For simplicity we assume zero transaction costs for licensing.

DATA GOES STALE: Panel A of Figure 1 shows the following situation. Because data has become stale, the value of AI has decreased and the demand curve shifts to the left (from dashed to solid). At the same price, end users now buy a smaller quantity. The orange area marks the welfare loss: both consumers and the AI firm would have been better off if fresh data had been available.

CREATORS GET COMPENSATED FOR FRESH DATA: Panel B of Figure 1 shows what happens when creators are compensated for providing new data, which improves quality and shifts demand outward again. Consumers are better off than in Panel A: they gain the area A. Yet higher prices mean that some of this gain is offset. Consumers transfer income to rightsholders (Π^C) and some leave the market, which creates deadweight loss (B). Consumers' maximum willingness to pay for new data is therefore $A - B - \Pi^C$ .The AI firm also benefits: its revenues increase by C compared to Panel A, and it is therefore willing to pay up to C for data access.

TOTAL GAINS FOR SOCIETY. Moving from the situation in Panel A to the situation in Panel B increases overall welfare by $A - B + C$. Everyone is better off when fresh data are supplied and creators are compensated. However, the costs fall more heavily on consumers than on the AI firm, since $A - B - \Pi^C > C$. In the European context, this means that enforcing copyright brings benefits to creators in the EU, but those benefits are mostly funded by EU consumers, rather than by (often foreign) AI companies.

WHAT THE FULL MODEL ADDS. More realistic assumptions and endogenous behaviour explain whether it's possible to get from the situation in Panel A to the situation in Panel B under a range of possible policy options: (1) Production side: how data volume translates into AI value. (2) Creator side: costs of creation and how creators respond to incentives (in a given policy framework). (3) Institutions: transaction costs for licensing and distributing income. (4) Policy options: EU-relevant regimes (opt-in, opt-out, statutory licensing, exceptions), compared in terms of (1) whether they generate enough new data to offset depreciation, and (2) which achieves this with minimal societal cost.

RESULTS. The analysis of the full model shows that a clear ranking of policy options in terms of total welfare exists. (1) Statutory licensing is on par or better than other approaches: it preserves incentives for new data creation and ensures a large flow of data, but is limited by transaction costs. (2) Opt-in can achieve a large enough flow of data, if the costs to perform opt-in and the perceived harm to creators are sufficiently low. (3) A broad exception ("TDM") without compensation cannot activate the flow of new data. (4) An exception with opt-out is worst because it not only activates the flow of new data, but also further reduces the amount of available data.

2.1. Model Setup and Economic Intuition

This section explains a simple but rigorous model that connects copyright policy for AI training data to outcomes that matter for society. The first outcome is how much useful data an AI system can access today, which we call the stock of works. The second outcome is whether there are strong incentives to produce new works tomorrow, which we call the flow. The third outcome is how the benefits created by AI are shared between users, AI firms, and creators, which we refer to as welfare.

The model keeps the number of moving parts small on purpose. Each part is tied to a symbol that appears in a short formula. The goal is not to capture every legal detail but to isolate the few forces that determine which policy works best from a societal perspective. Whenever we make a simplifying assumption, we explain why it is reasonable, what might be missing, and how conclusions would change if the assumption were relaxed.

2.1.1. How Data Turns into Value

AI services are more valuable when they can train on a broad and representative set of works and when these works are kept fresh by a steady flow of new creations. We capture this with

$$A = \theta(\sigma) D_S \mu^m, \mu = \{\frac{D_f}{\lambda D_S}, 1\}$$

Here A summarises how much users are willing to pay for AI output. The variable D_s is the stock of works already created. The variable D_f is the newly created flow per period. The variable λ is an obsolescence rate, determining how fast the stock of data loses informational value. The variable σ is the share of the entire stock (within a respective policy framework) is available for training. Through the function $\theta(\sigma) = \sigma^\gamma$, we model how representative the available share of the stock is of the entire stock. If γ is larger than one, leaving out parts of the stock hurts more than proportionally. The variable m tells us how sensitive AI value is to freshness. It is the key parameter to capture the importance of fresh data. The variable μ is the maintenance ratio. If $\mu = 1$, the flow is high enough to retain the value of the stock. If $\mu < 1$ the stock loses value.

This compact mapping captures two facts. Missing chunks of data reduce AI value of end users even when a lot of other data remains. A lack of fresh data reduces how valuable even the best old data is.[10]

2.1.2. How Prices and Quantities are Set

Users have standard downward sloping demand. We write it as

$$p(Q) = A - BQ$$

where p is price, Q is quantity, and B is a slope. This linear form is chosen because it lets us compute welfare in closed form. We allow the market structure that the AI firm finds itself in to be more or less competitive. We summarise this with a single conduct parameter κ between 0 and 1. Larger κ means more competitive behaviour and higher pass through of cost changes into prices.

Creators are compensated through a royalty rate r applied to the revenues of the AI firm. Equilibrium price and quantity can be written as

$$p(A,r) = \frac{1-\kappa}{1-\kappa r} A, \qquad Q(A,r) = \frac{\kappa(1-r)}{1-\kappa r}\frac{A}{B}$$

Gross sales, which are both firm revenue and the base for royalties, are

$$S(A,r) = pQ = \frac{A^2}{B}\rho(r), \qquad \rho(r) = \frac{\kappa(1-\kappa)(1-r)}{(1-\kappa r)^2}$$

[10] In this model set up, algorithmic progress and firms' retraining or architectural choices are treated as exogenous, so A changes only via coverage σ and freshness μ. Endogenising innovation would primarily rescale A (and likely lower the effective m), shifting the maintenance threshold and the welfare-maximising royalty rate quantitatively. Unless royalties exert a very strong negative elasticity on innovation, however, the welfare ranking is unlikely to change: broad-access, low-friction regimes with modest royalties (statutorily licensed access) remain preferred, with only extreme innovation sensitivities pushing the optimum toward a near-zero-royalty or exception corner.

NOTE: An ad valorem royalty, a percentage of sales, matches how many licensing arrangements are implemented in adjacent cultural markets. A single conduct parameter is a standard shortcut to represent how competitive the market is. If competition were stronger or weaker than assumed, only the scale of static price distortions would change, not the direction of our results. If demand were not linear, algebra would change, but the core trade-off would not. Policies that raise A through better coverage ($\sigma \rightarrow 1$) and better maintenance ($\mu \rightarrow 1$) tend to raise welfare, unless the royalty needed to fund maintenance becomes so large that static price losses dominate.

2.1.3. How The Royalty Pool Funds New Works and how Administrative Overhead is Created

Royalties build a pool that can restore incentives for the flow of new works. This pool is redistributed to creators, with an administrative overhead τ, through a centralised clearing and redistribution system. The net pool is

$$F = (1-\tau) r S(A, r)$$

In practice, transaction costs arises because most licensing at scale is handled by Collective Management Organizations (CMOs) that act as centralised hubs that aggregate uses, collect payments, match them to rightsholders, resolve disputes, audit reports, and remit funds across borders. Using CMOs can lower individual action and bargaining costs for millions of rightsholders and makes broad participation feasible, but it introduces an administrative overhead τ that reduces the net pool available for creators.

The cost of producing a flow D_f is modeled with a convex function $T(D_f) = \phi {D_f}^{\psi}$, which allows costs to rise faster when we push for more creation. Without funding, AI may lower private incentives to create, which pushes the flow down to a suppressed baseline ${D_f}^{AI}(0)$. An interpretation of this baseline is that this is the amount of works creators still produce even in the presence of AI and without being compensated for their works being included in training data. The royalty pool only needs to cover the gap back toward the level of data flow that ensures that the value of AI is maintained. We write that gap as

$$\Delta T(D_f) = \phi {D_f}^{\psi} - \phi \left({D_f}^{AI}(0)\right)^{\psi}.$$

2.1.4. How a Policy Framework can Change the Stock

The share of stock that is available for AI training, σ, depends on the policy framework as well as on action costs to seek rights clearing. Under opt in, the AI firm needs to locate rightsholders and clear usage. Under opt out, a creator must act to remove a work, which can be costly. In very large populations even tiny action costs reduce participation dramatically. This is well known in many other markets with default rules. We represent this with a smooth function that maps net benefits into a share between zero and one. The details of the function are not very important for our conclusions. What matters is that opt-in tends to yield low σ and opt-out tends to yield high σ, unless a collective or a platform actively reduces participation frictions.

2.1.5. Total Welfare

Total welfare aggregates the surpluses accruing to users, AI firms, and creators, and subtracts real resource costs. Transfers between actors (e.g. royalties) are neutral for welfare; what matters are static price effects, distributional consequences, and the resource costs of administering and sustaining creation.

Formally,

$$W = CS + \Pi^{AI} + \Pi^{C} - \Delta T(Df) - Admin,$$

with

$$CS = \frac{A^2}{2B}\frac{\kappa^2(1-r)^2}{(1-\kappa r)^2}, \ \Pi^{AI} = \frac{A^2}{B}\rho(r)(1-r), \ \Pi^{C} = \frac{A^2}{B}r\rho(r), \qquad Admin = \frac{A^2}{B}\tau\rho(r).$$

Here, CS denotes consumer surplus, Π^{AI} the operating profits of the AI firm, and Π^{C} the gross royalty income accruing to creators. The two resource costs are the administrative overhead from running the licensing system, and $\Delta T(D_f)$, the incremental cost of producing new works above the suppressed baseline.

