

# The Economics of Copyright in the Digital Age

**Working Paper****Author(s):**

Peukert, Christian  Windisch, Margaritha

**Publication date:**

2023-10

**Permanent link:**

<https://doi.org/10.3929/ethz-b-000639999>

**Rights / license:**

In Copyright - Non-Commercial Use Permitted

**Originally published in:**

Center for Law & Economics Working Paper Series 04/2023

# Center for Law & Economics Working Paper Series

Number 04/2023

## **The Economics of Copyright in the Digital Age**

**Christian Peukert  
Margaritha Windisch**

October 2023 (First version)

# The Economics of Copyright in the Digital Age

Christian Peukert<sup>1</sup> and Margaritha Windisch<sup>2</sup>

<sup>1</sup>HEC Lausanne

<sup>2</sup>ETH Zurich

October 25, 2023

## Abstract

Intellectual property rights are fundamental to how economies organize innovation and steer the diffusion of knowledge. Copyright law, in particular, has developed constantly to keep up with emerging technologies and the interests of creators, consumers, and intermediaries of the different creative industries. We provide a synthesis of the literature on the law and economics of copyright in the digital age, with a particular focus on the available empirical evidence. First, we discuss the legal foundations of the copyright system and developments of length and scope throughout the era of digitization. Second, we review the literature on technological change with its opportunities and challenges for the stakeholders involved. We give special attention to empirical evidence on online copyright enforcement, changes in the supply of works due to digital technology, and the importance of creative re-use and new licensing and business models. We then set out avenues for further research identifying critical gaps in the literature regarding the scope of empirical copyright research, the effects of technology that enables algorithmic licensing, and copyright issues related to software, data and artificial intelligence.

*Keywords:* Copyright, Digitization, Technology, Enforcement, Licensing, Business Models, Evidence

---

\*Peukert (corresponding author): christian.peukert@unil.ch, Windisch: margaritha.windisch@gess.ethz.ch. We are grateful to Stefan Bechtold, Christopher Buccafusco, Julia Cagé, Paul Heald, Jérémie Haese, Ulrich Kaiser, Tobias Kretschmer and Imke Reimers, as well as participants at seminars at ETH Zurich and the Munich Summer Institute 2022 for feedback and advice on earlier drafts.

## 1 Introduction

Intellectual Property (IP) rights are fundamental to how economies organize innovation and steer the diffusion of knowledge. Yet, while research in economics and management of innovation has largely focused on patents (Hall and Harhoff, 2012; Awaworyi Churchill et al., 2022), other IP rights, and copyright, in particular, have received considerably less attention.

Copyright law has continuously developed in response to emerging technologies and their challenges and opportunities not only for creators but also for intermediaries and consumers (Bently et al., 2010). For most of copyright history, storage, and distribution, technology for written works, sound recordings, photographs, and films were analog. With a broad scope and long protection terms, copyrights have played an essential but relatively passive role in the development of content industries, creating significant value for rightsholders but also a deadweight loss for society.

Digital technologies, perhaps starting with the introduction of the microchip and personal computers and much accelerated by the internet, have immensely challenged the copyright system (Greenstein et al., 2013). First, this led some jurisdictions to extend copyright law to new subject matters, for example, by adding protection for software and databases. Second, technology exacerbated the enforcement of *demand-side copyright infringement* through peer-to-peer file-sharing and unlicensed content platforms. Third, by substantially lowering the cost of creating, distributing, and promoting works, we have seen market entry at an unprecedented scale (Waldfoegel, 2017; Wu and Zhu, 2022; Aguiar et al., 2023). However, not all new works are original, but some represent the re-use of existing works which may be potentially infringing – in the following marked as *supply-side copyright infringement*.

In this paper, we provide a synthesis of the empirical literature on the law and economics of copyright in the digital age. First, we review important institutional cornerstones of copyright, focusing on the European Union (EU) and the United States (US), and also introduce the key developments in copyright law in the digital age. Second, we review the available empirical evidence, with particular attention to the economic impact of digital technologies and the varied interests of stakeholders in the digital economy. We structure our review around the issues of copyright infringement, enforcement, and digital business models and licensing. Third, having established what is known in the field, we continue to develop a research agenda. In particular, we highlight open questions about the heterogeneity and diversity of works and actors in the current copyright system. For example, evidence

is limited on economically significant application areas of copyright, such as software code, and with respect to the diverse socio-demographic backgrounds of creators. Further, as licensing technology continues to evolve, we argue that it becomes ever more critical to understand its effects on the involved stakeholders and overall welfare. We argue that copyright law plays a pivotal role in the next wave of technological innovation, acting as a gatekeeper concerning data and Artificial Intelligence (AI). For example, it will be important to study how (the absence of) copyright exceptions for training data affect AI innovation. As governments continue to regulate the digital world, copyright increasingly intersects with competition policy and privacy protection. Finally, we describe how digitization has brought researchers new tools and datasets that can help to answer important questions.

Our paper is an update and extension of prior overview work on the economics of copyright ([Landes and Posner, 1989](#); [Varian, 2005](#); [Towse, 2006](#); [Belleflamme and Peitz, 2010](#); [Handke, 2012](#); [Greenstein et al., 2013](#); [Belleflamme et al., 2014](#); [Peukert, 2019](#); [Windisch and Peukert, 2021](#)). The primary objective is to stimulate further research rather than simply providing a systematic literature review. We aim to consolidate the existing body of empirical evidence and put forward a research agenda to guide an informed discourse on the future of copyright law.

## **2 Foundations of copyright law**

### **2.1 Scope and terms of copyright**

Copyright law grants an exclusive bundle of rights to creators to control who can reproduce, distribute, adapt, perform, or exhibit their work. The leading international copyright treaty, the Berne Convention for the Protection of Literary and Artistic Works, adopted in 1886, serves as a baseline for most national copyright laws. Traditionally, copyright refers to the protection of all types of literary and artistic works that constitute the author's intellectual property, such as books, music, and movies. Nevertheless, due to technical developments for reproduction and copying and increasing economic globalization, the scope and degree of copyright protection have expanded over time. Since 1983, the US Copyright Act has protected computer programs (software) created by humans as literary works – provided they are original in the sense of being the author's intellectual creation. The EU legal framework for copyright and related rights, which now consists of two regulations and 11 directives, also has covered computer programs (source and object codes) since 2009 (Computer Programs Directive, 2009/24/EC). In comparison, technical effects and the software's functionality can only be protected by patent law, provided it is an invention, i.e., a product or process that offers a new way of doing

something or is a new technical solution to a problem. The EU additionally created a copyright protection scheme for electronic and non-electronic databases (Database Directive, 1996/9/EC). However, the scope extends only to the database structure – the “compilation” or “collective work.” Therefore, for the database to be protected, the materials (e.g., data) must be selected, coordinated, or arranged so that the resulting work in its entirety constitutes a new work. Furthermore, the individual elements must be arranged systematically or methodically and include the author’s intellectual creation.

Copyright is primarily justified as a means to provide monetary incentives for creating new works and the possibility to recoup the investments made by the creator in the creation process. Based on the utilitarian theory, mainly dominant in the US, the main goal of copyright is to promote social welfare by granting incentives for creation. The economic rationale is that design usually involves high fixed costs in the form of time and effort of the author, as well as the costs of editing, marketing, and publication. Similar to patent law, copyright law solves a public goods problem where creative works are under-produced due to the non-exclusive and non-rival nature of such goods. As ample empirical and anecdotal evidence suggests, monetary rewards are not the only incentive to create – intrinsic motivation is an important factor in spurring creativity. For example, copyright law systems in Continental Europe are often based on the labor theory by John Locke, according to which creators should be entitled to the fruits of their own (intellectual) labor, and the personality theory supported by Hegel and Kant, according to which the artists’ work reflects their personality. The importance of the latter is reflected by civil law countries, which grant authors an extensive set of so-called “moral rights” – the personality rights of an author (see [Towse, 2006](#) for a detailed discussion).

Empirical evidence provides support to the incentive rationale. At least at the extensive margin, exclusive rights are causally linked to innovative output. [Giorcelli and Moser \(2020\)](#) show that when certain regions of Italy received copyright protection due to Napoleon’s export of the French law, the quantity and quality of newly created operas increased substantially. Historical evidence, however, also suggests that copyright creates concentrated market structures where only the works of a few composers are regularly performed in the opera house ([Albinsson, 2021](#)). In the context of online publishing, [Li et al. \(2021a\)](#) show that better rights enforcement leads authors to increase their output. Copyright in creative works takes effect automatically upon their creation. As a general rule, the work’s author also owns the copyright in the original and derivative works. If two or more persons are involved in the creation process, they have a joint claim to the entitled rights unless the individual contributions

can be distinguished. Protection against unauthorized use in a particular country depends on the country's national laws. A registration or formality of the copyrighted works through an examination intellectual property office does not exist in the EU, in contrast to other IP rights. In the US, these works can be registered at the US Copyright Office to obtain statutory benefits within the US. For example, a registration or refusal of registration is required before an infringement suit may be filed in a US court and for claiming statutory damages.

The length of a copyright depends on the subject matter and jurisdiction – in most countries, it ends 70 years after the creator's death. Once copyright expires, creative works fall into the public domain and can legally be used without permission. Despite the public domain status of a work, however, moral rights may remain, which can lead to certain forms of use being inadmissible as an infringement of the author's personality rights. The current length is the result of historical term extensions. Evidence shows that, in particular, the UK Copyright Act of 1814 and the US Copyright Term Extension Act of 1998 have resulted in lower availability of works, higher prices of copyrighted content, and lower overall welfare because consumer surplus has decreased more than profits of copyright holders have increased (Heald et al., 2015; Li et al., 2018; Reimers, 2019).

## **2.2 Licensing and exceptions in copyright law**

Copyright owners frequently exploit their rights by granting licenses of some or all of their rights to others instead of assigning the copyrights to another party. License agreements allow licensees to distribute or modify the original work or create derivative works while the copyright owner keeps their copyright. First and foremost, this enables end-users to access and consume songs, movies, and other copyrighted works. Since creative activities often build upon existing works, either as a mere inspiration or by integrating parts into the new creation, creators might re-use (parts of) the work of others. For example, music recordings are often re-used and re-combined for sampling and remixing (Watson, 2017).

