FIU Law Review

Volume 14
Number 2 *Symposium: Intelligent Entertainment: Shaping Policies on the Algorithmic Generation and Regulation of Creative Works – Florida International University College of Law (Miami, Florida, 2019)*

Article 6

Fall 2020

Future of Copyright & Competition Law in Media

Dr. Matthew Eric Bassett
Professor, Barcelona Graduate School of Economics, me@mebassett.info

Online ISSN: 2643-7759

FUTURE OF COPYRIGHT & COMPETITION LAW IN MEDIA

Dr. Matthew Eric Bassett*

I. INTRODUCTION

In particular, this paper discusses how *data* could affect copyright and competition law. We shall begin with an overview of recent developments in the intersection of data, competition law, and privacy regulations in order to establish some common principles in the subject. Then we shall consider some scenarios from the film and entertainment industries and how they might interact more with such laws and regulations in the near future. In particular, we will consider the questions of "What if. . ."

1. audience data became the primary driver of content production and distribution decisions; and
2. content produced by AI and deep learning algorithms became the norm?

While these scenarios are necessarily speculative, we shall try to establish their relevance based on the state of the technology and your author's experience in helping film studios establish data-driven business models.

* Ph.D., Queen Mary University of London; Professor, Barcelona Graduate School of Economics; Research Fellow, Frankfurt Big Data Lab. This paper is an adaption of a keynote address given at the *FIU Law Review* Symposium, "Intelligent Entertainment: Algorithmic Generation and Regulation of Creative Works" held at Florida International University College of Law.

II. BACKGROUND ON AI, DATA AND COMPETITION LAW

The popular image of Artificial Intelligence is a robot-like C-3PO or Data from *Star Trek*, or a sinister computer-like SkyNet or HAL 9000. In reality, the term has been used to refer to a variety of programming techniques—symbolic computation and expert systems in the 1980s or statistics-driven deep learning algorithms of today. The vast majority of applications of artificial intelligence today are based on statistics, and that implies large amounts of data. This "modern" form of artificial intelligence consists of families of algorithms that take the data as input and produce as output a "calculator" for classification, prediction, or other tasks. Without going into technical details (and possibly by abusing analogies), these algorithms can be thought of as a form of compression: films, pictures, text messages, or other sources are represented in computers as the bits and bytes that make up data. These algorithms (and this paper will use the term "algorithm" to mean any artificial intelligence model, machine-learning model, or implementation thereof) try to extract a smaller set of numbers that go into a formula to perform whatever AI task the programmer was trying to achieve. In effect, they extract the numeric "essence" of the data with respect to whatever task the programmer had in mind. The effect of this is that many tasks that knowledgeable, trained humans would have performed can be replaced by these algorithms. Through the data, and metadata that humans have added, these algorithms have extracted the "essence" of much work in the knowledge economy and promise to accelerate economic growth in the same way the Industrial Revolution accelerated it when it replaced physical human labor with machines.

Just as the Industrial Revolution led to a few companies accumulating a monopoly through these machines and triggered modern anti-trust legislation, there is fear that the accumulation of data would allow new monopolies to form and must be regulated with a new perspective on competition law. Competition and antitrust regulation come from an unholy mix of politics, economics, and law. European regulators have been the keenest on regulating the use and collection of data from both the perspectives of competition and data privacy, which we illustrated in the following short, abbreviated timeline:

1. 2014—Directorate-General for Competition of the European Commission approves Facebook/WhatsApp merger without condition.[1]

[1] European Commission Press Release IP/14/1088, Mergers: Commission Approves Acquisition of WhatsApp by Facebook (Oct. 3, 2014).

2. 2016—EU Commissioner Vestager gives a speech acknowledging the relevance of data to competition law.[2] French and German competition authorities publish a joint report on big data.[3]
3. 2017—EU Commission carries out dawn raids as part of an investigation into alleged agreements by Polish banks to withhold data from Fintech rivals.[4] UK competition authority establishes Data team.[5]
4. 2018—GDPR comes into force.[6]
5. 2019—EU Commission President suggests that the Commission will consider "human and ethical implications of artificial intelligence."[7]

Arguably these latest developments illustrate how political competition law can become: The EU is said to be motivated by fear of large American and Chinese tech firms, the UK's Competition and Market Authority sees an opportunity in Brexit to establish itself as the leader in the field, and even major US Presidential candidates see a political opportunity in expanding the regulation of data. But this timeline also illustrates some principles that we shall see in our scenarios:

- Both firms (e.g., the Polish bank) and regulators (e.g., in the Facebook/Whatsapp deal) see data-drive as a defensive barrier to future competition.
- While algorithms are generally regarded as commodities, regulators see that they have a potential for misuse.