FEASIBILITY: Before discussing the welfare effects of different stylised policy options, it is useful to first discuss the feasibility of funding new data in sufficient quantity to counter the value depreciation of the stock of data. There exists a small positive royalty that raises enough funds to restore a sufficient flow of data. At that point, D_f reaches λD_s and the value discount disappears. Raising the royalty beyond that point does not increase *A* further in the baseline model. The optimal royalty therefore sits at, or just above, the threshold that funds maintenance. Below the threshold, the flow is underfunded, and the usefulness discount persists. Above the threshold, the only effect of a higher royalty is a larger static price wedge, larger transfers to creators, and larger administrative overhead.

When administrative overhead is small and the AI market is reasonably competitive, the small royalty needed to reach the threshold creates a static loss that is modest. This resembles the stylised illustration of the model discussed at the top of Section 3.

2.2. Stylised Policy Options

We now describe four regimes that exhaust the main choices in the current debate. Each regime plugs into the same welfare accounting. This makes comparisons transparent and fair.

AI Exception

The entire stock can be used for AI training. Representativeness is maximal, since $\sigma = 1$ and $\theta(1) = 1$. Rightsholders are not remunerated ($r = 0$), and there is also no funding for new creation, since $F = 0$. The flow falls to the suppressed baseline ${D_f}^{AI}(0)$ and the value discount µm applies. This policy option is similar in spirit to Art. 3 DSM but without restrictions on the type of AI developer (scientific, non-commercial and commercial).

AI Exception with opt Out

Per default all stock can be used for AI training. Only works that creators actively remove are excluded. In many real implementations the royalty rate is zero, which implies no funding for new creation. The value discount therefore remains. This policy option is similar in spirit to Art. 4 DSM.

Licensing Market ("opt in")

The default is that AI developers cannot access data. Only works that creators actively allow to be included, e.g. through licensing agreements with AI developers, can be used for AI training. Coverage of the potential stock is therefore low unless participation frictions are near zero. A royalty may be negotiated in principle. In practice, the reduced coverage lowers A through $\theta(\sigma)$ and the pool is smaller because it is applied to a smaller revenue base.

Statutory Licence

The entire stock can be used for AI training. A regulated royalty r is charged, and the net pool $F = (1-\tau)rS(A,r)$ is used to fund new creation. If the pool reaches the maintenance cost gap, the flow is restored to $D_f = \lambda D_s$, which removes the value discount. The key choice for policy is the level of r. The relevant target is the smallest r that reliably funds a sustained value of AI. Alternative instruments such as direct subsidies or innovation prizes could in principle also sustain new creation, but they require sustained public budgets and detailed central information, whereas statutory licensing ties incentives automatically to market size.

2.3. Results

2.3.1. Optimal Policies from Different Perspectives

The model makes clear that different actors may have different views about what policy is "optimal." Formal proofs of the statements below can be found in Peukert (2025a).

From the Perspective of Creators

Creators focus on their net surplus, which depends on the size of the royalty pool F relative to the incremental cost of creation. A higher royalty rate raises per-unit compensation but also reduces the size of the market and hence the royalty base. Their surplus is therefore maximised at an intermediate rate: too low a royalty underfunds new creation, while too high a royalty shrinks demand so much that the royalty pool collapses. In practice, creators tend to favour statutory licences with substantial royalties, but the model highlights that even from their perspective, there is an internal trade-off.

From the Perspective of AI Firms

AI firms maximise operating profits, which are highest when royalties are minimised. Exceptions or no-royalty opt-out regimes deliver the best short-run outcome for firms because they keep training data abundant and payments low. However, such policies depress the flow of new works, which gradually erodes the quality and usefulness of the stock. The model thus predicts a divergence between firms' short-term incentives and their longer-run interest in sustaining data freshness.

From the Perspective of Overall Society

Society cares about total welfare, defined as the sum of consumer surplus, creator surplus, and firm profits, net of administrative overhead and the cost of restoring creation incentives. From this perspective, the optimal policy is typically a statutory licence with a modest royalty: just large enough to sustain the flow at $D_f = \lambda D_s$, ensuring representativeness and freshness while limiting static distortions. Over-compensation reduces welfare by driving up prices and excluding users. When administrative costs

τ are high or pass-through κ is low, the distortions of royalties rise and the welfare optimum can shift toward the corner, close to an exception regime.

Summary

The divergence between these perspectives explains why industry positions in the policy debate differ. AI firms favour exceptions or no-royalty opt-outs that maximise their short-term profits, even if they risk underfunding future creation. Creators prefer statutory licensing with higher royalties, though their own surplus is maximised at an intermediate rate rather than at the highest possible one. From a social planner's perspective, a statutory licence with a modest positive royalty is usually optimal, as it balances representativeness and freshness against static distortions. However, when administrative overhead is high or pass-through from firms to consumers is weak, the welfare-maximising policy can shift toward very low or even zero royalties.

2.3.2. Calibration and Policy Simulation

We next use the calibrated parameters to run simulation experiments across the four policy regimes. The goal is to illustrate how the model works in practice and how different choices of parameters shift welfare outcomes. It is important to stress that the numerical results are not policy advice. They should be read only as stylised illustrations of the mechanisms in the model.

Parameters and Data Sources

Table 1 reports our parameter choices to calibrate the model. Each of these values was anchored in empirical evidence where possible (see Appendix A.1). Where evidence was weak or ranges were wide, we chose transparent benchmark values. For ease of comparison with the underlying studies and industry data, values in the calibration exercise are expressed in $. This choice is not substantive: results would be identical if expressed in € or another currency, as outcomes scale proportionally under the assumptions described in Appendix A.

Simulation Results

Table 2 decomposes total welfare (W) into consumer surplus (CS), AI firm profits (Π^{AI}), creator profits (Π^{C}), the real resource cost of additional creation beyond the suppressed baseline ($\Delta T(D_f)$, shown as a negative), and administrative overhead ($Admin$, also negative). Each column is a policy regime evaluated either at the welfare-maximising choice or at the choice that maximises a particular actor's surplus (AI firms or creators). We also report these private optima to show how aggregate welfare would shift if policy (or bargaining) were driven solely by one group's interests rather than by efficiency. The shaded columns are the politically and economically most realistic counterfactuals: (i) statutory licensing with a welfare-maximising royalty (col. 3), (ii) a licensing market with balanced bargaining power $\beta = 0.5$ (col. 4), and (iii) an exception with opt-out (col. 6), which mirrors existing "TDM with opt-out" designs.

Under our baseline calibration, statutory licensing (col. 3) delivers the highest total welfare at about $98.8 billion per year. Consumers capture the vast majority ($96.7b), AI firms earn around $2.3b in revenue, and creators are paid royalties of roughly $2.2b. These creator payments just cover the additional creation costs (about $2.2b), leaving them with $3m extra net income. Relative to that, the administrative overhead is large with $243m.

It's important to stress that under statutory licensing calibrated to the maintenance threshold, creators' net income is relatively small only after the additional cost of producing new works has been covered. In other words, the system is designed to make creators whole: they receive exactly the compensation needed to sustain a steady flow of fresh, high-quality works. This ensures that incentives are preserved without overcompensation that would primarily raise consumer prices and administrative costs. The aim is not to minimise creator returns, but to balance them at the level that secures continued creation while maximising overall welfare.

The second-best realistic outcome is a licensing market with equal bargaining power (col. 4), which yields total welfare of about $84.6b. Compared to this regime, statutory licensing increases total welfare by roughly $14.2b (about 14%). Consumers are the main beneficiaries, as their surplus rises by nearly $42b (from $54.9b to $96.7b). AI firms also gain substantially, with profits increasing by about $1b (from $1.3b to $2.3b). Creators, by contrast, see their net income fall from about $32.2b to nearly zero, as statutory licensing limits royalties to the maintenance requirement. Importantly, statutory licensing also reduces administrative overhead by more than $3.5b (from $3.8b to $0.2b). These shifts illustrate the central economic logic of the model: royalties above the maintenance threshold primarily redistribute surplus to creators while raising prices for consumers and inflating transaction costs, thereby lowering total welfare.

For context, an exception without opt-out (col. 5) produces total welfare of $79.8b, about $19b below statutory licensing. Allowing opt-out (col. 6) reduces welfare further to $73.9b, as some creators exit without restoring incentives for fresh data. A market regime that maximises creator surplus (col. 7) performs worst, with total welfare of just $63.6b. In this case, large transfers to creators ($43.0b) come at the expense of consumer surplus, which collapses to only $25.0b, while administrative costs rise to about $5.0b.