Evidence shows that the ability to re-use works critically depends on the availability and accessibility of copyrighted works. While the digitization of works facilitated re-use in general, copyrights and a lack of licensing often become an obstacle for future creators and end users. For example, Nagaraj (2018) shows that copyright restrictions drastically reduce citations to copyrighted works on Wikipedia, which results in lower-quality articles (as measured by page visits). Further, in a study of the market for Appropriation Art, which is characterized by varying degrees of legal uncertainty around copyright

infringement, [Cuntz and Sahli \(2023\)](#) show that higher perceived litigation risks decrease sales volume in art auctions and lead art dealers to relocate to jurisdictions with less legal uncertainty.

When comparing copyrighted works to those in the public domain, the monopoly power created by copyright results in higher prices and lower availability ([Li et al., 2018](#); [Flynn et al., 2019](#); [Buccafusco and Heald, 2013](#); [Heald, 2014, 2018](#)). Evidence also shows that the reverse is true: once copyright expires, works are more often performed ([Cuntz, 2023](#); [Watson et al., 2023](#)) and lower prices of accessing copyrighted works encourage follow-on innovation ([Biasi and Moser, 2021](#)). Frictions and inefficiencies in the licensing market might explain the conflicting results. Contacting and negotiating with rightsholders can be time-consuming, leading to high transaction costs for license acquisition ([Vuopala, 2010](#); [Akmon, 2010](#)). In some cases, the copyright owner cannot be identified, leading to so-called orphan works that cannot be licensed for use or re-use. The longer the copyright term, the more difficult it may be for potential licensees to locate the copyright owner. Because of a lack of empirical research, it is unclear, however, to what extent frictions in the licensing market and orphan works as an extreme case represent an economically important problem.

Copyright law also provides some exceptions that allow (limited) use of copyrighted works without the copyright owner's permission. Jurisdictions in the EU provide an exhaustive list of examples of fair use to restrict the copyright owner's rights scope, with exceptions, such as for private copying, scholarly work, parodies, news reporting, etc. The approach of the US is broader. Section 107 of the US Copyright Act identifies certain types of fair uses, such as criticism, comment, news reporting, teaching, scholarship, and research. It also provides a statutory framework for determining whether something is a fair use calling for four factors to be balanced: (i) purpose and character of the use, (ii) nature of the copyrighted work, (iii) amount and substantiality of the portion used in relation to the copyrighted work, (iii) effect of the use upon the potential market for or value of the copyrighted work. While fair use aims to promote the progress of science and useful arts, clear guidance on how to decide in fair use cases is missing, which has become more troubling as digital technology expands the universe of potential uses of copyrighted works ([Sag, 2005](#)). The doctrine is, therefore, often criticized as unpredictable and incoherent ([Beebe, 2007](#); [Goldstein, 2007](#); [Whitaker, 2019](#)).

The determination of copyright infringement has long been marked by confusion ([Balganesh and Menell, 2022](#)), especially in cases of re-use and the assessment of substantial similarity between two works ([Balganesh et al., 2014](#)). Empirical studies of litigation cases in the US and EU find that courts



often lack expertise in copyright matters, identifying unpredictable patterns of reasoning, concluding that litigation often differs from what legal experts in copyright matters would expect it to be (Cotropia and Gibson, 2013; Favale et al., 2015). In infringement cases, courts can rely on the opinions of experts in the specific context. Evidence from an online experiment where jurors are presented with expert testimony about the similarity of two songs shows that experts can indeed be helpful for jurors (Lee and Moshirnia, 2021), especially when jurors have little musical knowledge (see also Lund, 2011, Lund, 2013). However, an empirical study by Asay (2022) on substantial similarity decisions shows that courts rarely rely on experts in practice.

### 2.3 Digital market regulation and copyright

The copyright system has been under immense pressure in the digital age. The digitization of works such as music, movies, and books, starting in the late 1990s, enabled the lossless copying of existing works with different media and devices. The lower effort and costs of unlawful accessing, downloading, and copying copyrighted works further led to massive-scale copyright infringement on the *demand-side*, which affected revenues various creative industries, including software, music, and movies. 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 in music, movies, and software between 3.5% and 64% (Rob and Waldfogel, 2006, 2007; Hong, 2013; Yue, 2020). In response to decreasing revenues, creators and firms sought income in alternative business models, such as live concerts (Mortimer et al., 2012; Papies and van Heerde, 2017), freemium subscription services for software, music, and video (Haruvy et al., 2004; Aguiar and Waldfogel, 2018b; Lu et al., 2021) and by changing their IP strategies (Bradley and Kolev, 2023).

At the same time, cost reductions for the creation, distribution and promotion brought about by digital technology, the internet and open content platforms have enabled the massive-scale entry of works (Waldfogel, 2017; Wu and Zhu, 2022; Aguiar et al., 2023). Technology democratized the creation of content, giving a stage also to non-professional authors, musicians, software developers, etc. This resulted in new challenges to the copyright system because a large part of this so-called user-generated content may represent potentially unlawful re-use, creating issues of *supply-side* copyright infringement.

With most digital content being made available via intermediary online platforms, questions about

their liability regarding unlawfully uploaded content arise. Early on, internet service providers and content platforms were exempt from copyright liability, conditional on compliance with certain rules. The safe harbor provision in the US Digital Millennium Copyright Act (DMCA) of 1998 plays a vital role in dividing compliance burdens between online service providers and copyright owners. It intends to encourage online intermediaries and creators to cooperate closely to identify and remove copyright-infringing material. To do so, the DMCA opts for a Notice and Takedown (N&T) regime where copyright owners can send “takedown notices” to platforms requesting to remove infringing materials. Users can challenge the removal through a “counter-notice” that gives the target of the notice the ability to respond and ask a “putback”. The process avoids costly lawsuits unless a user challenges with a counter notice, in which case the rightsholder can continue the claim by filing an infringement suit. From the perspective of intermediary platforms, the DMCA can therefore be seen as a *passive N&T system*. Similarly, under the EU e-Commerce Directive (2000/31/EC), platforms are only liable for user-uploaded content if they become aware that the content is unlawful. The EU Directive on Copyright in the Digital Single Market (DCDSM) (2019/790) further addresses the transparency in relationships between copyright holders and online platforms, aiming for fairer remuneration for creators and rightsholders, particularly in the online environment and balancing bargaining power between creators and online platforms. Under Art. 17, perhaps the most controversial provision of the DCDSM, creates a legal basis for rightsholders to authorize the use of their works when uploaded by users and encompasses a series of “best efforts” obligations for the platform to obtain the authorization, ensure unavailability of specific protected content, and put in place N&T and notice and stay down mechanisms. As a result, it seems inevitable that platforms introduce *automated N&T systems* to prevent users from uploading copyright-infringing content. Certain types of internet service providers, such as mere quantity services, caching and hosting services, still maintain the traditional safe harbor principle.

The EU Digital Services Act (DSA), coming into effect in 2024, will replace the e-Commerce Directive to complement existing sector-specific legislation, such as the DCDSM. The DSA aims to improve content moderation on social media platforms to address concerns about illegal content by imposing additional rules for hosting providers and online platforms, depending on their size, and further defining their liability. This shows that N&T systems can also be understood in the broader context of content moderation beyond copyright enforcement. However, apart from legal requirements, there

can be economic incentives for platforms to moderate content. Theoretical work suggests that content moderation in multi-sided platforms supported by advertisement can increase users' engagement with ads (Jiménez Durán, 2022). Since a platform financed by advertising always wants to expand its user base, it is more inclined to moderate content, albeit less aggressively, than platforms based on a subscription model (Madio and Quinn, 2021). The sophistication of the technology used for content moderation, such as algorithms, also plays a role (Liu et al., 2022). This explains why platforms often offer private enforcement mechanisms that go beyond N&T and filtering mechanisms by additionally enabling licensed use. The prime example, which we will describe in detail below, is YouTube's "Content ID" program, which aims to detect infringing content automatically and then offers rightsholders the option to either block it or license it in exchange for monetary compensation. Such systems shift the burden of monitoring copyright infringement from copyright owners and courts to algorithms.

As we will discuss in detail below, technological advancements and the wide adoption of AI applications have raised new copyright questions, for example, because training data could be copyright infringing (Fiil-Flynn et al., 2022) or because AI output can be interpreted as a weighted recombination of input data, such as text, images, and recordings of voice and music, which may be copyright infringing.

### **3 Evidence on copyright in the digital age**

#### **3.1 Demand-side copyright infringement**

Traditionally, rightsholders can take action against copyright infringement through lawsuits. According to the US Copyright Act, rightsholders can seek statutory damages once a violation has been established without having to prove actual damages. In the EU, the rights holder can also bring claims for damages and cease and desist against the infringer within a given period, respective to the national law. The cases of Napster and other peer-to-peer file-sharing networks enabling unlicensed distribution of music in the early 2000s, however, illustrate the difficulty of identifying, prosecuting, and obtaining damages from massive-scale online copyright infringement (Landes and Lichtman, 2003; Electronic Frontier Foundation, 2008). The continued demand for online piracy (UK Intellectual Property Office, 2016; EU Intellectual Property Office, 2021) also suggests that individual-level lawsuits did not have the chilling effects that rights holders may have intended to create (see Eisend, 2019, for an overview of individuals' motivations for engaging in online piracy). 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). Evidence even shows that some consumers only discover the availability of online piracy websites because of media reports of a shutdown (Aguiar et al., 2018). One of the reasons why websites with unlicensed content exist is that running an ad-funded online piracy website can be a lucrative business. Hence, a relatively new enforcement approach is to “follow the money” and stop advertising revenues flowing to the operators of unlicensed websites (Circum, 2016). Evidence on the effectiveness of this type of enforcement is limited but sobering. A self-regulation effort led by the European Commission to convince the online advertising industry to stop funding copyright-infringing websites had limited effects, with only a small share of firms complying and no evidence of a decrease in demand for piracy websites (Batikas et al., 2019).

The available evidence on the effectiveness of public enforcement and government policies is mixed. Studies suggest that announcements and the introduction of stricter anti-piracy laws did not have a lasting effect on box office movie revenues in the US nor European countries (Orme, 2014; McKenzie, 2017; Bellégo and De Nijs, 2020). In the context of digital music, new anti-piracy laws in France and Sweden led to an increase in sales of about 25%, but this effect diminished after six months (Adermon and Liang, 2014; Danaher et al., 2014; Savelkoul, 2020). Evidence also points to distributional effects, suggesting that online copyright enforcement in the form of stricter laws and website shutdown disproportionately benefits more popular content (Peukert et al., 2017; Bellégo and De Nijs, 2020; Savelkoul, 2020).