2 Margrethe Vestager, Executive Vice President for the European Union Commission for a Europe Fit for a Digital Age, (Sep. 29, 2016), https://wayback.archive-it.org/12090/20191129222113/https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/big-data-and-competition_en.

3 AUTORITE DE LA CONCURRENCE & BUNDESKARTELLAMT, COMPETITION LAW AND DATA (May 10, 2016), https://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Berichte/Big%20Data%20Papier.pdf.

4 European Commission Memo/17/3761, Antitrust: Commission Confirms unannounced Inspections Concerning Access to Bank Account Information by Competing Services (Oct. 6, 2017), https://ec.europa.eu/commission/presscorner/detail/en/MEMO_17_3761.

5 Stefan Hunt, *CMA's New DaTA Unit: Exciting Opportunities for Data Scientists*, GOV.UK (Oct. 24, 2018), https://competitionandmarkets.blog.gov.uk/2018/10/24/cmas-new-data-unit-exciting-opportunities-for-data-scientists/.

6 2016 O.J. (L119/1) 99.

7 *See* Etienne Bassot, *The von der Leyen Commission's Priorities for 2019-2024*, EUR. PARLIAMENTARY RES. SERV. (Jan. 2020), https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/646148/EPRS_BRI(2020)646148_EN.pdf.

- The accumulation of data is creating new fronts for regulators to protect rights, especially the right to privacy (GDPR).

With these three principles in mind, let us consider the first scenario.

III. *WHAT IF* AUDIENCE DATA BECAME THE PRIMARY DRIVER OF CONTENT DISTRIBUTION AND PRODUCTION DECISIONS?

The question suggests a notion of content producers collecting esoteric data on its audiences—from what portions of a piece of content they fast-forward through or to their facial expression while watching the same—to produce content specifically designed to be more engrossing. Those data points are important, and we shall return to them. But in reality, data-driven content decisions could appear much more innocuous to viewers while being no less fraught with subtle competition issues. The idea of data-driven content production entered mainstream thought when Netflix announced that its decision to greenlight *House of Cards* came from a data-driven model. The misconception was that the IP (script, plot, et cetera) was generated by algorithms. In actuality, the IP behind *House of Cards* had existed decades before as a British television show in the 1990s and as a book before that. The difference is how Netflix decided to greenlight the content.[8] Television and film productions involve a significant investment of capital, and any investor must do their due diligence before proceeding. Therefore, each piece of potential content undergoes a financial analysis. Investors are inherently risk-averse, and so content with a "track record" tends to do better. This explains why Hollywood is more likely to back a well-known director as well as why they produce so many franchises and sequels. Franchises and sequels have an existing fan-base and a solid set of demographic and marketing data from which they can extrapolate. It is not too difficult to imagine a world where investors require all such decisions to undergo a data-driven financial analysis. Indeed, the promise of more reliable, consistent returns on investments would be hard to pass up, especially if it could be demonstrated that the data-driven approach could deliver. We can extrapolate from current times to imagine what this future greenlighting process could look like: Financiers would demand data on the content's potential audience (perhaps even anonymized data points about individuals' moving-watching habits and expendable income) before any production funds would be made available. Because data has such strong network effects, the relevant data would most

8 Articles like David Carr's *Giving Viewers What They Want*, N.Y. TIMES (Feb. 23, 2013), https://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html, appear to contradict what I am saying. I ask the readers to trust me on this point as I worked in the industry during this time and shared a panel discussion with a Netflix executive in 2014.

likely accumulate in one or two large firms. A firm that gets better at providing data for accurate underwriting is likely to see its content funded more frequently, thus acquiring more audience data. Data would become a firm's defensive barrier, allowing it to underwrite and fund content production and depriving other firms of the opportunity. So, a lack of regulation on the accumulation of data or its use for such decisions could lead to a single firm could acquiring a "natural" monopoly on the ability to produce content.

