Seton Hall University

eRepository @ Seton Hall

2022

Pirates, AI, and Privacy: The Use of AI in Combatting Media Piracy Online

Lauren P. Haberstroh

Pirates, AI, and Privacy: The Use of AI in Combatting Media Piracy Online
Lauren P. Haberstroh

I. Introduction

Throughout human history, wherever a business has sold pieces of art and entertainment for profit, people have found ways to obtain them for free. Piracy involving "the unauthorized reproduction or use of an invention or work of another...especially as constituting an infringement of patent or copyright,"[1] in Western societies can be traced as far back as the 17th century, where a young Wolfgang Amadeus Mozart allegedly attended the performance of Allegri's *Miserere* and later illegally transcribed the privately-owned piece's sheet music from memory.[2] Publishers and entities such as the Catholic Church attempted to control the circulation and content of texts after the advent of the printing press,[3] but it was inevitable that such new technology created for the primary purpose of sharing information would have a "dual effect" of creating an easy method by which copyright can be violated through the illegal sharing of materials.[4]

This "dual effect" was most vividly felt with the 1999 release of Napster, a digital music-sharing service that brought piracy to the mainstream.[5] Where physical copies of books, software, and media were once unreasonably difficult to copy and share without first purchasing a copy of said product, the advent of easily-accessible digital file sharing via the internet made obtaining copyrighted materials as simple as downloading a peer-to-peer (P2P) file sharing software and choosing from a lengthy list of songs and movies available for free.[6] Napster's

---

[1] *Piracy*, OXFORD ENGLISH DICTIONARY (3rd ed. 2006).
[2] EDWARD HOLMES, THE LIFE OF MOZART: INCLUDING HIS CORRESPONDENCE 69-70 (1845).
[3] JOE KARAGANIS ET. AL., MEDIA PIRACY IN EMERGING ECONOMIES 399 (Joe Karaganis ed. 2011).
[4] NATHAN FRISK, UNDERSTANDING ONLINE PIRACY: THE TRUTH ABOUT ILLEGAL FILE SHARING 9-10 (2009).
[5] Joel Waldfogel, *Copyright Protection, Technological Change, and the Quality of New Products: Evidence from Recorded Music since Napster*, 55 J.L. & ECON. 715, 715-16 (2012).
[6] *Id*. at 716.

reign would be short-lived; though the use of P2P file sharing was and is legal, the sharing of copyrighted materials was not. Napster's creator was sued by a litany of U.S. music record companies for contributory and vicarious infringement of the plaintiffs' copyrights in 2000.[7] The United States Court of Appeals for the Ninth Circuit denied his fair use defense and confirmed his conviction in 2001, holding that while the service could be used for legal means, Napster had both "actual and constructive" knowledge of its users direct infringement[8] and could prevent its users from engaging in direct infringement, which gave Napster a duty to do so.[9] The court enjoined Napster from hosting "uses of copyrighted material that [were] not fair use"[10] and Napster officially shut down its services on July 11, 2001.[11]

However, this modern form of piracy, often dubbed "digital piracy," could not so easily be stopped. Several P2P file-sharing programs sprung to life after Napster's shutdown, this time utilizing the decentralized P2P file-sharing protocol BitTorrent,[12] which, while legal, is often used in conjunction with sites that host magnet links to torrent files that contain copyrighted materials. Perhaps the most infamous of these is the Pirate Bay, a site that hosts user-uploaded links that can be used to torrent copyrighted material, though its founders argue that it does not host infringing files on its servers nor link to them on its website – it only hosts "trackers," or "files that tell BitTorrent apps which other app users to link to in order to download large files."[13] BitTorrent apps and the Pirate Bay, as well as sites providing a similar service, still operate today. It is also much easier to engage in digital piracy in the present year: torrent

---

[7] *A&M Records, Inc. v. Napster, Inc*, 114 F.Supp. 2d 896 (2000).
[8] *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1019 (2001).
[9] *Id*. at 1027.
[10] *Id*. at 1028.
[11] *Napster is to Remain Shut*, N.Y. TIMES, July 12, 2001 at 7.
[12] *About BitTorrent*, BITTORRENT, https://www.bittorrent.com/company/about-us/ (last visited Apr. 10, 2020)
[13] Alex Hern, *European Court of Justice Rules Pirate Bay is Infringing Copyright*, Guardian (Jun. 15, 2017) https://www.theguardian.com/technology/2017/jun/15/pirate-bay-european-court-of-justice-rules-infringing-copyright-torrent-sites (last visited Apr. 10, 2020).