As a result, theory suggests that when public enforcement is too weak, firms may find it optimal to invest in private enforcement to deter piracy (Banerjee, 2003; Sundararajan, 2004; Kiema, 2008; Ahn and Shin, 2010; Lu and Poddar, 2012). Evidence shows that private enforcement through negotiating settlement deals, which can be done at scale, can have positive welfare effects, both by reducing enforcement costs (e.g., legal fees, see Luo and Mortimer, 2017) and by increasing expanding the market (Luo and Mortimer, 2021). Software piracy, for example, has been largely curbed by the introduction of closed ecosystems, where applications can only be installed through a platform that is managed by the manufacturer (Miric and Jeppesen, 2020). However, the effectiveness of other technical enforcement measures is not transparent. By encrypting digital content and embedding digital watermarks, copyright owners mainly reduced the utility of those consumers willing to purchase digital content. When the music industry moved away from Digital Rights Management (DRM)

technology in 2007, music sales increased (Zhang, 2018). It is not clear whether this result holds in general, though. For example, in an empirical ex-ante policy analysis in the context of e-books, Kim and Leung (2021) show that eliminating DRM can only be profitable for a platform in the absence of pricing power of copyright holders.

Recognizing the huge difficulties of copyright enforcement in the digital age, some authors have proposed alternative policy instruments that ensure that producers' incentives to create products are sufficiently large (see Quintais, 2017 for a comprehensive overview). For example, Chen and Png (2003) use a theoretical model to compare the welfare implications of different governmental anti-piracy policies and conclude that a tax on copying is superior to fining consumers but also find that subsidizing original production is socially optimal. Furthermore, evidence from a discrete choice experiment shows that a copyright compensation system with a relatively low tax on internet access would keep revenues of rights holders fixed and simultaneously increase consumer surplus (Handke et al., 2016).

### 3.2 Supply-side copyright infringement

Despite their enormous practical importance, there is little empirical research on the effectiveness and effects of N&T systems. The available evidence focuses on the volume and determinants of takedown requests, suggesting that N&T systems can lead to over-enforcement, with implications for access to knowledge, freedom of speech, and justice (Gillespie, 2020). Several studies show that a small number of issuers is responsible for a large share of takedown requests received by platforms, which affects the accuracy of reports (Seng, 2015; Urban et al., 2017; Bar-Ziv and Elkin-Koren, 2018). The concentration of complainants appears to be related to the emergence of third-party services working on behalf of rightsholders. In the context of book publishers that use a third-party firm to find potentially infringing content online and file N&T notices on their behalf, e-book sales increase when infringing content is taken down. Still, the effectiveness varies with popularity, genre, and search frequency of individual titles (Reimers, 2016).

Large platforms often automate the N&T system through blanket licenses directly negotiated with rightsholders or collection societies<sup>1</sup> or by using algorithms that compare uploaded material to a database of original works. When the system detects a match, rightsholders can decide whether the content should be taken down or whether they want to receive monetary compensation. For example,

---

<sup>1</sup>see, e.g., Hiller (2016) and Kretschmer and Peukert (2020) for a discussion of music licensing on YouTube, and <https://www.musicbusinessworldwide.com/tiktok-and-universal-music-group-sign-global-licensing-deal/>.

re-use is at the core of the short-video platform TikTok. With the “sounds” feature, the audio track of existing videos can be used as a basis for a new video and with “duets”, users can put their new video next to an existing video (Valdovinos Kaye et al., 2021). Automated N&T systems have the potential to provide an efficient solution to massive-scale copyright enforcement, but they do not come without problems. While automated N&T regimes decide whether to approve or remove material based on content review and risk tolerance, the systems often include special takedown privileges for trusted senders, advanced profiling of users, restrictive terms of use, or modifications to search functions (Karaganis and Urban, 2015). Evidence shows differences in takedown rates between YouTube’s automated N&T system and the passive N&T system enabled by the DMCA. In particular, videos associated with film piracy and sports content are more often taken down, while video game publishers are largely not enforcing rights and tend not to request to take down gameplay streams (Gray and Suzor, 2020). Enforcement challenges are especially pronounced for live-streaming platforms, where copyright infringement response should ideally happen in real time. An immediate response requires additional resources to detect, analyze, shut down, and notify the streamer of the infringement, which is particularly challenging when live streams last several hours and the threshold above which a violation is considered to have occurred is only a few seconds.

Automated N&T is also challenging when it comes to exceptions in copyright law. For example, studies of parody videos on YouTube, an essential type of re-use that classifies as fair use in many jurisdictions, show that about 20% of videos are taken down, where production value and the genre of the original work are more important predictors of takedowns than the original’s commercial popularity (Erickson and Kretschmer, 2018; Erickson et al., 2019). The fact that derivative works are taken down is surprising in light of economic theory, which suggests that re-use can not only increase welfare but also benefit copyright hold (Arai and Kinukawa, 2014; Gans, 2015). Evidence in the context of music shows that re-use in the form of sampling and user-generated music videos can increase demand for original work, especially for less popular artists and novel re-uses of works who have not been widely re-used in the past (Watson, 2017; Schuster et al., 2019; Kretschmer and Peukert, 2020). Looking at millions of news articles, Cagé et al. (2020) show that copying is widespread in digital media, with news stories typically copied by other outlets within minutes. Original creators, however, receive more views, which seems to balance disincentives from de-facto weak copyright protection in this industry. In the context of designs for 3D printing, whether re-use enhances or decreases demand

for the original work is a function of the similarity between derivative works and original creations (Yilmaz et al., 2023).

It is important to recognize that automated N&T effectively allows online platforms to replace the legal use of copyrighted material with their own rules. The problem is that automated systems still lack an understanding of context, which often leads to the incorrect labeling of fair-use content or original material and the removal of allegedly inaccurate, defamatory, or misleading content that does not involve copyright at all (Bar-Ziv and Elkin-Koren, 2018). In addition, the non-transparent nature of automated filter mechanisms creates uncertainty, which can hinder follow-on innovation by affecting creative decisions and distribution strategies (Brøvig-Hanssen and Jones, 2021). Like in inefficient licensing markets, creators may innovate around existing ideas rather than seeking a license (Bechtold et al., 2015) or decide not to innovate at all.

### 3.3 Digital business models and licensing

Digital technology has not only enabled massive-scale copyright infringement but has also substantially reduced the cost of production, distribution, and advertising. As a result, the revenue needed to recoup up-front investments has decreased, countervailing the de-facto weakening of copyright. Since the beginning of the digital era, many industries have seen a tremendous increase in the supply of new products. The number of new songs, new films, new TV shows, new books, and comic books released each year has more than doubled (Waldfoegel, 2017; Kaiser et al., 2023), with high enough quality appreciated by experts and consumers (Waldfoegel, 2012; Aguiar and Waldfoegel, 2016, 2018a). As a reaction to the enforcement challenges with respect to *demand-side copyright infringement*, rightsholders have gradually shifted focus from fighting piracy to licensing content to digital services. Digital distribution technology, and corresponding business models, have arguably played their part in convincing consumers to consume licensed over unlicensed modes of consumption (Aversa et al., 2019). In what follows, we discuss the empirical literature on copyright licensing, both from the perspectives of licensing use to end-users, as well as licensing re-use to follow-on innovators. The latter seems particularly important in the context of user-generated content but also plays a major role in software development.

#### 3.3.1 Licensing use

The first broadly successful licensed digital music distribution service, iTunes, was introduced in 2003 (see Waldfoegel, 2010 for a detailed discussion). iTunes also pioneered the then revolutionary model of

unbundling songs from albums, which led to a steep decrease in the price per song and changed release strategies of record labels (Essling et al., 2017; Hiller and Walter, 2017). With advances in broadband technology and adoption, iTunes and other digital stores started to add movies and TV shows to their inventory. Evidence suggests that adopting such a business model played a role in decreasing online piracy. During a dispute between NBC and iTunes, which led to a temporary removal of NBC's TV shows from iTunes, demand for unlicensed copies (via the piracy technology BitTorrent) increased (Danaher et al., 2010). In a large field experiment run by Godinho de Matos et al. (2018), randomly selected households that previously used piracy technology received free access to a licensed service. Piracy usage only decreased for those households whose preferences (as measured by their consumption history) were more closely matched by the contents of the licensed service. Other consumers, with more obscure tastes, did not stop pirating, which suggests that the available inventory was not broad enough to convince users to switch to licensed content (see also Aguiar et al., 2018 who make the same point in a different empirical context). Online piracy is also related to catalog strategies of platforms. Cuntz and Bergquist (2022) find that broad availability of titles across video-streaming services is more effective in deterring online piracy and induces more investment in the quality of new movie productions compared to the exclusive distribution in one platform. In the context of books, Nagaraj and Reimers (2023) find that the availability of digital versions (via the Google Books project) can increase the sales of physical books. Sales increases occur more strongly for less popular books, with demand spilling over into non-digitized works by authors.

The introduction of on-demand streaming services has brought further changes to the revenue model of rightsholders (Hiller and Walter, 2017). Streaming platforms such as Spotify, YouTube and Netflix combine an extensive catalog of content into one bundle that consumers can access for free (subject to occasional advertising breaks) or by paying a subscription fee. This raises the question under which circumstances rightsholders would be willing to join such distribution platforms in the first place. Whether rightsholders are better off inside or outside the bundle varies substantially across different revenue-sharing agreements (Shiller and Waldfogel, 2013). This seems consistent with examples of artists such as Pink Floyd and Taylor Swift in dispute with streaming platforms about royalty rates (Waters et al., 2013; Ellis-Petersen, 2014), and continued disagreements between YouTube and the music industry (Sawers, 2016; Sweney, 2017; Hiller, 2016; Kretschmer and Peukert, 2020). However, evidence shows that – at least on average – on-demand streaming benefits rightsholders when compared



to the alternatives of transaction-based business models or online piracy. In the empirical context of Spotify’s entry into the music streaming market, [Aguiar and Waldfogel \(2018b\)](#) estimate that the lower revenues from selling less digital music in the transaction model (e.g., iTunes) are, on average, countered by the additional royalty income from streaming. In addition, licensed music consumption options, where consumers can choose between fee-based or free and ad-supported models, can reduce the demand for piracy ([Wlömert and Papiés, 2016](#); [Aguiar and Waldfogel, 2018b](#)). Evidence from the video streaming context suggests similar effects. For example, when Netflix was banned in Indonesia, search volume for pirated movies and TV shows increased by 20% relative to other countries, suggesting a net substitution of piracy for licensed services ([Lu et al., 2021](#)). Aggregate data further indicates 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](#)). The underlying reason may lie in differences in the release strategies across industries ([Smith et al., 2019](#)), where music is made available immediately on all distribution channels, but movies are first released in physical channels and only later through online media. Furthermore, licensing models can matter for the incentives to create. Evidence from online publishing shows that authors operating under quantity-based commission fees (similar to a royalty-based model) produce higher quality output compared to authors that operate with uniform commission fees (similar to upfront payments) ([Li et al., 2021b](#)).