It could be argued that this would not be bad for consumers. But one must also consider the sort of content for which this data would be available. As we have seen in today's world, such content means sequels and franchises, sometimes referred to as "known-IP." Not only would few firms be able to acquire the data necessary to conduct a financial analysis, but few firms would be able to acquire the rights to the underlying material. This would restrict the pool of "source material" for which data exists.

A. Content Distribution

Content production is only one side of the coin. The other major influential business process for the film and TV industries is content distribution. This side of the business tries to best answer the questions of when to release content, where to release content, and over which media. When it comes to the question of when content is distributed, film content distribution is already data driven. Your author has first-hand experience in establishing models that consider factors that would compete for audiences' attention in particular jurisdictions and using the algorithms to suggest the most financially optimal release date. But the questions of where and how are more complicated. To understand this, we must first discuss how film distribution works. The structure of cinema distribution in the US was established by the Hollywood Antitrust Case of 1948 that separated the operation of cinemas from film studios.[9] As a consequence, US film producers have had an indirect relationship with their audiences and distribution decisions have been largely driven by gut feeling (for dates) or personal relationships (cinema pricing). This setup is mostly followed around the world with the notable exceptions of France and South Korea, where studios and distributors are vertically integrated. As a consequence, cinemas and film distributors have a frequently antagonistic relationship. Distribution companies (major studios) have to convince cinemas to carry their content, and therefore offer them "rental" income, the price of which is typically negotiated weekly per piece of content. So, film distributors are competing

---

[9] *See generally* U.S. v. Paramount Pictures, Inc., 334 U.S. 131 (1948).

against each other for screen space and are bidding against each other via the rental price. The consumers are not the audience, but the exhibitors. And a decision about where to release a film becomes a decision about how much to bid in a particular cinema. A data-driven distribution strategy means that distributors and exhibitors would need data on the local demographics (including personal data as described before, perhaps acquired via loyalty programs with the cinema, which we already see today), along with data on the other *local* factors competing for audiences' attention. For both exhibitors and distributors to book the most optimal screens for the most optimal films, both would need to rely on the same dataset. This dataset would need to be consumed by an algorithm that would then suggests the optimal booking. At first glance this would sound as though it would eliminate anti-competitive behavior—especially price collusion. But datasets can be inherently, or even purposely, biased, and algorithms are black boxes that can disguise nefarious uses. Even before the advent of big data, as early as the 1990s, it was already established that it would be difficult for regulators to spot price collusion in digital systems.[10] Data-driven, machine-learning systems could be biased by their input data and thus could allow firms to engage in automated price collusion without any direct communication between firms. In this way, the use of machine learning algorithms for key business processes without regulation or inspection of the underlying data can create (or obfuscate) potential competition issues under today's statues. This has already been academic work on strategies for detecting some AI-enabled collusion.[11]

B. Advertisement-supported Content

The previous discussion centered around content that consumers pay for directly. But a large portion of media is financially supported, at least in part, by advertisements. It has been well-established for some years now that certain meta-data on a population can allow firms to produce highly-targeted, effective marketing materials.[12] It is perhaps less well-known that media companies have been developing algorithms to assess the emotional impact of the content they produce.[13] As is often the case in developing machine

10 Severin Borenstein, *Rapid Price Communications and Coordination: The Airline Tariff Publishing Case (1994)*, *in* THE ANTITRUST REVOLUTION: ECONOMICS, COMPETITION, AND POLICY 236 (J.E. Kwoka & L.J. White eds., 1999).

11 *See generally* Ariel Ezrachi & Maurice E. Stucke, *Artificial Intelligence & Collusion: When Computers Inhibit Competition*, 5 U. ILL. L. REV. 1775, 1775 (2017).

12 *See* Kashmir Hill, *How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did*, FORBES (Feb. 16, 2012, 11:02 AM), https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/.