Table 1: Exemplary calibration of model parameters

Parameter	Meaning	Value(s)	Source(s)
A, B	Demand intercept/slope	$A = 404,\ B = 834$	Brynjolfsson and Collis (2025)
κ	Pass-through rate	0.988	Calibrated to Brynjolfsson and Collis (2025), OpenAI revenues
δ_f	Flow reduction after AI	0.214	Peukert et al. (2025)
ϕ, ψ	Production cost	$\phi = 1.64,\ \psi = 1.001$	Aguiar and Waldfogel (2018a), US inflation data
λ	Depreciation rate	0.015	Peukert et al. (2024), Larivière et al. (2008), Cheng et al. (2024)
m	Maintenance elasticity	0.5	Assumed (domain sensitivity)
$x_{\text{in}}, x_{\text{out}}$	Opt-in/out frictions	$0.018,\ 0.010$	Shcherbakov et al. (2025)
γ	Representativeness curvature	2.88	Shcherbakov et al. (2025)
τ	Licensing overhead	0.10	EU CMO transparency reports

Table 2: Exemplary calibration results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Component	Market $\max(\Pi^{AI})$	Market $\max(W)$	Statutory $\max(W)$	Market $\beta = 0.5$	Exception no opt-out	Exception opt-out	Market $\max(\Pi^{C})$
W	98821.44	98821.44	98821.44	84620.99	79810.75	73936.35	63595.22
CS	96712.53	96712.53	96712.53	54925.26	77918.00	72182.92	25041.10
Π^{AI}	2349.29	2349.29	2349.29	1334.22	1892.75	1753.43	608.29
Π^{C}	2190.65	2190.65	2190.65	34367.78	0.00	0.00	45150.14
$\Delta T(D_f)$	-2187.62	-2187.62	-2187.62	-2187.62	0.00	0.00	-2187.62
Π^{C}_{net}	3.03	3.03	3.03	32180.16	0.00	0.00	42962.52
Admin	-243.41	-243.41	-243.41	-3818.64	0.00	0.00	-5016.68

Takeaway

Among realistic policy options, statutory licensing dominates. It secures maintenance-level creation at low administrative cost, keeps consumer access broad, and---in our calibration---generates about \$14b more total welfare than the next-best regime with balanced bargaining power.

Interactive Policy Playground

To avoid dependence on any single parameter set, we provide an open, web-based simulation tool accompanying this briefing: https://copyright-policy.streamlit.app/. The dashboard allows policymakers, researchers, and stakeholders to vary key inputs — including demand parameters, pass-through rates, licensing overhead, and creation cost elasticities — and see how the welfare-optimal policy changes. This transparency means that the conclusions reported here are robust not because of a single "correct" calibration, but they can be challenged across a wide range of plausible parameter values.

2.4. Extensions

2.4.1. Carve-outs for Scientific Research

An important extension of the model concerns whether non-commercial scientific research should enjoy special treatment under copyright policy. In practice, many jurisdictions already provide broader exceptions for research use, motivated by the idea that early-stage knowledge creation produces positive externalities that spill over into downstream innovation. We can make this intuition precise by modelling a research tier as an upstream actor whose access to data at $t = 0$ shifts the demand intercept A for commercial AI services at $t = 1$. Broader research access thus raises future consumer surplus and firm profits, but may undermine near-term incentives for creators unless downstream commercialisation is coupled with a funding mechanism.

The trade-off is straightforward. Granting a research exception today ($\sigma_R = 1$) improves representativeness and raises the stock of usable data. This shifts commercial demand upward at $t = 1$, magnifying the surplus block $g(\tilde{r})$ and increasing overall welfare. The cost is the governance or compliance burden $C_R(\sigma_R)$ and, if no additional mechanism is introduced, the absence of royalty funding to close the creation gap $\Delta T(D_f)$. By contrast, limiting research to the same inclusion rules as commercial actors ($\sigma_R = \tilde{\sigma}$) preserves creator incentives in the short run but leaves future surplus unrealised.

The efficiency of carve-outs for scientific research depends on three conditions. First, research productivity and spillover intensity must be high enough that expanding access materially shifts the value of AI. Second, the intertemporal weight must not be too small, so that policymakers value long-run gains. Third, the downstream funding gap must be credibly addressed. One instrument is a reach-through levy on research-derived commercialisation: because it only applies if and when knowledge capital is monetised (with some probability), distortions are limited to realised markets. When a substantial share of commercial applications builds on research outputs, administrative leakage is limited, and product-market competition is strong, even a light reach-through levy can restore maintenance-level funding without eroding static surplus.

The implication is that carve-outs for scientific research can be welfare-improving if paired with downstream mechanisms that link eventual commercialisation back to creator funding. This logic provides a rationale for policies such as temporary embargo windows, licensing rules that require fair and equal access to research outputs on reasonable terms, or earmarked levies on research-derived sales. In each case, the goal is the same: enable broad access for research today while ensuring that, when spillovers mature into commercial value tomorrow, creators are compensated at least up to the maintenance threshold.

2.4.2. The Role of Competition Policy

So far, we have analysed copyright policy in isolation. Yet, in practice, its effects depend critically on how competitive the downstream AI market is. Competition matters because it determines how strongly royalties are passed through to users and how much surplus firms can extract. If markets are competitive ($\kappa \to 1$), pass-through is smooth, margins are thin, and the static loss from a given royalty is limited. When markets are concentrated ($\kappa \to 0$), the same royalty creates larger distortions. This complementarity implies that copyright and competition policy work best in tandem: effective antitrust enforcement makes moderate royalties easier to justify, because the static cost of funding is lower. For statutory licensing to be most effective in practice, it should therefore be accompanied by both low administrative overhead in the licensing system and vigorous competition in the product market.

If firms collude tacitly or explicitly, the effective κ falls, static losses increase, and the range of royalties that enhance welfare narrows. The model thus predicts that competition policy and copyright policy reinforce each other, with each becoming more effective when the other is also in place.

2.5. Geopolitical Considerations

So far, we have treated copyright and competition policy as domestic instruments. In reality, most AI development is concentrated in the US, while many data subjects and users are in the EU or other jurisdictions. This creates geopolitical tension: foreign firms supply local markets, while local users may care disproportionately about the use and preservation of local data. An extension of our model illustrates how these forces interact.

EU users may place more value on AI systems trained on EU sources than on models relying on non-EU data. In this case, EU welfare depends strongly on coverage and maintenance of EU works. A royalty that finances continued EU creation can be justified even if suppliers are foreign. If EU consumers attach higher value to local data, it is in the commercial interest of foreign firms to incorporate and maintain that data to remain attractive in the EU market. The more salient local data are to users, the stronger the case for policies that secure their preservation.

When the EU market is served by US firms, their competitive conduct shapes the welfare trade-off. If competition is vigorous, a moderate royalty to fund local creation entails only a small static loss. However, royalties only make sense if the funds collected translate into new creation at a reasonable cost. High administrative leakage or weak responsiveness of creators raise the threshold at which royalties improve welfare. Where clearance costs are low and creators respond strongly, royalties can be set more confidently; where costs are high and responses are muted, exceptions or subsidies may dominate.

The geopolitical lesson is that copyright policy cannot be separated from international market structure. For the EU, the case for royalties on local data strengthens with the salience of that data, but weakens if foreign AI markets are concentrated or administrative systems are leaky. Effective policy therefore requires coordination: copyright policy to fund creation, competition policy to keep markets open, and diplomatic or trade instruments to ensure that foreign firms supplying the EU both use relevant local data and comply with licensing rules.

Statutory licensing may also support the EU's digital sovereignty. Because most AI development occurs outside the EU, there is a risk that EU works could be underused, excluded, or incorporated without fair returns. A harmonised EU-wide framework would give creators predictable compensation and help ensure that EU content remains represented in global training datasets. It would also reduce fragmentation across Member States and provide foreign firms with a clear, transparent set of rules. While further evidence is needed, statutory licensing could complement other EU instruments, such as the Digital Markets Act (DMA) and the Digital Services Act (DSA), in shaping how global actors engage with the EU's digital and cultural resources.

3. PRACTICAL CHALLENGES

KEY FINDINGS

Applying our model for practical simulations requires empirical inputs that remain uncertain. Three dimensions are particularly important: how additional data translate into economic value, how strongly AI performance depends on a continuous flow of fresh data, and what the marginal cost of creative production looks like. Each shapes the optimal policy choice, yet reliable evidence is scarce. Regulators could help close these gaps by mandating disclosures from AI firms, supporting systematic benchmarking, and facilitating researcher access to royalty and earnings data. For both AI firms and creators, greater transparency is in their own interest, as it helps align regulation with technological and market realities.

The analysis also shows that implementation error in setting royalties is asymmetric. Overshooting the optimal fee is relatively benign, with welfare losses only of second order. Undershooting, by contrast, directly reduces creator funding and leads to first-order welfare losses. Statutory licensing remains preferable to opt-in as long as royalties are set within a feasible interval around the maintenance-restoring threshold. This interval narrows when markets are concentrated, administrative leakage is high, or creative output is costly, and it widens when competition is strong, administration lean, and data freshness highly salient.

The policy implication is clear: regulators should err slightly on the side of higher royalties, keep administrative overhead low, and invest in evidence generation to reduce uncertainty around the core parameters that determine optimal policy.

The model offers a transparent framework for comparing copyright policy regimes, but applying it in practice requires empirical inputs that are difficult to obtain. Three questions are particularly important: how data map into economic value, how strongly value depends on the flow of fresh data, and what the cost curve of creative production looks like. Each of these dimensions shapes the optimal policy choice, yet all face significant measurement challenges. Regulators could help close these gaps by requiring systematic reporting from firms or by facilitating data-sharing arrangements with researchers. For AI firms and the creative sector, transparency may be in their own interest: providing credible evidence on how value is generated, and what the cost of creation is, can help ensure that regulation is aligned with technological and market realities.

3.1. Data Limitations

3.1.1. How is Data Mapped into Value?

At the core of the framework lies the link between the stock of data and the usefulness of AI systems. Conceptually, broader and more representative datasets increase the value of AI outputs, but the precise mapping is poorly understood. Scaling laws suggest diminishing returns, yet these estimates depend on model architecture, task, and training regime. In practice, policymakers lack reliable evidence on how much an additional unit of data contributes to consumer surplus. Without this mapping, it is difficult to determine whether royalties that fund more creation generate real welfare gains or simply redistribute surplus.