### **3.3.2 Licensing re-use**

In the context of entertainment formats, such as music and video, platforms implementing automated N&T systems switched focus from pure enforcement to enabling a digital licensing market. For example, around the time YouTube launched the affiliate program in 2009, in which the platform agreed to share advertising revenue with contributors, YouTube introduced its “Content ID” system, where rightsholders are notified of infringing content and can choose to block the allegedly infringing video or monetize the content by sharing advertising revenue. YouTube reports that over 90% of rightsholders choose monetizing, which has accumulated to \$7.5bn in advertising revenue paid to rightsholders from claims through the system since 2016 ([YouTube, 2021](#)). Over 9,000 partners use the Content ID system, including large network broadcasters, movie studios, and record companies. In the second half of 2021, the Content ID system made close to 760mn claims, of which only 0.5% were disputed, suggesting that it has become feasible to license millions of derivative works with negligible

cost. The Content ID system is just one example of digital technology that has spurred the introduction of market mechanisms and allowed to circumvent the institutions that would traditionally be involved in the settling of copyright infringement cases (see [Towse, 2017](#) for an overview of the history of music publishing and licensing), thereby potentially enhancing the efficiency of licensing markets. However, to date, no empirical studies estimate the welfare effects of algorithmic licensing.

Software is another important area of licensing the re-use of copyrighted works. While many legal principles and guidelines of copyright law apply to software, several unique matters arise concerning re-use. In the traditional software distribution model, only the machine-readable code is licensed to users, which protects against third-party re-use of source code. In contrast, the Open-Source Software (OSS) model builds on the idea that source code should be freely available to develop code collaboratively, often including thousands of contributors. The economic incentives behind OSS are not the same as the traditional incentives under copyright law that arise from exclusive rights. In some cases, the owners of copyright on OSS do not even execute their copyright and put their code into the public domain. In fact, as of 2020, about 46% of the top 1 million most popular projects hosted on the code hosting platform GitHub do not specify a license to govern potential commercial applications despite the fact that the OSS community has developed a variety of license models<sup>2</sup>. Permissive licenses grant users the right to copy, modify and use the source code as desired and pass it on together with the source code, whereas restrictive licenses are more similar to traditional copyright in that they limit the extent to which code can be reused in commercial products. Hence, depending on the original creators' license, re-using open-source software in proprietary works may still constitute an infringement of copyright.

Looking at thousands of open source projects, [Lerner and Tirole \(2005\)](#) find that projects geared toward end-users tend to have restrictive licenses, while those oriented toward developers are less likely to do so. Projects with restrictive OSS licenses are negatively associated with developer interest but positively related to the interest of non-developer users and project administrators ([Subramaniam et al., 2009](#)). Evidence also shows that the degree of IP enforcement can substantially affect developer activity in OSS projects ([Wen et al., 2013, 2016](#)). Generally, the literature suggests that complementary assets matter when for-profit organizations opt for OSS as the development mode. Firms with large stocks of software patents or hardware trademarks are more likely to release OSS products,

---

<sup>2</sup>See <https://openweaver.medium.com/beware-over-half-of-the-github-public-repositories-are-not-open-source-licensed-23c7d2b5b621>, accessed October 25, 2023.

whereas firms with large stocks of software trademarks are less likely to release OSS products (Fosfuri et al., 2008). Projects like Linux, Apache, and BIND OSS provide essential technology that powers a broad range of core applications in the digital economy and therefore create tremendous welfare effects (Greenstein and Nagle, 2014; Murciano-Goroff et al., 2021). However, given the economic importance of OSS, it is surprising how little empirical research has looked at the intersection of copyright and OSS and the intersection of copyright and software in general.

## 4 Avenues for further research

Important technological and regulatory changes have a broad intersection with copyright law, leading to a range of interesting trade-offs that demand more academic work to unravel. It is critical to derive normative policy advice from a solid base of empirical evidence that appreciates the effects of copyright law in the digital age. First and foremost, we, therefore, advocate for carefully evaluating recent and future developments in regulatory frameworks and rules, such as the EU’s efforts to regulate digital markets concerning IP protection and content moderation, especially from the perspective of copyright. In what follows, we highlight a few areas that are not addressed in the empirical literature, but seem important to study. We discuss how the scope of empirical copyright research should be expanded in the future, the need to understand the effects of algorithmic licensing and blockchain technology, as well as providing evidence on copyright issues related to AI. We argue that digitization provides exciting new data sources for researchers.

### 4.1 Expanding the scope of empirical copyright research

Copyright applies to a diverse set of creative works created by various actors. Academic research has made significant progress in advancing the understanding of the fundamental trade-offs of copyright law, for example, by showing that we cannot expect considerable incentive effects from further extending copyright terms (see Heald, 2021 for a comprehensive review of the evidence base). However, the setting for most empirical work on copyright focuses on the industries of books, music, and film. Only a few empirical studies look at software, databases, performance art, and photography. Since these areas are treated differently in copyright law, e.g., with respect to a shorter protection period, the external validity of the existing evidence base might be diminished.

Another understudied area, yet critical for society, is defined by issues related to the socio-demographic diversity of creators and copyright holders (see Craig, 2021 for a review of the limited evidence base). Studies surface gender gaps of varying sizes across subject areas of copyright (Brauneis and Oliar,

2016). While in 1970, women published a third as many books as men, by 2020, women produced the majority of books, creating welfare gains for a wide range of consumers (Waldfoegel, 2023). Evidence also shows that age and gender are important predictors of engagement in creative re-use (Sinnreich and Latonero, 2014). Future research is needed to investigate closely how and why these differences exist, grow or shrink, and how and why this warrant changes in copyright law.

Expanding the scope of empirical copyright research also includes increasing the study of legal doctrines. Most importantly, research should focus on fair use and derivative works, especially in its interaction with technology (see below). Another area is the exception in copyright law for private copying. In the US, according to 17 USC §1008 and the Audio Home Recording Act of 1992, non-commercial copying of digital and analog music recordings by consumers does not constitute copyright infringement. Similarly, the 2001 EU Copyright Directive introduced an exception for private copying, which was implemented by most European jurisdictions (Directive 2001/29/EC). However, the private copying exception may lead to economic harm for copyright owners. In practice, most countries impose a system of levies on consumers via intermediaries, such as manufacturers, importers, or distributors, through a tax or charge on media and devices that enable reproduction. Although levy schemes collect more than a billion EUR in fees each year (BIEM/CISAC/Stichting de ThuisKopie, 2021), it is surprising that there is very limited academic work on the effects of levy schemes. Few existing studies analyze the economic effects of copyright levies theoretically and descriptively (Kretschmer, 2011; Poort and Quintais, 2013; Legros and Ginsburgh, 2013), but there is no empirical evidence on welfare effects. Future research could take advantage of the differences in levy schemes across countries to understand their effects on rightsholders, intermediaries, and consumers. Especially the issue of dynamic effects, for example, whether additional income from the levy system leads to more creative work and higher quality, seems important to approach empirically.

*Summary and exemplary research questions:* Academic research on copyright law has largely neglected specific industries, the socio-demographic diversity of creators, and important legal details like fair use and private copying, which calls for further empirical investigation of the potential impact of copyright law. Examples of open questions include:

- Do the established results in the literature (e.g., on incentive effects of copyright) also hold regarding copyright-protected works that have not been studied much, e.g., software, databases, performance art, and photography?

- How can we characterize the socio-demographic diversity of creators and copyright holders? Does copyright law represent the interests of all socio-demographic groups accordingly, and if not, what should be changed?
- What are the static and dynamic welfare effects of copyright exceptions, e.g., for private copying or fair use? How have welfare effects changed with digital technology, and does copyright law need to adapt because of these changes?

## 4.2 Algorithmic licensing and blockchain technology

Even in a de-facto weakened copyright system, we have seen an increase in creative output as technological advancements and online platforms have simultaneously reduced entry barriers. As discussed, it seems plausible that automated enforcement, such as the N&T systems, and algorithmic licensing, such as YouTube’s Content-ID, have played a major role in the explosion of user-generated content. Further empirical work is needed to establish this as a fact and to quantify the resulting welfare effects, for example, in a similar methodological fashion as prior work ([Waldfoegel and Reimers, 2015](#); [Waldfoegel, 2017](#); [Aguilar and Waldfoegel, 2018a](#); [Reimers and Waldfoegel, 2021](#)).

In the future, algorithmic licensing using blockchain technology – across different platforms and end-users – might become an important topic to study ([Bodó et al., 2018](#)). Blockchain technology – a decentralized ledger that allows to store information in a transparent fashion securely – can provide the technological infrastructure to not only unambiguously match licensor and licensee but also to automatically enforce license payments using so-called smart contracts ([Catalini and Gans, 2017](#)). Blockchain technology can be seen as the next generation of DRM solutions. The decentralized storage of ownership information and a substantially reduced power of intermediaries are the key differentiating factors from previous DRM solutions, with Non-fungible Tokens (NFTs) as one example. Especially for digital artwork, music, film, and video games, NFTs offer a way to create, sell, and collect digital goods by relating them to tokens which can then be sold on digital marketplaces, making NFTs a potentially more efficient tool to license to the use of original work. Further, NFT platforms offer creators of NFTs the opportunity to claim their droit de suite rights and resale royalties on subsequent sales of their works, which certain countries, including the US, do not traditionally offer ([Evans, 2019](#)). More research is necessary to better understand the role of the relationship between creators and intermediaries since there is variation across types of content and types of platforms. With market-enabling solutions, such as technology that enables automated enforcement and algorithmic licensing,

the question is whether the traditional institutions in the copyright system (e.g., national collection societies that represent the interests of some national artists) will or should become obsolete. Therefore, studying the relative benefits and costs of the copyright system and technology-based mechanisms is crucial, particularly regarding the welfare effects of derivative works under traditional and algorithmic licensing.