download clients are easy to obtain online and are often free,[14] and virtual private networks

(VPNs) or internet proxies that utilize VPN technologies can be purchased to mask a user's

internet protocol (IP) address so that copyright holders cannot trace the illegal download of their

materials to a user via that user's Internet Service Provider (ISP).[15]

Today, U.S. citizens pirate books, video games, movies, songs, and television shows with

wild abandon. U.S. book publishers lose $300 million in income annually to eBook piracy.[16] The

U.S. Chamber of Commerce's Global Innovation Policy Center estimates that global piracy of

U.S. digital videos (i.e. films and episodes of television shows) has cost the country around $29.2

billion in losses per year,[17] while piracy of digital music costs the U.S. $12.5 billion in output

annually.[18] Video games, sold either digitally online or physically on a disc or cartridge, are

often distributed with digital rights management (DRM) tools in an effort to prevent piracy, but

the U.S. still lost approximately $74 billion in sales revenue in 2014 due to video game piracy.[19]

These figures are staggering, but how is a copyright holder supposed to monitor and prevent

---

[14] One of the most well-known torrent client providers, BitTorrent, offers downloads for a free basic version of the client as well as paid versions which provide VPNs, remove advertisements from the program, and block malware, depending on the version purchased. *BitTorrent Classic*, BITTORRENT, https://www.bittorrent.com/products/win/bittorrent-classic-free/ (last visited Apr. 11, 2021).

[15] Like BitTorrent clients, VPNs are also provided for free or for a monthly fee with additional services provided by various companies. Stefan Larsson, et. al, *Law, Norms, Piracy and Online Anonymity: Practices of De-Identification in the Global or File Sharing Community*, 6 J. RES. IN INTERACTIVE MKT. 260, 263 (2012).

[16] Adam Rove, *U.S. Publishers are Still Losing $300 Million Annually to eBook Piracy*, FORBES (Jul. 28, 2019) https://www.forbes.com/sites/adamrowe1/2019/07/28/us-publishers-are-still-losing-300-million-annually-to-ebook-piracy/?sh=39255732319e (last visited Apr. 11, 2021).

[17] DAVID BLACKBURN, ET. AL., IMPACTS OF DIGITAL VIDEO PIRACY ON THE U.S. ECONOMY 12 (2019).

[18] STEPHEN E. SIWEK, THE TRUE COST OF SOUND RECORDING PIRACY TO THE U.S. ECONOMY 1 (2007).

[19] Luke Graham, *Can Video Game Piracy be Stopped in Two Years?*, CNBC (Jan. 14, 2016) https://www.cnbc.com/2016/01/14/can-video-game-piracy-be-stopped-in-two-years.html (last visited Apr. 10, 2021).

piracy of its copyrighted materials when 93%[20] of America's total population of 300 million[21] uses the internet daily via two thousand different ISPs?[22]

Artificial Intelligence (AI) may provide a solution for the tidal wave of digital piracy that copyright holders face. AI generally refers to a computer's ability to exhibit intelligent behavior.[23] Alan Turing grappled with the concept of machine intelligence as early as the 1950s, but it was not until the late 1990s and early 2000s that important goals of AI were truly achieved, such as the defeat of a human world champion chess player by IBM's AI *Deep Blue*.[24] This paper will mostly focus on machine learning (ML), a branch of AI concerned primarily with giving training data to an algorithm (a set of statistical processing steps) with the goal of producing a specific output; a "trained, accurate" algorithm is called a "machine learning model."[25] Machine learning is currently used in a variety of ways: Chatbots utilize it with natural language processing to provide automated responses to a person's questions,[26] and Walmart combines it with the Internet of Things (IoT)[27] to track inventory and allow customers to pay for certain items in-store via Walmart's phone app.[28] Put simply, machine learning excels at taking large amounts of information and filtering out unnecessary data to achieve an intended result (i.e. extracting features from images to highlight inconsistencies) without need for human

---

[20] *Internet/Broadband Fact Sheet*, PEW RESEARCH CENTER (Apr. 7, 2021) https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ (last visited Apr. 10, 2021).