13 *Contextual Moments*, 4SALES, https://www.4sales.com/contextual-moments (last visited April 20, 2020). Your author personally knew the executive managing the project.

learning models, the issue or opportunity here is not in the existence of these two particular datasets, but their combination. If meta-data from audiences' behavior while watching a piece of content—say data from a device's camera so that an algorithm can infer a person's attention or emotional state from their facial expression, as well as what portions of the content they find most engrossing—can be combined with data on the emotional impact of the content itself then an algorithm could learn the "right" moment to advertise a specific product. For example, meta-data from one source could suggest, via an algorithm, that a person's parent recently passed away and that said parent had a fondness for a particular vehicle. While the same person watches a piece of content, an algorithm observes that they are particularly engrossed and open to persuasion, so the content is interrupted for an advertisement of that same vehicle. This example might be far-fetched or distasteful, but it illustrates what could be achieved for advertisers, and more importantly, illustrates the issues around data. Society takes the view that the vulnerable should be protected from manipulation or predatory trade practices, yet it is not clear how existing laws and data protection regulations could be used to prevent persons from being emotionally manipulated by content. This issue becomes more fraught when one considers algorithm-generated content.

IV. *WHAT IF* CONTENT PRODUCED BY AI AND DEEP LEARNING ALGORITHMS BECOME THE NORM?

This scenario is much more speculative than the last. So, let us begin by establishing that the principles of it are very much reality and not science fiction. [14] The reader is no doubt aware that deep learning algorithms are able to generate photo-realistic videos, superimpose faces from still images onto videos of other people, and use those same still photos to create videos of the subject moving and talking. We have also seen this technology used to re-create the appearances of deceased actors as their younger selves in modern cinema[15] It should also be noted that other deep learning algorithms have demonstrated the capability to re-create written words that closely match the tone of a particular author or speaker. For instance, if your author used the preceding few sentences as a "seed" for one such algorithm, it would result in the following text:

---

14 THIS PERSON DOES NOT EXIST, https://www.thispersondoesnotexist.com/ (last visited April 20, 2020). Some "friends" of your author. These photographs were generated by an algorithm and do not represent real human beings.

15 *See* Rebecca Hawkes, *From Rogue One's Peter Cushing to Audrey Hepburn: 6 Stars Who Were Digitally Brought Back to Life*, THE TELEGRAPH (Dec. 20, 2016, 7:20 PM), https://www.telegraph.co.uk/films/0/rogue-ones-peter-cushing-audrey-hepburn-stars-digitally-brought/.

> "There were six men in the room and I'm going to tell you what they were saying." These algorithms are very powerful. One can imagine using them for voice recognition systems to identify human speech. Or for example, using neural networks to determine how to make a face from different images or a word from one language to another. However, there are two ways that these deep learning algorithms can go wrong. One is that they can end up generating random noise.[16]

Further, we have already discussed how film and television content must go through a risk-averse greenlighting process. Since talented actors have a loyal audience for which data can be gathered, it is not hard to conclude that a greenlight process would favor content with established talent. Let us imagine a world where algorithms could recreate their likeness in arbitrary roles. The likeness of a particular actor would be a valuable piece of Intellectual Property (as, indeed, it is today). But it is less clear what would consist of infringement of another person's intellectual property. For instance, if producers found that the personality of a particular comedian was popular (or favorable to the greenlighting algorithms discussed previously), they could use a text generation algorithm such as the one demonstrated to generate content that matches the comedian's personality and use a video-generation algorithm to create a video of a non-existent person (who resembles the original comedian) delivering the text. They could, in effect, capture the personality of an actor and use it to generate new original content. If comedians, actors, and talent agencies could see their "likeness" easily reproduced, then they may seek to protect it, and that would require expanding copyright to include personality and other social skills. We have already seen the beginnings of such copyright issues in the music industry in a recent music copyright case involving a Katy Perry song, Dark Horse. The jury's verdict in the court case (as your author understood it) suggested that even content generated by random algorithms (never mind intelligent algorithms) could be protected by copyright.[17]

With new concepts of copyright and intellectual property come new possibilities to infringe. Consider if an algorithm were to create the likeness of a non-existent person (like those people pictured above), and that non-existent person resembled closely enough an existing actor (or even another

---

16 *See* Alec Radford et al., *Language Models Are Unsupervised Multitask Learners*, OPENAI (2019), https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf. This was implemented at TALK TO TRANSFORMER, https://talktotransformer.com/. Your author took the first suggestion and did not edit it.