*Summary and exemplary research questions:* Technological advancements and online platforms have reduced entry barriers, leading to an increase in creative output despite a weakened copyright system. Further research is needed to understand the role of copyright in the context of algorithmic licensing, and blockchain technology, as well as to study the relationship between creators, intermediaries, and the resulting welfare effects. Some of the open questions we identify include:

- What are the welfare effects of algorithmic licensing? What is the role of platforms and specific technologies (e.g., ContentID vs NFT)? Can we quantify potential efficiency gains?
- How do automated enforcement mechanisms affect cumulative innovation in copyright-related industries? Which kinds of works do not exist (in socially large enough quantities) because of copyright restrictions?
- What are the limits and frictions of algorithmic licensing technology, especially given the incentives of platforms and their business models? Can we characterize the rate of type I and type II errors (flagging lawful content as unlawful vs. not flagging unlawful content as unlawful) of N&T systems? What are the characteristics of traditional copyright institutions that cannot be replaced by technology?
- What are the distributional effects of automated licensing, both at the micro- and macro-level? How does algorithmic licensing affect the (international) convergence or divergence of culture, either by enabling the (automated) co-production of works or by allowing international digital trade in licenses?

### **4.3 Issues related to data and artificial intelligence**

#### **4.3.1 Copyright in the data economy**

Data plays an increasingly important role in the digital economy. Private firms and public organizations do not only collect information about consumers through connected personal devices, but more

and more sensors monitor industrial machines and public infrastructure. Data is the key input factor in Machine Learning (ML) and AI technologies that enable innovation, either by discovering knowledge (Bianchini et al., 2022), increasing productivity (Wu et al., 2020; Yang, 2022), or by expanding existing or serving new markets (Bessen et al., 2022; Rammer et al., 2022).

The ever-growing abundance of information available on the Internet, including unstructured data from text, images, sounds, and video, can nowadays be automatically extracted at scale thanks to technological innovation. Therefore, making copies of existing works, which may be copyright-protected, is indispensable and unavoidable to create training data for ML algorithms. However, the right balance between the interests of copyright holders and access to information seems difficult to establish without empirical evidence (see Margoni, 2021 for an overview).

In the EU, according to the DCDSM, the reproduction and extraction of works in order to carry out text and data mining are possible without prior authorization from the rightsholder for purposes of scientific research, provided that access to the works in question is lawful and that the copies may be stored in a secure environment for no longer than is necessary for scientific research (Art. 3(2)). “Lawful access” covers access to content under contractual arrangements (e.g., subscriptions or open access licenses), as well as to “content that is freely available online” (see Recital 14). For non-research uses, rightsholders remain able to forbid or license the use and reproduction of their works. In 2016, the US Court of Appeals for the Second Circuit stated in *Author’s Guild v. Google* (804 F.3d 202 (2d Cir. 2015) 116 U.S.P.Q.2d (BNA) 1423) that training an algorithm on copyrighted data without getting the copyright owner’s permission first is legal under the fair use doctrine, pointing out the significant public benefits and advances in “the progress of the arts and sciences while respecting the rights of authors and other creative individuals, and without adversely impacting the rights of copyright holders.” In the case of *hiQ Labs, Inc. v. LinkedIn Corp.*, the US Court of Appeals for the Ninth Circuit ruled that scraping data from public sections of the social network website LinkedIn is legal and affirmed its decision in 2022. However, later in 2022, after a review of the decision, it ruled that hiQ had breached LinkedIn’s User Agreement (*hiQ Labs v. LinkedIn Corp.*, 938 F.3d 985 (9th Cir. 2019) and 31 F.4th 1180 (9th Cir. 2022)). Currently pending lawsuits in the US against numerous AI-system owners (e.g., Stable Diffusion and Midjourney) allege, among other claims, direct copyright infringement by creating unauthorized copies of their copyrighted works and using such copies as training data.

While underlying algorithms and applications are used globally, important differences exist across jurisdictions with respect to copyright exceptions for research and data mining (Fiil-Flynn et al., 2022). These differences, however, may create quasi-experimental settings that let researchers study how more or less restrictive copyright systems causally affect innovation in AI and data-driven and AI-enabled innovation.

In the context of data, copyright law has intricate interactions with other areas of the law. Intellectual property rights, particularly copyright, also intersect with privacy and competition policy. Competition policy instruments that mandate firms to share data with competitors can limit the scope of IP protection available to rights holders (Graef and Prüfer, 2021). Examples from the EU can be found embedded in the EU’s General Data Protection Regulation (GDPR) of 2018, the Digital Markets Act (DMA), which will come into force in 2024 and the proposed Data Act. When copyright allows actors to withhold data for the training of AI applications, innovation may be slowed down because inadequate training datasets can lead to biased outcomes of ML models (Vidgen and Derczynski, 2020; Shimron et al., 2022). Related evidence shows that limited access to personal data, for example because of data protection laws, can lead ML applications to produce lower quality predictions, leading to fewer clicks on personalized recommendations (Peukert et al., 2023) and more concentrated consumer choice (Sun et al., 2023).

*Summary and exemplary research questions:* Data is vital for innovation in AI and ML technologies, but striking a balance between copyright protection and access to information is challenging, especially when copyright law intersects with competition and privacy policy. Empirical evidence is needed to guide pressing policy issues. Exemplary research questions include:

- What are the dynamic effects of the research exemption in copyright law, compared to jurisdictions where such exemption does not exist? Important empirical settings to study this question include academic research across all disciplines, as well as applied research and development of data-driven innovation, e.g., foundation models of AI.
- Under which circumstances can a copyright system without research exemptions provide higher quality training data for AI applications because rightsholders do not have incentives to strategically withdraw data or works that can be used as training data?
- Do we need to extend copyright law to include ML model parameters (the result of costly



training)? Or will DRM-like systems (see e.g., [Vyas et al., 2023](#); [Zhong et al., 2023](#)) provide efficient private solutions to generate incentives for firms to invest in training large complex models?

- How does the (non-) existence of research exemptions in national copyright law affect the inclusion of works created by actors with diverse socio-demographic backgrounds in training datasets? What are the interactions between data protection law and copyright law with respect to the bias they may introduce to ML models via training datasets?

### 4.3.2 Copyright and AI-generated works

In the past few years, the perhaps most striking innovation developments regarding copyright and the creation of new content have taken place with generative AI. Recent advances in ML have opened up new possibilities to produce literary, musical, and artistic works, and AI has become a new tool to express ideas for hundreds of millions of users. The applications are many and vary across the creative domains. For example, WaveNet was initially developed to help a machine learning algorithm create seamless artificial voices and has learned to make music itself by analyzing voice material. Current AI systems can also quickly create customized royalty-free background music, jungles, and even entire albums. Also, some video game environments are no longer created by programmers but by the program itself based on predetermined rules and algorithms (e.g., in *No Man's Sky*). In visual art, machines can emulate human thought and decide how to alter an image based on a given algorithm. With AI art generators, such as DALL-E, Stable Diffusion, and Midjourney, users can create artwork by entering a prompt, choosing different styles, and further modifying the generated outputs to express their ideas.

As the technology is reaching a point where it is difficult to distinguish works created by humans from works generated with AI ([Jakesch et al., 2023](#)), the importance of human influence on creative works is put to the test. At present, AI-generated works are not protected by copyright in most jurisdictions because of the missing element of “human authorship” which is seen as the guiding principle ([Hristov, 2016](#); [Ramalho, 2017](#)). Consequently, these works directly enter the public domain. The ongoing development of AI and its ability to make critical, creative decisions either with or without human intervention gives us reason to rethink our understanding of originality and ownership and evolve copyright law.

When no protection is granted, the output generated by AI might be considered fair use of the

underlying processing and recombination of training data. Since the AI is doing much more than copying bits and pieces of existing works and is creating new works based on the techniques learned from studying the copyrighted prior work, much like an art student learns to paint in particular styles by studying existing artwork in that style. However, output substantially similar to original works might still be copyright infringing.

The case of a song that uses AI-generated voices of the popular artists “Drake & The Weeknd” released in 2023 further highlights a dilemma of online copyright enforcement for platforms. Under pressure from the record label that released the artists’ original works, streaming platforms such as Spotify and YouTube removed the song, citing copyright infringement. Given the legal uncertainty, accepting such infringement claims is tricky for Google, which is YouTube’s parent organization and a leader in the development of AI. As discussed in [Patel \(2023\)](#), by agreeing with the record label and treating AI-generated music as copyright infringing, Google endangers its own development of generative AI products (which may have much broader applications) under the fair use argument. On the other hand, by not agreeing with the record label, they risk lawsuits and potentially lose necessary licenses for music on YouTube, one of the biggest content categories.

Even in countries where AI-generated work is protected by copyright, e.g., in the UK, it is not clear who the author of the output is – several eligible personas come into question, such as the architect, the programmer or the user of the model, who have each contributed necessary input for the final work to be created.

Depending on whether copyright protection for AI-generated works is granted has further implications on the question of potential copyright infringement regarding the training data, making it more likely that copying original images for training data is legal because the resulting AI software is used to create new, non-infringing artwork.

*Summary and exemplary research questions:* The innovation in generative AI has led to significant developments in copyright-related works, raising questions about human influence, originality, and ownership. The unclear status of AI-generated works within copyright law, and different approaches for classification depending on the jurisdiction raises legal uncertainties for platforms and organizations involved in AI development. Evidence-based policy advice is needed to fine-tune copyright law for the age of AI, for example, by answering the following questions:

- How can we characterize the demand for AI-generated works? Under which circumstances are

consumers willing to substitute machine-created content for human-created content? Answers to such questions can help to calibrate welfare models and provide ex-ante guidance for policy, for example, on the trade-off between copyright exemptions and incentives to create.