A promising path forward would be structured disclosure by AI firms on training data usage, combined with systematic benchmarking. Regulators could mandate or incentivise the release of model performance metrics as a function of training data subsets. Independent researchers could design "data ablation" studies—removing or adding specific types of data to observe marginal value—and run controlled competitions that standardise such tests across firms.

3.1.2. How Important is Data Flow for Value?

The model also requires an estimate of how quickly data depreciate and how sensitive AI performance is to the arrival of new material. In fast-moving domains such as news, advertising, or real-time platforms, the half-life of value is short, making continuous inflows of data critical. In slower-moving domains such as literature or scientific research, depreciation is far less severe. The elasticity of maintenance, denoted by m, governs how much welfare is lost when fresh data fall short. Calibrating m is challenging: it varies across sectors and may evolve as models become more capable of generalising from existing corpora. This uncertainty directly affects the size of the welfare gains from statutory licensing or other interventions.

Here, regulators could play a coordinating role by collecting firm-level evidence on how frequently models are retrained, which domains see rapid decay, and what fraction of model performance depends on recency. Researchers could complement this with longitudinal studies comparing model accuracy against evolving benchmarks, or natural experiments that exploit sudden shifts in data availability (for example, platform shutdowns, changes in copyright enforcement, or major news cycles).

3.1.3. What is the Shape of the Cost Function of Creations?

Finally, optimal policy depends on how costly it is to produce new creative works. If marginal costs rise steeply, even modest royalties may suffice to elicit additional output. If costs are flat or declining due to digital tools, much higher royalties might be required to restore maintenance. Empirical evidence on these cost curves is scarce. Most studies focus on average revenues or labour income in creative industries, but what matters here is the marginal responsiveness of creation to monetary incentives. Measuring this requires careful natural experiments or detailed microdata, which are rarely available. Without credible estimates of the cost function, it is difficult to know whether policy-induced transfers are sufficient to sustain creation or whether they risk overshooting and generating large deadweight losses. To improve the evidence base, regulators could facilitate anonymised access to royalty distribution data from collective management organisations or platform-level creator earnings. Researchers could build on this by analysing quasi-experiments such as changes in payment schemes (e.g., Spotify's new "artist-centric" model, YouTube's revenue-sharing adjustments) or variations in grant and subsidy programs. Large-scale surveys of creators that link reported effort, costs, and revenues to observed market outcomes would also be valuable, especially when combined with administrative tax or income records.

3.2. Notes on Enforcement

Our model shows: if AI quality deteriorates as data become outdated, a phenomenon known as "concept drift" in the machine learning literature, and if users value model performance, then it is in the interest of AI firms to maintain a steady inflow of fresh, high-quality data. Aligning incentives through appropriate compensation mechanisms can therefore benefit creators, AI firms and end users in the long run. Ensuring compensation, through a statutory licensing regime, is therefore a common long-term interest, which should simplify enforcement across jurisdictions.

The idea of statutory licensing as a solution to balance the interests of rightsholders, providers of innovative technology and consumers / end users is not new. It is useful to revisit the arguments historically brought against statutory licensing, as exemplified by the court's opinion in a case of rightsholders versus a file-sharing platform (A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004 (9th. Cir., 2001)), in the context of AI:

- *Argument:* Statutory licensing would allow the infringer to escape penalties for injunction violations, statutory damages, or criminal liability.
  Counterargument: This is not a general argument against introducing statutory licensing. It seems very possible to introduce a statutory licensing regime only for future works, making it independent of the enforcement of rights on existing works.
- *Argument:* The infringer would gain the option to continue operations by simply paying royalties instead of facing consequences for infringement.
  Counterargument: The empirical evidence discussed in section 1 shows that AI is very different from historical cases of online piracy, as it creates large net benefits for society while using copyrighted works as input. Hence, a regime that allows to "continue operations" is in the best interest of society.
- *Argument:* Copyright holders would be compelled to engage with a company that has profited from their intellectual property without consent. Rightsholders would lose the ability to refuse licensing their works to the infringer.
  Counterargument: If AI outputs are biased when trained on non-representative data, a system that structurally allows withdrawal of data risks making training data non-representative and therefore introducing bias into AI outputs. When a rightsholder exercises opt-out, they do not take into account that this decreases value for society. A compulsory royalty, set by an independent authority, internalises this externality.
- *Argument:* Statutory licensing would erode bargaining power and prevent copyright holders from negotiating the terms and value of any potential licence.
  Counterargument: Bargaining power is not desirable once we take all stakeholders into account. Reducing the bargaining power of rightsholders is beneficial for overall welfare. With large externalities, the welfare benefits of a compulsory licensing regime with a royalty rate set by an independent authority can be large. As shown in the simulation above, full bargaining power of rightsholders (column 7 of Table 2), compared to a statutory licensing regime with a welfare-optimal royalty rate (column 3 of Table 2), would decrease overall welfare by $35bn. Put differently, the benefits to rightsholders are smaller than the costs to AI firms and consumers.

3.3. Implementation Error: Setting the Optimal Licence Fee

Even if statutory licensing is desirable in principle, its performance hinges on choosing the right ad valorem royalty. A fee that is too low fails to finance maintenance of new works; a fee that is too high raises prices and administrative leakage. Let

$$g(r) = \frac{\kappa^2(1-r)^2}{2(1-\kappa r)^2} + \rho(r)(1-r) - \tau r\rho(r), \qquad \rho(r) = \frac{\kappa(1-\kappa)(1-r)}{(1-\kappa r)^2}$$

Static welfare at demand intercept A is $(A^2/B)g(r)$; creation costs are accounted for via $\Delta T(\cdot)$. The maintenance-restoring royalty $\overline{r}$ is the threshold at which the flow of new works just reaches λD_s, and the intercept caps at $\overline{A} = \theta(1)D_s$.

3.3.1. Tolerance Around the Optimal Fee

Suppose the regulator sets $r = \overline{r} + \varepsilon$ with $\varepsilon > 0$. Because coverage is already complete, the intercept A remains fixed at $\overline{A}$, and the welfare loss is quadratic in ε:

$$W^{stat}(r+\varepsilon) = W^{stat}(\overline{r}) - \frac{\overline{A}^2}{2B}|g''(\overline{r})|\varepsilon^2 + o(\varepsilon^2),$$

Statutory licensing continues to dominate opt-in provided

$$\varepsilon \leq \varepsilon^+_{max} \equiv \sqrt{\frac{2B\Delta}{\overline{A}^2|g''(\overline{r})|}},$$

where $\Delta = W^{stat}(\overline{r}) - W^{in,*}$ is the statutory advantage over the licensing market ("opt-in") at the correctly calibrated threshold. Overshooting is therefore locally benign: welfare losses are second order in the deviation.

By contrast, if the regulator undershoots and sets $r = \overline{r} - \varepsilon$, the flow of new works falls short. The resulting funding gap ΔF directly reduces maintenance and hence A, leading to a first-order welfare loss. A conservative bound is

$$\varepsilon \leq \varepsilon^+_{min} \equiv \frac{\Delta F_{max}}{(1-\tau)\frac{\overline{A}^2}{B}[\rho(\overline{r}) + \overline{r}\rho'(\overline{r})]},$$

where

$$\Delta F_{max} = \frac{\Delta\phi\psi\lambda^{\psi}D_s^{\psi}}{\left(\frac{2\overline{A}}{B}\right)g(\overline{r})m}, \qquad \rho'(r) = -\frac{\kappa(\kappa-1)(\kappa r - 2\kappa + 1)}{(\kappa r - 1)^3}.$$

Undershooting is therefore more dangerous than overshooting: welfare losses arise linearly through a shortfall in creator funding.

3.3.2. When does the Optimal Policy Switch?

Statutory licensing weakly dominates opt-in whenever

$$r \in [\overline{r} - \varepsilon^-_{max}, \overline{r} + \varepsilon^+_{max}].$$

This interval shrinks when markets are concentrated (low κ), when administrative leakage is high (τ large), or when creation is expensive (large ϕ, ψ). It expands when competition is strong, administration lean, and maintenance salient (m and γ large). These comparative statics highlight that implementation error is more tolerable in environments where coverage and freshness matter strongly to users and where product markets are competitive.

4. RECOMMENDATIONS AND CONCLUSIONS

This briefing has traced the evolution of copyright in the digital era, examined the economics of data and AI, and developed a model that compares alternative regulatory regimes. The overarching lesson is that copyright policy for AI must be judged by its effects on total welfare: the combined benefits to consumers, firms, and creators, net of administrative and real resource costs. The key priority is to safeguard the *flow* of new, high-quality creative data that sustains the long-term value of AI. Retroactive compensation for the existing *stock* of works is less important, because that stock is fixed.

4.1. Core Policy Insights

1. **Focus on flow, not stock**

   The main policy concern is dynamic efficiency. Abolishing copyright for training data would maximise short-run access, but it risks depressing incentives for future creation, as early evidence from stock-photo and content platforms already suggests. Conversely, policies that over-compensate creators raise prices for end consumers, create transaction costs, and reduce data access without generating additional human creative expression. The challenge is to find an institutional design that restores creation incentives up to the maintenance threshold, while avoiding excess transfers and administrative overhead.