[21] *U.S. and World Population Clock*, UNITED STATES CENSUS BUREAU (Apr. 10, 2021 2:05 PM), https://www.census.gov/popclock/ (last visited Apr. 10, 2021).

[22] *The Complete List of Internet Service Providers in the U.S.*, BROADBAND NOW, https://broadbandnow.com/All-Providers (last visited Apr. 11, 2021).

[23] *Artificial Intelligence*, OXFORD ENGLISH DICTIONARY (3rd ed. 2006).

[24] Rockwell Anyoha, *The History of Artificial Intelligence*, HARVARD UNIVERSITY (Aug. 28, 2017) https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/ (last visited Apr. 10, 2021).

[25] IBM Cloud Education, *Machine Learning*, INTERNATIONAL BUSINESS MACHINES INC. (Jul. 15, 2020) https://www.ibm.com/cloud/learn/machine-learning (last visited Apr. 11, 2021).

[26] *Id*.

[27] "The Internet of Things" refers to concept of everyday objects sending data to and receiving data from a database via microchips which provide network connectivity. *Internet*, OXFORD ENGLISH DICTIONARY (3rd ed. 2006).

[28] Bernard Marr, *How Walmart is Using Machine Learning AI, IoT and Big Data to Boost Retail Performance*, BERNARDMARR.COM, https://bernardmarr.com/default.asp?contentID=1181 (last visited Apr. 10, 2021).

intervention.[29] The U.S. is already seeing machine learning used by ISPs. In response to shifts in internet usage caused by COVID-19 isolation, Verizon plans to use machine learning to monitor its users' network usage so that it can better respond to the unprecedented increase in online gaming, video traffic, and streaming via planned network maintenance.[30]

This paper will discuss why the haphazard application of AI in preventing piracy, while convenient and beneficial to copyright holders, poses risks to the average end-user's privacy and ability to create art and express opinions. Part II will briefly discuss the history of the United States' approaches to copyright protection, concluding with an examination of the balance between preventing copyright infringement with allowing internet users to express themselves through speech and works. Part III will present pro-piracy arguments, discussing its use in preserving art, distributing educational resources, and increasing a product's publicity and future sales. Part IV of this paper will describe the current and speculated ways in which AI can detect piracy, specifically the unauthorized streaming or sharing of copyrighted videos, and how digital pirates continue to circumvent these methods. Part V will discuss and compare the United States' and European Union's approaches to privacy rights and how an end-user's privacy will be endangered by the application of AI in scanning various websites to detect and prevent piracy It will also contrast these arguments with the harmful effects that piracy can have on a business or industry. Part VI will conclude, reiterating the importance of societal and legal change before AI can be used to regulate content more efficiently on the internet.

---

[29] IBM, *supra* note 25.
[30] Kyle Wiggers, *How ISPs are Using AI to Address the Coronavirus-driven Surge in Traffic*, THE MACHINE (Mar. 27, 2020), https://venturebeat.com/2020/03/27/how-isps-are-using-ai-to-address-the-coronavirus-driven-surge-in-traffic/ (last visited Apr. 10, 2021).

II. <u>Rules of the Seas: Copyright Law and Speech on the Internet</u>

Digital piracy and its prevention are planted squarely in the center of several competing issues. It infringes on the rights of copyright holders, who want to take whatever measures necessary to prevent detect and stop it, but what if such measures trample on the rights of innocent people who are merely suspected of pirating digital media? What if a copyright holder alleges copyright infringement where there was fair use instead? Do prominent video-hosting sites such as YouTube have an obligation to facilitate free speech on their platform, despite being a private company? Should they? These questions have only grown more prevalent as AI and ML models have been employed in detecting copyrighted materials hosted on private online platforms. The European Union (E.U.) and the U.S. still attempt to stop digital copyright infringement through new and proposed legislation that hampers and encourages the development of AI, respectively.