- What are the effects of generative AI on the industrial organization of creative industries, and how do these effects depend on the allocation of ownership rights?
- What is the role of business models and licensing markets to maximize welfare with respect to AI-generated works? More broadly, can we design market mechanisms that align the economic incentives of machines with socially desirable goals?

#### 4.4 New data sources

Digitization has not only changed institutions and market behaviors around copyright-related issues but opened up a world of opportunities for empirical researchers to monitor and measure constructs and behavior. With the vast amount of data available on the internet, researchers are increasingly using web scraping as the method of choice to access untapped sources of information for empirical research. We now also have access to resources that help anyone, even without much technical knowledge, to write simple scripts that automatically collect unstructured information from websites and parse it into quantifiable data (Boegershausen et al., 2022; Miric et al., 2023).

In addition, a wide range of structured databases are available and well-suited to answer copyright-related research questions. While already available since the early 2000's, databases covering the notice & takedown system seem under-utilized by the community of empirical copyright researchers. A case in point is the Lumen database, which provides researchers with access to takedown notices and other legal complaints related to online content, covering a variety of platforms, including Google Search.<sup>3</sup> Community-led efforts to catalog publications of copyright-related products have led to immense databases on releases and meta information around music (e.g., MusicBrainz), films (e.g. IMDB), video games (e.g., Mobygames), and books (e.g., Goodreads). And more generally, application programming interfaces (APIs) of content platforms are valuable tools for researchers to access data on copyright-related issues. For instance, the YouTube Research API allows researchers to collect data on user behavior, such as views, likes, and comments, as well as copyright-related availability information of content (see, e.g., Kretschmer and Peukert, 2020). Other platforms, such as Twitch and TikTok,

---

<sup>3</sup>See <https://www.lumendatabase.org/>, accessed October 25, 2023.

provide similar data access methods, which represent ample opportunities for empirical research on understudied copyright-related phenomena in video (game) content and the copyright economics of live streams.

For research on copyright-related issues in software, the open-source development platform Github provides detailed information about individual users and their coding activities, but also about the software code itself, alongside information on dependencies on prior work (i.e., pre-existing libraries) and licensing choices (see e.g., [Gambardella and Hall, 2006](#); [Vendome, 2015](#)). Through GitHub's API, millions of data points are freely available, but several projects (e.g., GHArchive) provide historical data alongside GitHub's own initiatives that support academic research with data.<sup>4</sup>

Technological advancements have also led government bodies to digitize historical records and publish micro-level data. For instance, in 2022, the US copyright office released a data set containing copyright registration records, copyright renewal records, and recorded document records from 1978 to 2021.<sup>5</sup> This includes author information, types of works registered, and publication status. A large range of questions can be studied with this dataset, hopefully, also applications of copyright that have not received much attention in the literature, such as performing arts, visual arts, or mask work (a process in the creation of integrated circuits and semiconductor chips). In addition, the data set enables researchers to study copyright registrations of unpublished works. Data collected and made available by [Yuvaraj et al. \(2022\)](#) additionally covers copyright termination notices, which can be helpful, e.g., to study the empirical importance of orphan works. Moreover, disclosure requirements in laws can provide researchers with access to data. The Freedom of Information Act (FOIA) in the United States, for instance, provides individuals with the right to access records held by federal agencies ([Walterscheid, 1989](#)). The EU DSA includes provisions related to data disclosure requirements, which are designed to improve transparency and accountability in the digital ecosystem. Under the DSA, digital platforms will be required to provide certain information to users, authorities, and researchers upon request. For instance, on the criteria and processes used to moderate content, including details on how decisions are made to remove or restrict access to content, as well as data on the use and distribution of copyrighted content.

---

<sup>4</sup>See, e.g., <https://github.blog/2021-08-31-request-for-proposals-defining-standardized-github-metrics/>, accessed October 25, 2023.

<sup>5</sup>See <https://copyright.gov/policy/women-in-copyright-system/>, accessed October 25, 2023.

## 5 Conclusions

In this paper, we have provided a review of some fundamental aspects of copyright law, with a particular focus on changes brought about by digital technology. Our non-exhaustive synthesis of the empirical literature highlights significant progress in understanding the fundamentals of copyright and the challenges and opportunities of copyright in the digital age. It is now an established fact that online copyright enforcement did not contribute to a decrease in demand-side copyright infringement that was enabled by file-sharing and unlicensed streaming platforms but that online piracy was indeed at least partially curbed by new business models and licensing.

However, surprisingly little empirical work studies the precise settings in which licensing is used at scale in online platforms, e.g., in the context of open-source software or audiovisual entertainment content. A big open question remains whether automated licensing, which can be thought of as the technologically-enhanced and market-enabling refinement of the legal construct of N&T, is equally effective in generating follow-on innovation as it seems to be in enforcing copyright vis-a-vis plain unlawful copying. In addition, the welfare effects of platforms such as YouTube, expanding their power by making their own rules and primarily taking over and privatizing copyright enforcement, are not clear.

In general, we emphasize the pressing need for further research at the intersection of law and economics to offer normative guidance to policymakers. We particularly highlight copyright's pivotal role in the forthcoming wave of digital technological advancements. This new wave is characterized by intricate issues surrounding data and IP rights concerning intermediary and final outputs of ML models. Notably, as regulatory approaches in consumer protection, piracy, and competition policy continue to intertwine with copyright law, understanding and navigating these complex interactions becomes increasingly crucial for policymakers and stakeholders alike.

## References

- Adermon, A., and Liang, C.-Y. (2014). “Piracy and Music Sales: The Effects of an Anti-Piracy Law.” *Journal of Economic Behavior & Organization*, 105, 90–106.
- 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., Reimers, I., and Waldfogel, J. (2023). “Platforms and the transformation of the content industries.” *Journal of Economics & Management Strategy*.
- Aguiar, L., and Waldfogel, J. (2016). “Even the losers get lucky sometimes: New products and the evolution of music quality since napster.” *Information Economics and Policy*, 34, 1–15.
- 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*, forthcoming.
- Ahn, I., and Shin, I. (2010). “On the optimal level of protection in drm.” *Information Economics and Policy*, 22(4), 341–353.
- Akmon, D. (2010). “Only with your permission: how rights holders respond (or don’t respond) to requests to display archival materials online.” *Archival Science*, 10(1), 45–64.
- Albinsson, S. (2021). “Avoiding silent opera: the ‘grand’ performing right at work in nineteenth century paris.” *European Journal of Law and Economics*, 51(1), 183–200.
- Arai, Y., and Kinukawa, S. (2014). “Copyright infringement as user innovation.” *Journal of Cultural Economics*, 38(2), 131–144.
- Asay, C. D. (2022). “An empirical study of copyright’s substantial similarity test.” Available at SSRN 4013095.
- Aversa, P., Hervas-Drane, A., and Evenou, M. (2019). “Business model responses to digital piracy.” *California Management Review*, 61(2), 30–58.
- Awaworyi Churchill, S., Luong, H. M., and Ugur, M. (2022). “Does intellectual property protection deliver economic benefits? a multi-outcome meta-regression analysis of the evidence.” *Journal of Economic Surveys*, 36(5), 1477–1509.
- Balganesh, S., Manta, I. D., and Wilkinson-Ryan, T. (2014). “Judging similarity.” *Iowa L. Rev.*, 100, 267.
- Balganesh, S., and Menell, P. S. (2022). “Proving copying.” *William & Mary Law Review*, 64.
- Banerjee, D. S. (2003). “Software piracy: a strategic analysis and policy instruments.” *International Journal of Industrial Organization*, 21(1), 97–127.
- Bar-Ziv, S., and Elkin-Koren, N. (2018). “Behind the scenes of online copyright enforcement: Empirical evidence on notice & takedown.” *Conn. L. Rev.*, 50, 339.

- 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.
- Bechtold, S., Buccafusco, C., and Sprigman, C. J. (2015). “Innovation heuristics: Experiments on sequential creativity in intellectual property.” *Ind. LJ*, 91, 1251.
- Beebe, B. (2007). “Does judicial ideology affect copyright fair use outcomes: Evidence from the fair use case law.” *Colum. JL & Arts*, 31, 517.
- Belleflamme, P., and Peitz, M. (2010). “Digital piracy: theory.”
- Belleflamme, P., Peitz, M., et al. (2014). “Digital piracy: an update.” *Encyclopedia of Law and Economics, Springer (in print)*, UCL.
- Bellégo, C., and De Nijs, R. (2020). “The unintended consequences of antipiracy laws on markets with asymmetric piracy: The case of the french movie industry.” *Information Systems Research*, 31(4), 1064–1086.
- Bently, L., Deazley, R., and Kretschmer, M. (2010). *Privilege and property: essays on the history of copyright*. Open Book Publishers.
- Bessen, J., Impink, S. M., Reichensperger, L., and Seamans, R. (2022). “The role of data for ai startup growth.” *Research Policy*, 51(5), 104513.
- Bianchini, S., Müller, M., and Pelletier, P. (2022). “Artificial intelligence in science: An emerging general method of invention.” *Research Policy*, 51(10), 104604.
- 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.
- BIEM/CISAC/Stichting de ThuisKopie (2021). “Private copying global study.”
- Bodó, B., Gervais, D., and Quintais, J. P. (2018). “Blockchain and smart contracts: the missing link in copyright licensing?” *International Journal of Law and Information Technology*, 26(4), 311–336.
- Boegershausen, J., Datta, H., Borah, A., and Stephen, A. T. (2022). “Fields of gold: Scraping web data for marketing insights.” *Journal of Marketing*, 86(5), 1–20.
- Bradley, W. A., and Kolev, J. (2023). “How does digital piracy affect innovation? evidence from software firms.” *Research Policy*, 52(3), 104701.
- Brauneis, R., and Oliar, D. (2016). “Copyright’s race, gender and age: A first quantitative look at registrations.” *GWU Law School Public Law Research Paper*, (2016-48).
- Brøvig-Hanssen, R., and Jones, E. (2021). “Remix’s retreat? content moderation, copyright law and mashup music.” *New Media & Society*, 14614448211026059.
- 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.
- Cagé, J., Hervé, N., and Viaud, M.-L. (2020). “The production of information in an online world.” *The Review of Economic Studies*, 87(5), 2126–2164.
- Catalini, C., and Gans, J. S. (2017). “Some simple economics of the blockchain.” *SSRN Working Paper 2874598*.