2. **Statutory Licensing Appears Optimal**

   Among realistic options, a statutory licence with a modest royalty outperforms licensing markets, or copyright exceptions with or without opt-out. It secures wide access to the stock of data, funds enough new works to offset value depreciation of existing data and minimises transaction costs. Our calibration exercise suggests that statutory licensing generates about $14 billion more annual welfare than the next best alternative. Opt-out regimes are particularly problematic, as they shrink access without restoring incentives for fresh data. Licensing markets ("opt-in") can work in principle but suffer from high clearance costs and fragmented coverage of the available stock of data.

3. **End Users Fund Creator Incentives**

   A central distributional insight is that the benefits of statutory licensing are financed primarily by end users rather than by AI firms. While creators benefit from restored incentives and firms from sustained data quality, the incidence of royalties falls largely on end users. In the European context, this means that consumer welfare should be at the centre of the policy debate, rather than being treated as residual.

4. **Complementary Policies Can Matter**

   The effectiveness of copyright policy depends on the institutional environment. Strong competition policy ensures that royalties are passed through efficiently and that firms cannot absorb rents at consumers' expense. Lean and transparent collective management systems reduce administrative leakage. Carve-outs for non-commercial research can be welfare-enhancing if paired with downstream mechanisms (e.g. levies on commercialisation) that reconnect eventual surplus back to creators. Ensuring high-quality AI has an international component: when data subjects are in one jurisdiction and model developers in another, policies must ensure that local data are preserved and that local creators are compensated to benefit from global innovation.

4.2. Recommendations for Policymakers

Based on these insights, we offer four practical recommendations:

1. Avoid opt-out and adopt statutory licensing as the default framework. It is the most robust option for aligning private incentives with societal welfare. The royalty rate should be set at the lowest level that restores maintenance of creative supply, and it should be periodically reviewed.
2. Prioritise lean and transparent administration. Before expanding the role of intermediaries, policymakers should carefully assess technical solutions that can automate authenticity proof and royalty allocation. Emerging standards such as the Coalition for Content Provenance and Authenticity (C2PA) or a training-data analogue to YouTube's ContentID demonstrate that automated tools can drastically reduce clearance costs and improve attribution. Where these solutions are not yet feasible, CMOs provide a necessary fallback: they aggregate rights, handle cross-border transactions, and make participation possible for millions of creators. However, reliance on CMOs should be recognised as a second-best option: their higher administrative costs and limited transparency can dilute incentives. Regulation should therefore set strict standards for lean operations, disclosure, and auditing, while encouraging the gradual uptake of automated systems as they mature.
3. Integrate copyright with competition and science policy. This increases the welfare gains from licensing by ensuring that royalties are passed through efficiently and that dominant firms cannot absorb rents at consumers' expense. Copyright policy should be seen in relation to digital market regulation, such as the DMA and DSA. Carve-outs for scientific research can be beneficial if coupled with mechanisms that ensure downstream commercialisation feeds back into creator incentives.
4. Address measurement gaps through disclosure and evidence. Regulators should consider mandating systematic reporting on training data, retraining frequency, and royalty distributions. This would allow policymakers and researchers to calibrate market simulations more precisely and adjust rules and optimal policies to evolving market realities.

Comparison with Related Study

These recommendations align closely with those of the European Parliament study on "Generative AI and Copyright: Training, Creation, Regulation" (Lucchi, 2025), which focuses on the legal aspects. Both analyses recommend against opt-out as a basis for commercial GenAI training and converge on a statutory EU-wide remuneration mechanism implemented through collective management. Both emphasise that transparency must be backed by verifiable audits and that item-level tracking is unrealistic, advocating data-driven allocation instead. The difference lies in emphasis: the Parliament study anchors its recommendation in legal coherence, calling for a statutory exception with non-waivable equitable remuneration, while our analysis anchors in welfare optimisation, showing that any statutory framework calibrated to the maintenance threshold maximises total welfare. Our model adds quantitative calibration and tolerance bands for royalty-setting; the Parliament study complements this with institutional design proposals such as the Three-Pillar Accountability Test and an AI & Copyright Unit. Taken together, the two studies provide a coherent policy roadmap: fund the *flow*, keep access broad, and ensure lean, auditable, and enforceable administration.

4.3. Conclusion

The debate on copyright and AI is often framed as a zero-sum conflict between creators and technology firms. Our analysis shows that this view is misleading. The majority of welfare gains from AI accrue to consumers, and policy should be designed to preserve these gains while ensuring that creators continue to supply the fresh, high-quality data on which future progress depends. Statutory licensing, carefully designed and implemented, offers a pragmatic compromise: it secures broad access, sustains incentives, and minimises costs. In doing so, it provides a viable path for Europe and beyond to balance cultural production with technological innovation in the age of AI.

REFERENCES

- Adermon, A., and Liang, C.-Y. (2014). "Piracy and Music Sales: The Effects of an Anti-Piracy Law." *Journal of Economic Behaviour & Organization*, *105*, 90–106.
- Aguiar, L., and Chen, Z. (2024). "Let that sync in: The effect of music reuse on product discovery." *Working Paper*, cESifo Working Paper No. 11249, Posted: August 13, 2024.
- Aguiar, L., Claussen, J., and Peukert, C. (2018). "Catch me if you can: Effectiveness and consequences of online copyright enforcement." *Information Systems Research*, *forthcoming*.
- Aguiar, L., and Martens, B. (2016). "Digital music consumption on the internet: Evidence from clickstream data." *Information Economics and Policy*, *34*, 27–43.
- Aguiar, L., and Waldfogel, J. (2018a). "Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music." *Journal of Political Economy*, *126*(2), 492–524.
- Aguiar, L., and Waldfogel, J. (2018b). "Streaming reaches flood stage: Does spotify stimulate or depress music sales?" *International Journal of Industrial Organization*, *57, 278-307*.
- Ahmed, N., Wahed, M., and Thompson, N. C. (2023). "The growing influence of industry in ai research." *Science*, *379*, 884–886.
- Athey, S., Mobius, M., and Pal, J. (2017). "The impact of news aggregators on internet news consumption." *Working Paper*.
- Bairathi, M., Lambrecht, A., and Rao, A. (2024). "Social media, music consumption, and cross-platform spillover effects." *Working Paper*, available at SSRN: https://ssrn.com/abstract=4959753, Posted: October 24, 2024.
- Batikas, M., Claussen, J., and Peukert, C. (2019). "Follow the money: Online piracy and self-regulation in the advertising industry." *International Journal of Industrial Organization*, *65*, 121–151.
- Bergantiños, G. and Moreno-Ternero, J.D., 2025. Revenue sharing at music streaming platforms. *Management Science*, 71(10), 8097-8993.
- Bently, L. (1994). "Copyright and the death of the author in literature and law." *Mod. L. Rev.*, *57*, 973.
- Biasi, B., and Moser, P. (2021). "Effects of copyrights on science: Evidence from the WWII book republication program." *American Economic Journal: Microeconomics*, *13*(4), 218–60.
- Bick, A., Blandin, A., and Deming, D. J. (2025). "The rapid adoption of generative AI." Working paper.
- Brynjolfsson, E., and Collis, A. (2025). "AI's overlooked $97 billion contribution to the economy." *Wall Street Journal*.
- Brynjolfsson, E., Collis, A., Liaqat, A., Kutzman, D., Garro, H., Deisenroth, D., Wernerfelt, N., and Lee, J. J. (2023). "The digital welfare of nations: New measures of welfare gains and inequality." *Working Paper*, *31670*.

- Brynjolfsson, E., Eggers, F., and Gannamaneni (now Collis), A. (2018). "Measuring welfare with massive online choice experiments: A brief introduction." *AEA Papers and Proceedings*, *108*, avinash Gannamaneni now publishes as Avinash Collis.
- Brynjolfsson, E., Li, D., and Raymond, L. (2025). "Generative AI at work." *The Quarterly Journal of Economics*.
- Buccafusco, C., and Heald, P. J. (2013). "Do bad things happen when works enter the public domain?: Empirical tests of copyright term extension." *Berkeley Technology Law Journal*, 1–43.
- Burtch, G., Lee, D., and Chen, Z. (2024). "Generative AI Degrades Online Communities." *Communications of the ACM*, *67*(3), 40–42.
- Calzada, J., and Gil, R. (2020). "What do news aggregators do? evidence from google news in spain and germany." *Marketing Science*, *39*(1), 134–167.
- Caoui, E. H., and Galasso, A. (2024). "Fractional ownership and copyright licensing: Evidence from the music industry." *Working Paper*, available at SSRN: [https://ssrn.com/abstract=5016224, Posted: November 5, 2024](https://ssrn.com/abstract=5016224).
- Cheng, J., Marone, M., Weller, O., Lawrie, D., Khashabi, D., and Van Durme, B. (2024). "Dated data: Tracing knowledge cutoffs in large language models." *arXiv preprint arXiv:2403.12958*.
- Cheng, M. M., Ofek, E., and Yoganarasimhan, H. (2025). "The value of silence: The effect of UMG's licensing dispute with Tiktok on music demand." *Working Paper*, Harvard Business School and University of Washington, Current version: June 23, 2025.
- Cooper, A. F., Gokaslan, A., Ahmed, A., Cyphert, A. B., De Sa, C., Lemley, M. A., Ho, D. E., and Liang, P. (2025). "Extracting memorized pieces of (copyrighted) books from open-weight language models." In *Proceedings of the ICML 2025 Workshop on Reliable and Responsible Foundation Models*.
- Cottier, B., Rahman, R., Fattorini, L., Maslej, N., Besiroglu, T., and Owen, D. (2024). "The rising costs of training frontier ai models." *arXiv preprint arXiv:2405.21015*.
- Danaher, B., Dhanasobhon, S., Smith, M. D., and Telang, R. (2010). "Converting Pirates Without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy." *Marketing Science*, *29*(6), 1138–1151.
- Danaher, B., Hersh, J., Smith, M. D., and Telang, R. (2020). "The effect of piracy website blocking on consumer behavior." *MIS Quarterly*, *44*(2), 631 – 659.
- Danaher, B., Huang, Y., Smith, M. D., and Telang, R. (2014a). "An Empirical Analysis of Digital Music Bundling Strategies." *Management Science*, *60*(6), 1413–1433.
- Danaher, B., and Smith, M. (2014). "Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Movie Sales." *International Journal of Industrial Organization*, *33*, 1–8.