17 Mark Savage, *Katy Perry Wins Reversal of Dark Horse Plagiarism Verdict*, BBC NEWS (March 28, 2020), https://www.bbc.co.uk/news/entertainment-arts-51942894.

non-existent person generated by an algorithm that has already been used in revenue-generating content) that its owners could reap the benefits of the actor's fan base, would this constitute a violation of intellectual property rights? If generated, non-existent people were to become commonplace or if the "likeness" of someone was protected by law or regulation, then each generated item is at risk of a violation. As discussed at the start of this paper, the algorithms that generate these non-existent persons are a direct product of the data used to train them. Presumably, this data would be past revenue-generating content. Therefore, one would expect that these algorithms would generate similar non-existent persons. This could evolve into a scenario where firms accumulate the intellectual property associated with valuable actors, et cetera, in much the same was as tech companies accumulate patents: as a method of mutually assured destruction.[18] So, agencies that represent talent would compete to acquire as much intellectual property for as many persons as they could so they would always have a strong offensive strategy if they ever accidentally infringed. Indeed, hidden in that last sentence was the idea that talent agencies would be representing intellectual property—whether it is the likeness of a real person or not—or even algorithms themselves. Additionally, firms would also compete to acquire the "training data," that is, original revenue-generating content. A firm that held a large back catalog of films and television series would have a significantly easier time producing new content just from that data. This again highlights the principle of data accumulating into a handful of firms and denying other firms the possibility to compete.

V. CLOSING REMARKS

This paper began with three principles of how data pertains to competition, and these examples were meant to illustrate how those same principles apply to the film and entertainment industries. Some of these issues around data and competition could be addressed, naively, by forcing large firms to share data sets with smaller firms. This would create an entirely different set of risks, especially around violating consumers' right to privacy.

---

18 Wendy Seltzer, *Software Patents and/or Software Development*, 78 BROOK. L. REV. 929, 933–934 (2013). In particular,

> When Google announced its agreement a few weeks later to acquire Motorola Mobility Inc. for $12.5 billion, the acquisition of a major mobile hardware manufacturer was widely read as a purchase of a defensive portfolio of mobile software patents as a means to protect Google's Android mobile operating system. . . From a strategic perspective, there were plenty of reasons for Google not to enter the hardware business, but it apparently became clear that without patents to counter-assert against patent attackers, Android would lose the confidence of other hardware makers and fail.

Id.

Indeed, most consumer data protection attempts to restrict firms from sharing with third parties without the consent of the data subject. Yet deep learning models could provide a middle ground. "Transfer learning" is a process whereby a deep learning algorithm is trained a primary data set for a general task (e.g., identifying which part of a photograph is in the focus of a camera), and then the trained-algorithm is re-trained on a smaller, more specific data set for a more precise use case (e.g., identifying a face in a photograph). As the trained model is, in essence, a long list of numbers, this could be shared with minimal risk to data subjects. Many deep learning practitioners already share models this way.[19]

It is also worth keeping a healthy sense of skepticism with these examples. While they were certainly not exhaustive, they were hypothetical and quite speculative; many of these issues may never arise. Machine learning, big data, et cetera has an extraordinary capacity to both be frighteningly effective while at the same time quite underwhelming. For instance, a firm your author helped start, Gower Street Analytics, was credited with using its advanced datasets and algorithms to pick the release date for Michael Moore's last film. The chosen date for the said political documentary was a September date six weeks before an election. Arguably a film industry veteran could have made that call without computer assistance.

Finally, it is generally thought that the purpose of competition law is to ensure consumers have choice (in price, quality, et cetera).[20] The earlier example of emotional advertising was meant to highlight that some potential data abuses go behind protecting consumers. Competition law is so intertwined with politics because it is about regulating power. The data firms collect on consumers, audiences, or others gives them the power to influence consumer behavior and public opinion as well as the market. However skeptical one might be about the business value of machine learning, it is increasingly important to consider how to effectively regulate it.

---

19 *See* Sinno Jialin Pan & Qiang Yang, *A Survey on Transfer Learning*, 22 IEEE TRANSACTIONS ON KNOWLEDGE & DATA ENGINEERING 1345, 1345 (2010).

20 *See generally* Paul H. Brietzke, *Robert Bork, The Antitrust Paradox: A Policy at War with Itself*, 13 VAL. U. L. REV. 403 (1979).