- Chen, Y., and Png, I. (2003). “Information Goods Pricing and Copyright Enforcement: Welfare Analysis.” *Information Systems Research*, 14(1), 107–123.
- Circum (2016). “Examination of the” follow-the-money” approach to copyright piracy reduction.” <https://www.canada.ca/en/canadian-heritage/services/copyright-policy-publications/follow-money-piracy.html>.
- Cotropia, C. A., and Gibson, J. (2013). “Copyright’s topography: An empirical study of copyright litigation.” *TEx. L. REv.*, 92, 1981.
- Craig, C. (2021). “Copyright gender – evidencing the connections.” *CREATE 21 for 2021 Blog*, <https://www.create.ac.uk/blog/2021/12/17/21-for-2021-copyright-gender-evidencing-the-connections/>.
- Cuntz, A. (2023). “Grand rights and opera reuse today.” *Oxford Economic Papers*, 75(1), 206–232.
- Cuntz, A., and Bergquist, K. (2022). “Exclusive content and platform competition in latin america.” *Information Economics and Policy*, 60, 100989.
- Cuntz, A., and Sahli, M. (2023). “Intermediary liability and trade in follow-on innovation.” *Journal of Cultural Economics*, 1–42.
- 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., 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. (2014). “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.
- Eisend, M. (2019). “Explaining digital piracy: a meta-analysis.” *Information Systems Research*, 30(2), 636–664.
- Electronic Frontier Foundation (2008). “Riaa v. the people: Five years later.” <https://www.eff.org/wp/riaa-v-people-five-years-later>.
- Ellis-Petersen, H. (2014). “Taylor Swift takes a stand over Spotify music royalties.” <https://www.theguardian.com/music/2014/nov/04/taylor-swift-spotify-streaming-album-sales-snub>, last accessed: April 12, 2018.
- Erickson, K., and Kretschmer, M. (2018). “This video is unavailable.” *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, 9, 75.
- Erickson, K., Kretschmer, M., and Mendis, D. (2019). “An empirical approach to the public domain.” In *The Innovation Society and Intellectual Property*, Edward Elgar Publishing.
- Essling, C., Koenen, J., and Peukert, C. (2017). “Competition for attention in the digital age: The case of single releases in the recorded music industry.” *Information Economics and Policy*, 40, 26–40.



- EU Intellectual Property Office (2021). “Online copyright infringement in the european union.” *EUIPO Report*, doi:10.2814/505158.
- Evans, T. M. (2019). “Cryptokitties, cryptography, and copyright.” *AIPLA QJ*, 47, 219.
- Favale, M., Homberg, F., Kretschmer, M., Mendis, D., and Secchi, D. (2015). “Copyright, and the regulation of orphan works: A comparative review of seven jurisdictions and a rights clearance simulation.” Available at SSRN 2613498.
- 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., et al. (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 copyright’s under use hypothesis across australia, new zealand, the united states and canada.” *University of New South Wales Law Journal*, The, 42(4), 1215–1253.
- Fosfuri, A., Giarratana, M. S., and Luzzi, A. (2008). “The penguin has entered the building: The commercialization of open source software products.” *Organization science*, 19(2), 292–305.
- Gambardella, A., and Hall, B. H. (2006). “Proprietary versus public domain licensing of software and research products.” *Research Policy*, 35(6), 875–892.
- Gans, J. S. (2015). “Remix rights and negotiations over the use of copy-protected works.” *International Journal of Industrial Organization*, 41, 76–83.
- Gillespie, T. (2020). “Content moderation, ai, and the question of scale.” *Big Data & Society*, 7(2), 2053951720943234.
- 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.
- Godinho de Matos, M., Ferreira, P., and Smith, M. D. (2018). “The effect of video-on-demand on piracy: Evidence from a household level randomized experiment.” *Management Science*, forthcoming.
- Goldstein, P. (2007). “Fair use in context.” *Colum. JL & Arts*, 31, 433.
- Graef, I., and Prüfer, J. (2021). “Governance of data sharing: A law & economics proposal.” *Research Policy*, 50(9), 104330.
- Gray, J. E., and Suzor, N. P. (2020). “Playing with machines: Using machine learning to understand automated copyright enforcement at scale.” *Big Data & Society*, 7(1), 2053951720919963.
- Greenstein, S., Lerner, J., and Stern, S. (2013). “Digitization, innovation, and copyright: What is the agenda?” *Strategic Organization*, 11(1), 110–121.
- Greenstein, S., and Nagle, F. (2014). “Digital dark matter and the economic contribution of apache.” *Research Policy*, 43(4), 623–631.
- Hall, B. H., and Harhoff, D. (2012). “Recent research on the economics of patents.” *Annu. Rev. Econ.*, 4(1), 541–565.

- Handke, C. (2012). “A taxonomy of empirical research on copyright-how do we inform policy?” *Review of Economic Research on Copyright Issues*, 9(1), 47–92.
- Handke, C., Balazs, B., and Vallbé, J.-J. (2016). “Going means trouble and staying makes it double: the value of licensing recorded music online.” *Journal of Cultural Economics*, 40(3), 227–259.
- Haruvy, E., Mahajan, V., and Prasad, A. (2004). “The effect of piracy on the market penetration of subscription software.” *The Journal of Business*, 77(S2), S81–S107.
- Heald, P. (2021). “Term of copyright: Optimality and reality.” *CREATE 21 for 2021 Blog*, <https://www.create.ac.uk/blog/2021/06/18/21-for-2021-term-of-copyright-optimality-and-reality/>.
- Heald, P., Erickson, K., and Kretschmer, M. (2015). “The valuation of unprotected works: a case study of public domain images on wikipedia.” *Harv. JL & Tech.*, 29, 1.
- Heald, P. J. (2014). “How copyright keeps works disappeared.” *Journal of Empirical Legal Studies*, 11(4), 829–866.
- Heald, P. J. (2018). “Copyright reversion to authors (and the rosetta effect): An empirical study of reappearing books.” *J. Copyright Soc’y USA*, 66, 59.
- Hiller, R. S. (2016). “Sales Displacement and Streaming Music: Evidence from YouTube.” *Information Economics and Policy*, 34, 16–26.
- Hiller, R. S., and Walter, J. (2017). “The rise of streaming music and implications for music production.” *SSRN Working Paper 2670976*.
- 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.
- Hristov, K. (2016). “Artificial intelligence and the copyright dilemma.” *Idea*, 57, 431.
- Jakesch, M., Hancock, J. T., and Naaman, M. (2023). “Human heuristics for ai-generated language are flawed.” *Proceedings of the National Academy of Sciences*, 120(11), e2208839120.
- Jiménez Durán, R. (2022). “The economics of content moderation: Theory and experimental evidence from hate speech on twitter.” *Available at SSRN*.
- Kaiser, F., Cuntz, A., and Peukert, C. (2023). “Batman Forever? The Role of Trademarks for Reuse in the US Comics Industry.” *Research Policy*.
- Karaganis, J., and Urban, J. (2015). “The rise of the robo notice.” *Communications of the ACM*, 58(9), 28–30.
- Kiema, I. (2008). “Commercial piracy and intellectual property policy.” *Journal of Economic Behavior & Organization*, 68(1), 304–318.
- Kim, J.-H., and Leung, T. C. (2021). “Eliminating digital rights management from the e-book market.” *Information Economics and Policy*, 57, 100935.
- Kretschmer, M. (2011). “Private copying and fair compensation: An empirical study of copyright levies in europe.” *Independent Report Commissioned by the UK Intellectual Property Office*.
- 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., and Lichtman, D. (2003). “Indirect liability for copyright infringement: Napster and beyond.” *Journal of economic perspectives*, 17(2), 113–124.
- Landes, W. M., and Posner, R. A. (1989). “An economic analysis of copyright law.” *The Journal of Legal Studies*, 18(2), 325–363.
- Lauinger, T., Szydowski, 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, E., and Moshirnia, A. (2021). “Do experts matter? a study of the effect of musicologist testimony in music cases.” *Forthcoming, University of Illinois Law Review*, 2022.
- Legros, P., and Ginsburgh, V. A. (2013). “The economics of copyright levies on hardware.” *Review of Economic Research on Copyright Issues*, 10(1), 20–35.
- Lerner, J., and Tirole, J. (2005). “The scope of open source licensing.” *Journal of Law, Economics, and Organization*, 21(1), 20–56.
- Li, X., Liao, C., and Xie, Y. (2021a). “Digital piracy, creative productivity, and customer care effort: Evidence from the digital publishing industry.” *Marketing Science*, 40(4), 685–707.
- 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.
- Li, X., Shi, M., and Zhao, C. Y. (2021b). “Incentivizing mass creativity: An empirical study of the online publishing market.” *Available at SSRN 3842153*.
- Liu, Y., Yildirim, P., and Zhang, Z. J. (2022). “Implications of revenue models and technology for content moderation strategies.” *Marketing Science*.
- 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.
- Lu, Y., and Poddar, S. (2012). “Accommodation or deterrence in the face of commercial piracy: the impact of intellectual property rights protection.” *Oxford Economic Papers*, 64(3), 518–538.
- Lund, J. (2011). “An empirical examination of the lay listener test in music composition copyright infringement.” *Va. Sports & Ent. LJ*, 11, 137.
- Lund, J. (2013). “Fixing music copyright.” *Brook. L. Rev.*, 79, 61.
- Luo, H., and Mortimer, J. H. (2017). “Copyright enforcement: Evidence from two field experiments.” *Journal of Economics & Management Strategy*, 26(2), 499–528.
- Luo, H., and Mortimer, J. H. (2021). “Infringing use as a path to legal consumption: Evidence from a field experiment.” *Journal of Economics & Management Strategy*, 1–20.
- Madio, L., and Quinn, M. (2021). “Content moderation and advertising in social media platforms.” *Available at SSRN 3551103*.
- Margoni, T. (2021). “Computational uses.” *CREATE 21 for 2021 Blog*, <https://www.create.ac.uk/blog/2021/10/08/21-for-2021-computational-uses/>.