- Danaher, B., Smith, M. D., Telang, R., and Chen, S. (2014b). "The effect of graduated response anti-piracy laws on music sales: evidence from an event study in France." *Journal of Industrial Economics*, *62*(3), 541–553.
- del Rio-Chanona, M., Laurentsyeva, N., and Wachs, J. (2023). "Are Large Language Models a Threat to Digital Public Goods? Evidence from Activity on Stack Overflow." *arXiv preprint arXiv:2307.07367*.
- Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2023). "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality." *Working Paper*, Harvard Business School and Wharton School Research Paper, Last revised: September 27, 2023.
- Demirci, O., Hannane, J., and Zhu, X. (2025). "Who is AI replacing? the impact of generative ai on online freelancing platforms." *Management Science*.
- Duch-Brown, N., Peukert, C., and Reggiani, C., Schaefer, M. (2025). "Does AI displace or complement web traffic?" *Working Paper*.
- Doshi, A. R., and Hauser, O. P. (2024). "Generative AI enhances individual creativity but reduces the collective diversity of novel content." *Science advances*, *10*(28), eadn5290.
- Erickson, K., and Kretschmer, M. (2018). "This video is unavailable." *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, *9*, 75.
- EU Intellectual Property Office (2021). "Online copyright infringement in the European Union." *EUIPO Report, doi:10.2814/505158*.
- Fiil-Flynn, S. M., Butler, B., Carroll, M., Cohen-Sasson, O., Craig, C., Guibault, L., Jaszi, P., Jütte, B. J., Katz, A., Quintais, J. P., Margoni, T., de Souza, A. R., Sag, M., Samberg, R., Schirru, L., Senftleben, M., Tur-Sinai, O., and Contreras, J. L. (2022). "Legal reform to enhance global text and data mining research." *Science*, *378*(6623), 951–953.
- Flynn, J., Giblin, R., and Petitjean, F. (2019). "What happens when books enter the public domain? Testing copyrights under use hypothesis across Australia, New Zealand, the United States, and Canada." *University of New South Wales Law Journal, The*, *42*(4), 1215–1253.
- Frey, B.S. and Jegen, R., 2001. Motivation crowding theory. *Journal of Economic Surveys*, 15(5), pp. 589-611.
- George, L.M. and Hogendorn, C., 2020. Local news online: Aggregators, geo-targeting and the market for local news. *The Journal of Industrial Economics*, 68(4), pp.780-818.
- Geroski, P. A. (2000). "Models of technology diffusion." *Research Policy*, *29*(4-5), 603–625.
- Giorcelli, M., and Moser, P. (2020). "Copyrights and creativity: Evidence from Italian opera in the Napoleonic age." *Journal of Political Economy*, *128*(11), 4163–4210.
- Givon, M., Mahajan, V., and Muller, E. (1995). "Software Piracy: Estimation of Lost Sales and the Impact on Software Diffusion." *Journal of Marketing*, *59*(1), 29–37.

- Goldberg, S. G., and Lam, H. T. (2025). "Generative AI in equilibrium: Evidence from a creative goods marketplace." *Working Paper*.
- Handke, C., Guibault, L., and Vallb´e, J.-J. (2021). "Copyright's impact on data mining in academic research." *Managerial and Decision Economics*, *42*(8), 1999–2016.
- Hartmann, J., Exner, Y., and Domdey, S. (2025). "The power of generative marketing: Can generative AI create superhuman visual marketing content?" *International Journal of Research in Marketing*, *42*, 13–31.
- Heald, P. J. (2008). "Testing the over-and under-exploitation hypotheses: Bestselling musical compositions (1913-32) and their use in cinema (1968-2007)." *U of Chicago, Public Law Working Paper*, (234).
- Heald, P. J. (2014). "How copyright keeps works disappeared." *Journal of Empirical Legal Studies*, *11*(4), 829–866.
- Hoffmann, M., Boysel, S., Nagle, F., Peng, S., and Xu, K. (2025). "Generative ai and the nature of work." *Working Paper*, Harvard Business School and CESifo Working Paper Series, Last revised: April 30, 2025.
- Hong, S.-H. (2013). "Measuring the effect of Napster on recorded music sales: difference-in-differences estimates under compositional changes." *Journal of Applied Econometrics*, *28*(2), 297–324.
- Huang, H., Fu, R., and Ghose, A. (2023). "Generative AI and content-creator economy: Evidence from online content creation platforms." *SSRN Working Paper 4670714*.
- Hui, X., Reshef, O., and Zhou, L. (2024). "The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market." *Organization Science*, *35*(6).
- Jia, S., and Nagaraj, A. (2025). "Cloze encounters: The impact of pirated data access on llm performance." *Working Paper*, NBER Working Paper No. w33598, Last revised: April 7, 2025.
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. (2020). "Scaling laws for neural language models." *arXiv preprint arXiv:2001.08361*.
- Kretschmer, T., and Peukert, C. (2020). "Video killed the radio star? online music videos and recorded music sales." *Information Systems Research*, *31*(3), 776–800.
- Landes, W. M., and Posner, R. A. (1989). "An economic analysis of copyright law." *The Journal of Legal Studies*, *18*(2), 325–363.
- Lariviere, V., Eric Archambault, and Gingras, Y. (2008). "Long-term variations in the aging of scientific literature: from exponential growth to steady-state science (1900–2004)." *Journal of the American Society for Information Science and Technology*, *59*(2), 288–296.
- Lauinger, T., Szydlowski, M., Onarlioglu, K., Wondracek, G., Kirda, E., and Kruegel, C. (2013). "Clickonomics: Determining the effect of anti-piracy measures for one-click hosting." In *Proceedings of the NDSS Symposium, San Diego*.

- Lee, P. (2025). "Synthetic data and the future of AI." *Cornell L. Rev.*, *110*, 1.
- Lei, X., 2023. Pro-rata vs User-centric in the music streaming industry. Economics Letters, 226, p.111111.
- Lemley, M. A., and Casey, B. (2020). "Fair learning." *Tex. L. Rev.*, *99*, 743.
- Li, X., MacGarvie, M., and Moser, P. (2018). "Dead poets' property—how does copyright influence price?" *The RAND Journal of Economics*, *49*(1), 181–205.
- Lin, S. (2024). "From Creation to Caution: The Effect of AI on Online Art Market." *mimeo*.
- Liu, J.-H., Zhou, T., Zhang, Z.-K., Yang, Z., Liu, C., and Li, W.-M. (2014). "Promoting cold-start items in recommender systems." *PloS one*, *9*(12), e113457.
- Liu, X., Sun, T., Xu, T., Wu, F., Wang, C., Wang, X., and Gao, J. (2024). "SHIELD: Evaluation and defence strategies for copyright compliance in LLM text generation." In Y. Al-Onaizan, M. Bansal, and Y.-N. Chen (Eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 1640–1670, Miami, Florida, USA: Association for Computational Linguistics.
- Longpre, S., Mahari, R., Chen, A., Obeng-Marnu, N., Sileo, D., Brannon, W., Muennighoff, N., Khazam, N., Kabbara, J., Perisetla, K., et al. (2024). "A large-scale audit of dataset licensing and attribution in AI." *Nature Machine Intelligence*, *6*(8), 975–987.
- Lu, S., Rajavi, K., and Dinner, I. (2021). "The effect of over-the-top media services on piracy search: Evidence from a natural experiment." *Marketing Science*, *40*(3), 548–568.
- Lu, S., Wang, X., and Bendle, N. (2020). "Does piracy create online word of mouth? An empirical analysis in the movie industry." *Management Science*, *66*(5), 2140–2162.
- Lucchi, N. (2025). "Generative ai and copyright: Training, creation, regulation." Study PE 774.095, European Parliament, Policy Department for Justice, Civil Liberties and Institutional Affairs, Brussels, requested by the Committee on Legal Affairs (JURI).
- Lunney Jr, G. S. (2001). "The death of copyright: Digital technology, private copying, and the digital millennium copyright act." *Virginia Law Review*, 813–920.
- McCabe, M. J., and Snyder, C. M. (2018). "Open access as a crude solution to a hold-up problem in the two-sided market for academic journals." *The Journal of Industrial Economics*, *66*(2), 301–349.
- Minassian, A. (1997). "The death of copyright: Enforceability of shrinkwrap licensing agreements." *UCLA L. Rev.*, *45*, 569.
- Moreau, F., Wikström, P., Haampland, O. and Johannessen, R., 2024. Alternative payment models in the music streaming market: A comparative approach based on stream-level data. *Information Economics and Policy, 68*, p.101103.
- Nagaraj, A. (2018). "Does copyright affect reuse? Evidence from Google Books and Wikipedia." *Management Science*, *64*(7), 3091–3107.
- Nagaraj, A., and Reimers, I. (2023). "Digitization and the market for physical works: Evidence from the Google Books project." *American Economic Journal: Economic Policy*.

- Noy, S., and Zhang, W. (2023). "Experimental evidence on the productivity effects of generative Artificial Intelligence." *Science*, *381*(6654), 187–192.
- Oberholzer-Gee, F., and Strumpf, K. (2007). "The effect of file sharing on record sales: An empirical analysis." *Journal of Political Economy*, *115*(1), 1–42.
- Peukert, C. (2025a). "The Economics of Data for Generative AI: Trade-offs and Optimal Copyright Policy" *Working Paper*, available at SSRN: https://ssrn.com/abstract=5532959.
- Peukert, C. (2025b). "Copyright and the dynamics of innovation in Artificial Intelligence." In *Proceedings of the 58th Hawaii International Conference on System Sciences*.
- Peukert, C., Abeillon, F., Haese, J., Kaiser, F., and Staub, A. (2025). "AI and the dynamic supply of training data." *Working Paper*, available at SSRN: https://ssrn.com/abstract=4807979.
- Peukert, C., Claussen, J., and Kretschmer, T. (2017). "Piracy and box office movie revenues: Evidence from Megaupload." *International Journal of Industrial Organization*, *52*, 188–215.
- Peukert, C., Sen, A., and Claussen, J. (2024). "The editor and the algorithm: Recommendation technology in online news." *Management Science*, *70*(9), 5627–6482.
- Peukert, C., and Windisch, M. (2025). "The economics of copyright in the digital age." *Journal of Economic Surveys*, *39*(3), 877–903.
- Pollock, R. (2009). "Forever minus a day? calculating optimal copyright term." *Review of Economic Research on Copyright Issues*, *6*(1), 35–60.
- Poort, J., Leenheer, J., van der Ham, J., and Dumitru, C. (2014). "Baywatch: Two approaches to measure the effects of blocking access to The Pirate Bay." *Telecommunications Policy*, *38*, 383–392.
- Quinn, M., and Gutt, D. (2025). "Heterogeneous Effects of Generative AI on Knowledge Seeking in Online Communities" *Journal of Management Information Systems*.
- Reimers, I. (2019). "Copyright and generic entry in book publishing." *American Economic Journal: Microeconomics*, *11*(3), 257–84.
- Rob, R., and Waldfogel, J. (2006). "Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students." *Journal of Law and Economics*, *49*(4), 29–62.
- Rob, R., and Waldfogel, J. (2007). "Piracy on the Silver Screen." *Journal of Industrial Economics*, *55*(3), 379–395.
- Schaefer, M., and Sapi, G. (2023). "Complementarities in learning from data: Insights from general search." *Information Economics and Policy*, *65*, 101063.
- Shcherbakov, V., Dalaud, I., and Peukert, C. (2025). "AI needs better data than the law allows." *Working Paper*, available at SSRN: https://ssrn.com/abstract=5296505.
- Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., and Gal, Y. (2024). "Ai models collapse when trained on recursively generated data." *Nature*, *631*(8022), 755–759.

- Teutloff, O., Einsiedler, J., Kässi, O., Braesemann, F., Mishkin, P., and del Rio-Chanona, R. M. (2025). "Winners and losers of generative AI: Early evidence of shifts in freelancer demand." *Journal of Economic Behavior and Organization*, *235*, 106845.
- Towse, R. (1996). "Market value and artists' earnings." In: *The value of culture: On the relationship between economics and arts*, edited by Klamer, A. 96–107.
- Towse, R., 2000. Cultural economics, copyright and the cultural industries. Társadalom és gazdaság Közép-és Kelet-Európában/Society and Economy in Central and Eastern Europe, pp.107-134.
- Tushnet, R. (2009). "Economies of desire: Fair use and marketplace assumption." *Wm. & Mary L. Rev.*, *51*, 513.
- UK Intellectual Property Office (2016). "Online copyright infringement tracker – latest wave of research mar 16 – may 16." *Policy report*.
- Valavi, E., Hestness, J., Ardalani, N., and Iansiti, M. (2022). "Time and the Value of Data." *arXiv preprint arXiv:2203.09118*.
- Vuopala, A. (2010). "Assessment of the orphan works issue and costs for rights clearance." Tech. rep., European Commission, DG Information Society and Media, Unit E4 Access to Information.
- Waldfogel, J. (2017). "How digitization has created a golden age of music, movies, books and television." *Journal of Economic Perspectives*, *31*(3), 195–214.
- Watson, J. (2017). "What is the value of re-use? complementarities in popular music." *NET Institute Working Paper*, (17-15).
- Watson, J., MacGarvie, M., and McKeon, J. (2023). "It was 50 years ago today: Recording copyright term and the supply of music." *Management Science*, *69*(1), 351–376.
- Wlömert, N., and Papies, D. (2016). "On-Demand Streaming Services and Music Industry Revenues – Insights from Spotify's Market Entry." *International Journal of Research in Marketing*, *33*, 314–327.
- Wlömert, N., Papies, D., Clement, M., and Spann, M. (2024). "Frontiers: The interplay of user-generated content, content industry revenues, and platform regulation: Quasi-experimental evidence from youtube." *Marketing Science*, *43*(1), 1–12.
- Yilmaz, E. D., Naumovska, I., and Aggarwal, V. A. (2023). "AI-driven labor substitution: Evidence from Google Translate and ChatGPT." *Working Paper*.
- YouTube (2021). "Copyright Transparency Report." *https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-22_2021-7-1_2021-12-31_en_v1.pdf*.
- Yue, Y. (2020). "The effects of movie piracy on box-office revenue: an empirical analysis of the Chinese movie market." *Journal of Applied Economics*, *23*(1), 618–655.
- Zhao, J., Wang, T., Abid, W., Angus, G., Garg, A., Kinnison, J., Sherstinsky, A., Molino, P., Addair, T., and Rishi, D. (2024). "Lora land: 310 fine-tuned llms that rival gpt-4, a technical report." *arXiv preprint arXiv:2405.00732*.

- Zhou, E., and Lee, D. (2024). "Generative artificial intelligence, human creativity, and art." *PNAS nexus*, *3*(3), pgae052.
- Zimmerman, D. L. (2011). "Copyrights as incentives: Did we just imagine that?" *Theoretical Inquiries in Law*, *12*(1), 29–58.

A. TECHNICAL APPENDIX

A.1 Calibration

A.1.1. Economic Value of AI and Firm Conduct

We obtain an estimate of A (effectively with $r = 0$) using the underlying data of Brynjolfsson and Collis (2025). The incentive-compatible WTA experiment gives us parameters for the demand function. We estimate a linear demand function ($O = A - B\theta$, where θ is the share of AI users that are willing to keep AI at a given offer O) and obtain estimates $\hat{A} = 404.498$ and $\hat{B} = 834.151$.

We then calibrate our values of A and κ such that, at $r = 0$, we get the same consumer surplus estimate as Brynjolfsson and Collis (2025) and firm revenues close to \$2.25bn (OpenAI claims global annual revenues of \$12bn with about 15% of users in the US and a market share of ChatGPT of about 80%). The parameters we set are $A = 404$, $B = 834$ and $\kappa = 0.988$.

Since the data we obtained from Collis is based on surveys among US consumers, we could adjust it to the average WTA of a range of digital services by consumers in EU countries (namely, Germany, France, Belgium, Spain, Ireland and Romania) reported in Brynjolfsson et al. (2023). The average of these EU countries is 64% of the US average, with a broad range from 26% (Romania) to 82% (Belgium).

A.1.2. Change in Flow After AI

We calibrate the parameter δ_f using data from a natural experiment on the stock photography platform Unsplash (Peukert et al., 2025). The authors estimate that the flow of data (uploads to the platform) decreased by

$$\delta_f \equiv 1 - \frac{D_f^0 - D_f^{AI}(0)}{D_f^0} = 1 - 0.786 = 0.214,$$

when Unsplash introduced a data-sharing program for commercial AI research. Specifically, the estimate of δ_f comes from the relative difference in uploads between users whose photos were included in AI training and those who were not, using the elasticity reported after August 2022 (when generative AI services became available; see column 2 of Table 5 in Peukert et al. (2025): $0.786 = (2.9038 - 1.0056 - 0.4054)/(2.9038 - 1.0056)$).

A.1.3. Production Cost

To calibrate the cost parameters, we use estimates of the fixed cost of music production from Aguiar and Waldfogel (2018a). The authors estimate that, in 2011, the fixed cost of producing a song was \$1.14. With US inflation data, we can express this in 2025 units as $\$1.14 * 1.436 = \$1.64$. To mimic the fixed cost per unit nature (i.e., near-constant marginal cost) of this estimate, we set $\phi = 1.64$ and $\psi = 1.001$.