- McKenzie, J. (2017). “Graduated response policies to digital piracy: Do they increase box office revenues of movies?” *Information Economics and Policy*, 38, 1–11.
- Miric, M., and Jeppesen, L. B. (2020). “Does piracy lead to product abandonment or stimulate new product development?: Evidence from mobile platform-based developer firms.” *Strategic Management Journal*, 41(12), 2155–2184.
- Miric, M., Jia, N., and Huang, K. G. (2023). “Using supervised machine learning for large-scale classification in management research: The case for identifying artificial intelligence patents.” *Strategic Management Journal*, 44(2), 491–519.
- Mortimer, J. H., Nosko, C., and Sorensen, A. (2012). “Supply responses to digital distribution: Recorded music and live performances.” *Information Economics and Policy*, 24(1), 3–14.
- Murciano-Goroff, R., Zhuo, R., and Greenstein, S. (2021). “Hidden software and veiled value creation: Illustrations from server software usage.” *Research Policy*, 50(9), 104333.
- 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*.
- 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.
- Orme, T. (2014). “The short- and long-term effectiveness of anti-piracy laws and enforcement actions.” *Journal of Cultural Economics*, 38, 351–368.
- Papies, D., and van Heerde, H. J. (2017). “The dynamic interplay between recorded music and live concerts: The role of piracy, unbundling, and artist characteristics.” *Journal of Marketing*, 81(4), 67–87.
- Patel, N. (2023). “Ai drake just set an impossible legal trap for google.” *The Verge*, <https://www.theverge.com/2023/4/19/23689879/ai-drake-song-google-youtube-fair-use>.
- Peukert, C. (2019). “The next wave of digital technological change and the cultural industries.” *Journal of Cultural Economics*, 43(2), 189–210.
- 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. (2023). “The editor and the algorithm: Recommendation technology in online news.” *Management Science*, *forthcoming*.
- 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.
- Poort, J., and Quintais, J. P. (2013). “The levy runs dry: A legal and economic analysis of eu private copyright levies.” *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, 4, 205.
- Quintais, J. P. (2017). *Copyright in the age of online access: alternative compensation systems in EU law*. Kluwer Law International BV.

- Ramalho, A. (2017). “Will robots rule the (artistic) world? a proposed model for the legal status of creations by artificial intelligence systems.” *A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems* (June 13, 2017).
- Rammer, C., Fernández, G. P., and Czarnitzki, D. (2022). “Artificial intelligence and industrial innovation: Evidence from german firm-level data.” *Research Policy*, 51(7), 104555.
- Reimers, I. (2016). “Can private copyright protection be effective? evidence from book publishing.” *The Journal of Law and Economics*, 59(2), 411–440.
- Reimers, I. (2019). “Copyright and generic entry in book publishing.” *American Economic Journal: Microeconomics*, 11(3), 257–84.
- Reimers, I., and Waldfogel, J. (2021). “Digitization and pre-purchase information: the causal and welfare impacts of reviews and crowd ratings.” *American Economic Review*, 111(6), 1944–71.
- 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.
- Sag, M. (2005). “God in the machine: A new structural analysis of the fair use doctrine in copyright law.”
- Savelkoul, R. (2020). “Superstars vs the long tail: How does music piracy affect digital song sales for different segments of the industry?” *Information Economics and Policy*, 50, 100847.
- Sawers, P. (2016). “YouTube strikes landmark music licensing deal in Germany after more than 7 years of blocked videos.” <https://venturebeat.com/2016/11/01/youtube-gema-germany-google-deal/>, last accessed: July 25, 2018.
- Schuster, M., Mitchell, D., and Brown, K. (2019). “Sampling increases music sales: An empirical copyright study.” *American Business Law Journal*, 56(1), 177–229.
- Seng, D. (2015). *An Empirical Study of Dmca Takedown Notices*. Stanford University.
- Shiller, B., and Waldfogel, J. (2013). “The challenge of revenue sharing with bundled pricing: An application to music.” *Economic Inquiry*, 51(2), 1155–1165.
- Shimron, E., Tamir, J. I., Wang, K., and Lustig, M. (2022). “Implicit data crimes: Machine learning bias arising from misuse of public data.” *Proceedings of the National Academy of Sciences*, 119(13), e2117203119.
- Sinnreich, A., and Latonero, M. (2014). “Tracking configurable culture from the margins to the mainstream.” *Journal of Computer-Mediated Communication*, 19(4), 798–823.
- Smith, M. D., Telang, R., and Zhang, Y. (2019). “I want you back: The interplay between legal availability and movie piracy.” *International Journal of the Economics of Business*, 26(1), 199–216.
- Subramaniam, C., Sen, R., and Nelson, M. L. (2009). “Determinants of open source software project success: A longitudinal study.” *Decision Support Systems*, 46(2), 576–585.
- Sun, T., Yuan, Z., Li, C., Zhang, K., and Xu, J. (2023). “The value of personal data in internet commerce: A high-stake field experiment on data regulation policy.” *Management Science*, forthcoming.

- Sundararajan, A. (2004). “Managing digital piracy: Pricing and protection.” *Information Systems Research*, 15(3), 287–308.
- Sweney, M. (2017). “Music industry goes to war with YouTube.” <https://www.theguardian.com/business/2017/apr/15/music-industry-youtube-video-streaming-royalties>, last accessed: July 25, 2018.
- Towse, R. (2006). “Copyright and artists: a view from cultural economics.” *Journal of economic surveys*, 20(4), 567–585.
- Towse, R. (2017). “Economics of music publishing: copyright and the market.” *Journal of Cultural Economics*, 41(4), 403–420.
- UK Intellectual Property Office (2016). “Online copyright infringement tracker – latest wave of research mar 16 – may 16.” *Policy report*.
- Urban, J. M., Karaganis, J., and Schofield, B. L. (2017). “Notice and takedown: Online service provider and rightsholder accounts of everyday practice.” *J. Copyright Soc’y USA*, 64, 371.
- Valdovinos Kaye, D. B., Rodriguez, A., Langton, K., and Wikstrom, P. (2021). “You made this? i made this: Practices of authorship and (mis) attribution on tiktok.” *International Journal of Communication*, 15, 3195–3215.
- Varian, H. R. (2005). “Copying and copyright.” *Journal of economic perspectives*, 19(2), 121–138.
- Vendome, C. (2015). “A large scale study of license usage on github.” In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, vol. 2, 772–774, IEEE.
- Vidgen, B., and Derczynski, L. (2020). “Directions in abusive language training data, a systematic review: Garbage in, garbage out.” *Plos one*, 15(12), e0243300.
- Vuopala, A. (2010). “Assessment of the orphan works issue and costs for rights clearance.” *European Commission, DG Information Society and Media, Unit E4 Access to Information*, 42.
- Vyas, N., Kakade, S., and Barak, B. (2023). “Provable copyright protection for generative models.” *arXiv preprint arXiv:2302.10870*.
- Waldfoegel, J. (2010). “Music file sharing and sales displacement in the itunes era.” *Information economics and policy*, 22(4), 306–314.
- Waldfoegel, J. (2012). “Copyright Protection, Technological Change, and the Quality of New Products: Evidence from Recorded Music since Napster.” *Journal of Law and Economics*, 55(4), 715–740.
- Waldfoegel, J. (2017). “How digitization has created a golden age of music, movies, books and television.” *Journal of Economic Perspectives*, 31(3), 195–214.
- Waldfoegel, J. (2023). “The welfare effect of gender-inclusive intellectual property creation: Evidence from books.” *NBER Working Paper*.
- Waldfoegel, J., and Reimers, I. (2015). “Storming the gatekeepers: Digital disintermediation in the market for books.” *Information economics and policy*, 31, 47–58.
- Walterscheid, E. C. (1989). “Access to federally funded research data under the freedom of information act.” *Rutgers Computer & Tech. LJ*, 15, 1.

- Waters, R., Gilmour, D., and Mason, N. (2013). “Pink Floyd: Pandora’s Internet radio royalty ripoff.” <https://www.usatoday.com/story/opinion/2013/06/23/pink-floyd-royalties-pandora-column/2447445/>, last accessed: April 12, 2018.
- 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.
- Wen, W., Ceccagnoli, M., and Forman, C. (2016). “Opening up intellectual property strategy: Implications for open source software entry by start-up firms.” *Management Science*, 62(9), 2668–2691.
- Wen, W., Forman, C., and Graham, S. J. (2013). “The impact of intellectual property rights enforcement on open source software project success.” *Information Systems Research*, 24(4), 1131–1146.
- Whitaker, A. (2019). “Shared value over fair use: Technology, added value, and the reinvention of copyright.” *Cardozo Arts & Ent. LJ*, 37, 635.
- Windisch, M., and Peukert, C. (2021). “Copyright, re-use and digital business models.” *CREATE 21 for 2021 Blog*, <https://www.create.ac.uk/blog/2021/06/11/21-for-2021-copyright-re-use-and-digital-business-models/>.
- Wlömert, N., and Papiés, 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.
- Wu, L., Hitt, L., and Lou, B. (2020). “Data analytics, innovation, and firm productivity.” *Management Science*, 66(5), 2017–2039.
- Wu, Y., and Zhu, F. (2022). “Competition, contracts, and creativity: Evidence from novel writing in a platform market.” *Management Science*, 68(12), 8613–8634.
- Yang, C.-H. (2022). “How artificial intelligence technology affects productivity and employment: firm-level evidence from taiwan.” *Research Policy*, 51(6), 104536.
- Yilmaz, E. D., Naumovska, I., and Miric, M. (2023). “Does imitation increase or decrease demand for an original product? understanding the opposing effects of discovery and substitution.” *Strategic Management Journal*, 44(3), 639–671.
- YouTube (2021). “Copyright Transparency Report.” [https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-22\\_2021-7-1\\_2021-12-31\\_en\\_v1.pdf](https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-22_2021-7-1_2021-12-31_en_v1.pdf), last accessed: August 17, 2021.
- 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.
- Yuvaraj, J., Giblin, R., Russo-Batterham, D., and Grant, G. (2022). “Us copyright termination notices 1977–2020: Introducing new datasets.” *Journal of Empirical Legal Studies*, 19(1), 250–292.
- Zhang, L. (2018). “Intellectual property strategy and the long tail: Evidence from the recorded music industry.” *Management Science*, 64(1), 24–42.
- Zhong, H., Chang, J., Yang, Z., Wu, T., Arachchige, P. C. M., Pathmabandu, C., and Xue, M. (2023). “Copyright protection and accountability of generative ai: Attack, watermarking and attribution.” *arXiv preprint arXiv:2303.09272*.