A.1.4. Maintenance Parameters

To calibrate the depreciation rate λ, we draw on empirical evidence from both high- and low-drift domains. In high-drift environments such as online news and advertising, data value decays extremely quickly: for example, in news recommendation the half-life of click value has been estimated at only 17 hours (Peukert et al., 2024). By contrast, in low-drift environments such as scientific knowledge or large language model (LLM) training corpora, useful information persists for years. Bibliometric studies find that the median citation half-life in many fields is between five and ten years (Lariviere et al., 2008), while recent analyses of LLM knowledge cut-offs show that models trained on data up to 2019 remain useful for general tasks in 2025, implying an effective half-life of three to five years (Cheng et al., 2024).

Formally, if the empirical half-life of usefulness is H (measured in clock time) and the model period length is P, then after H/P periods the stock is halved:

$$(1-\lambda)^{H/P} = \frac{1}{2} \Rightarrow \lambda = 1 - 2^{-\frac{P}{H}}.$$

Equivalently, in continuous time with $D(t) = D(0)e^{\delta t}$ one has $\delta = ln\,2/H$ and the discrete per-period depreciation is $\lambda = 1 - e^{\delta P} = 1 - 2^{-P/H}$.

This mapping allows direct calibration of λ from observed half-lives. With monthly periods, half-lives of $H = 36 - 60$ months (3–5 years, as in LLM cutoffs) imply $\lambda \approx 0.019 - 0.012$, while $H = 60 - 120$ months (5–10 years, as in scientific citations) imply $\lambda \approx 0.012 - 0.006$.

It is interesting to note that such magnitudes also coincide with the growth rate of content platforms. In Peukert et al. (2025), the rate of new uploads to Unsplash is quantified as 0.52 over a period of 3 years, which implies a monthly growth of 0.014. The maintenance elasticity *m* governs how sensitively value responds to shortfalls in the flow of new data. When *m* > 1, recency is critical: even small shortfalls in fresh data strongly reduce the value of the stock, as in domains where user preferences or external conditions shift rapidly. When *m* < 1, by contrast, partial maintenance is already sufficient to preserve most of the stock's usefulness, which better reflects environments where knowledge is more stable and incremental updates suffice. In this sense, *m* captures whether maintenance is a strong complement to the stock or instead provides diminishing returns. Based on this reasoning, we chose $\lambda = 0.015$ and $\mu = 0.5$ as default parameter values to calibrate the model.

A.1.5. Cost of opt-in and opt-out

We normalise the decision scale to the total stock of available data D_s, so that the latent index for each regime is D_s times the net benefit of taking action. With action-side logits, the participation equations are

$$\tilde{\sigma}^{in} = \Lambda(D_s[r_{in} + \delta - x_{in}]), \qquad \tilde{\sigma}^{out} = \Lambda(D_s[\delta - r_{out} - x_{out}]),$$

with $\Lambda(z) = 1/(1 + e - z)$.

For a given scale D_s, these two observed participation rates pin down the required friction parameters x_{in} and x_{out}:

$$x_{in} = (r_{in} + \delta - \frac{1}{D_s} logit\ (\tilde{\sigma}^{in})), \qquad x_{out} = (\delta - r_{out} - \frac{1}{D_s} logit\ (1 - \tilde{\sigma}^{out})),$$

where $logit\ (p) = \ln\,(p/(1-p)$.

We calibrate the model using observed participation shares reported by Shcherbakov et al. (2025). In their empirical analysis of the LAION dataset, approximately 7 in 10,000 rightsholders actively opted in, while about 2 in 100 exercised the right to opt out. We interpret these as $\sigma^{in} = 0.0007$ and σ^{out}= 0.02, respectively, while at time of the analysis it is safe to assume that $r_{out} = 0$ and $r_{in} = 0$. Hence, we obtain $logit(0.0007) \approx -7.264$, $logit(0.02) \approx -3.892$ and

$$x_{in} = \delta + \frac{7.264}{404.498} \approx \delta + 0.018, \qquad x_{out} = \delta + \frac{3.8921}{404.498} = \delta + 0.010.$$

From this formulation, it becomes clear that the net effect of AI (harm/benefit) is simply a shift parameter. For calibration, we set $\delta = 0$.

A.1.6. Representativeness after opt-in and opt-out

To obtain a calibration of γ, the parameter that governs the curvature of representativeness loss, we exploit the empirical relationship between the percentage of works and the percentage of semantic clusters in which these works are best described, as documented in Shcherbakov et al. (2025) (Fig. 3, Panel A). Let $r \in (0,1]$ denote the percentage of works (the fraction of works up to a given percentile), and let $S(r) \in (0,1]$ denote the corresponding cumulative share of clusters. Following a Pareto-type approximation, the mapping between the two can be written as

$$lnr = \gamma \, lnS(r) + \varepsilon(r),$$

so that γ corresponds to the slope in a log–log regression without intercept. Intuitively, γ measures how quickly additional mass concentrates in the upper tail of the distribution: $\gamma > 1$ indicates strong curvature (high concentration), while $\gamma \simeq 1$ corresponds to proportional growth.

We implement this calibration in two steps. First, using shares across the entire LAION dataset $S(r)$ (blue curve in Fig. 3A in Shcherbakov et al., 2025), we estimate

$$lnr = \gamma_0 \, lnS(r) + \varepsilon_0(r).$$

Second, to account for licensing restrictions, we replace $S(r)$ with the cumulative licence-weighted share $S^{\ell}(r)$ (orange curve in Figure Fig. 3A in Shcherbakov et al., 2025) and estimate

$$lnr = \gamma_{\ell} \, lnS^{\ell}(r) + \varepsilon_{\ell}(r).$$

The difference $\Delta\gamma \equiv \gamma_{\ell} - \gamma_0$ quantifies the incremental curvature attributable to licensing frictions: if licensable material is disproportionately concentrated in the head or underrepresented in the tail, then $S^{\ell}(r)$ bends more strongly, yielding $\Delta\gamma > 0$.

Based on this data, we set $\gamma = 2.88$ as the default parameter value in our calibration.

A.1.7. Administrative Cost of Licensing System

To obtain a calibration of τ, the per-unit administrative cost of a licensing, we constructed an EU-wide dataset of collective management organisations (CMOs) and their administrative costs in three stages.

First, we assembled the population of CMOs in the EU by merging (i) official national registers maintained under the EU Collective Rights Management (CRM) Directive's competent authorities with (ii) pan-European membership directories of sectoral umbrellas (authors: GESAC; performers: AEPO–ARTIS; reprography/text: IFRRO; visual arts: EVA; audiovisual authors: SAA). We de-duplicated entities by Member State and mapped each society to a single primary sector (music authors, performers/neighbouring rights, reprography/text, visual arts, audiovisual authors, producers/publishers, or levy/retransmission). Second, for each CMO, we attempted to retrieve the most recent transparency report and/or annual accounts and extracted line-items capturing the organisation's administrative burden. We distinguish two constructs: *(a) management deduction/commission* (the pre-distribution skim retained from collections; terms include *Verwaltungskostenabzug*, *prelevement de gestion*, *descuento de gestion*), and *(b) gross administrative/operating expenditure* (the total bureaucracy cost recorded in the profit-and-loss statement; e.g., *Verwaltungskosten*, *charges de fonctionnement*, *gastos de gestion/administracion*). When both were disclosed, we recorded both and maintained their distinct meanings. We then computed (or adopted) an overhead percentage following a strict hierarchy: (1) the CMO's *published KPI* (e.g., cost-to-income, *Kostenquote*), preserving the KPI's stated denominator (collections only; collections plus financial result; or other base); (2) if absent, the *reported management-deduction rate* (deduction divided by collections); (3) otherwise an *implied ratio* defined as gross administrative expenditure divided by collections.

Finally, we normalised all absolute amounts to EUR. For reports in non-euro currencies (e.g., CZK, HUF, PLN), we converted using the relevant European Central Bank annual average rate for the reporting year.

We successfully obtained data (reflecting 2024) from 19 entities: ADAGP, AKM, Buma/Stemra, CEDRO, GEMA, GVL, IMRO, Kopiosto, OSA, PlayRight, Reprobel, SACEM, SGAE, SPEDIDAM, STIM, Teosto, VG Bild-Kunst, VG WORT. Compared to our full list of 163 CMOs, the CMOs for which we successfully obtained information, we have the strongest representation of Germany (40%), France (38%), Finland (33%), Spain (30%), and Belgium (29%). Our coverage is concentrated in music authors, reprography, and some performers, where we capture several major organisations. By contrast, film/AV, phonogram producers, visual artists, and press publishers are barely represented, and in most countries outside the top five, our coverage remains below 15%. While our dataset is strong on the large music and reprography CMOs in Western Europe, it leaves gaps in other repertoires and regions.

The average overhead across the covered CMOs is 10.1%. We therefore set $\tau = 0.1$ as the default parameter value in our calibration.

This in-depth analysis, commissioned by the European Parliament's Policy Department for Justice, Civil Liberties and Institutional Affairs at the request of the Committee on Legal Affairs, examines how copyright policy should respond to artificial intelligence (AI). It combines historical lessons from digital markets, insight on the economic value of data, and a formal model to study welfare effects. It assesses economic effects of various policy options, including an exception, an exception with opt-out, licencing market ("opt-out") and statutory licencing, in a search for optimal policy.

PE 778.859

IUST/2025/B/JURI/IC/022

Print ISBN 978-92-848-3114-2 | doi:10.2861/8762101 | QA-01-25-237-EN-C
PDF ISBN 978-92-848-3113-5 | doi:10.2861/0246137 | QA-01-25-237-EN-N