A. <u>Copyright Law in the United States and the European Union</u>

The origins of U.S. copyright law can be traced to Article I, section 8 of the U.S. Constitution, which gives Congress the authority to enact legislation that "promote[s] the Progress of Science and the useful Arts, by securing for a limited Times to Authors and Investors the exclusive Right to their respective Writings and Discoveries."[31] After Congress passed the first copyright act in 1790, [32] the list of works protected by copyright expanded to encompass musical compositions, etchings, engravings, photographs, moving pictures, and sound recordings

---

[31] U.S. CONST. art. I, § 8, cl. 8.
[32] The Copyright Act of 1790 offered limited rights to copyright holders, giving them limited rights to 'printing, reprinting, publishing, and vending" for only fourteen years. 1 STAT. 124, 1 CONG. CH. 15.

over the course of 190 years.[33] The modern list of protected works is now quite extensive, containing graphic works, pantomimes, choreographic works, and more.[34]

Congress forever changed American copyright law was in 1998, when, in response to a growing concern over copyright protections on the internet, it enacted the Digital Millennium Copyright Act to implement two 1996 treaties of the World Intellectual Property Organization.[35] 17 U.S.C.S. §512, also known as the DMCA safe harbor provision, exempts an ISP from liability for copyright infringement that takes place over its network if: (1) the transmission of copyrighted materials was not done by or at the direction of the ISP; (2) "the transmission, routing, provision of connections, or storage" occurs automatically without the ISP selecting the material; (3) the ISP does not manually select the recipients of the copyrighted material; (4) the ISP doesn't make a copy of the infringing material, store it on its system, or create a copy on its system or network that is accessible to anyone besides the intended recipient for a longer period than reasonably necessary for its transmission, and; (5) the copyrighted material is transmitted through the network/system without being modified.[36] Additionally, an ISP will not be liable for storing copyrighted material if it does not have: (1) actual knowledge that the material on the network is infringing; (2) in the absence of actual knowledge, the ISP is not aware of facts or circumstances from which infringing activity is apparent, or; (3) upon obtaining knowledge of copyright infringement, the ISP acts "expeditiously" to remove the material or access to it.[37] The United States Court of Appeals for the Second Circuit found that the "actual knowledge" described by the DMCA referred to whether the provider "subjectively" knew of a specific

---

[33] Leon Solomon, *Fair Users or Content Abusers: The Automatic Flagging of Non-Infringing Videos by Content ID on YouTube*, 44 HOFSTRA L. REV. 237, 240 (2015).
[34] *Id*. at 241.
[35] DIGITAL MILLENNIUM COPYRIGHT ACT, 1998 Enacted H.R. 2281, 105 Enacted H.R. 2281, 112 Stat. 2860.
[36] 17 U.S.C.S. §512(a)(1)-(5).
[37] 17 U.S.C.S. §512(c)(1)(A).

infringement, while the "red flag provision" (whether the ISP is aware of facts or circumstances that would make infringing activity apparent) referred to an objective standard, namely whether the ISP was subjectively aware of facts that would have made infringement "objectively obvious to a reasonable person."[38]

One limitation on the deference given to copyrighted-content holders, a defense used by Napster's creator in court,[39] is fair use under 17 U.S.C. §107. The fair use doctrine lists four guiding factors that a court must consider when determining if a use of copyrighted material is protected: (1) the purpose and character of the use, including whether such use is of a commercial nature or is for educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market or value of the copyrighted work.[40] While the fair use doctrine applies to works of commentary or criticism, a work is not immediately protected by the doctrine just because it fits into one of those categories – it must be sufficiently transform the original work that it is commenting on.[41] While the wholesale reproduction of a work can be transformative if placed in a "new context to serve a different purpose," the secondary use must benefit society by "imbuing the original with new function or meaning."[42]

The E.U. similarly struggles with preventing the unauthorized digital sharing of copyrighted materials, and this struggle involves the three main E.U. copyright law directives:

---

[38] *Viacom Int'l, Inc. v. YouTube, Inc.*, 676 F.3d 19, 31 (2d Cir. 2012).
[39] *A&M Records*, *supra* note 10.
[40] 17 U.S.C. §107.
[41] *Brammer v. Violent Hues Prods.*, 922 F.3d 255, 263 (4th Cir 2019).
[42] *Id*.

the Copyright Term Directive[43], the Information Society Directive,[44] and, most importantly, the Directive on Copyright in the Digital Single Market (hereafter "the directive").[45] The E.U.'s official website states that the purposes of the directive are to adapt key exceptions to copyright protection to the digital and cross-border environment; to ensure wide access to content and improve the E.U.'s licensing practices, and; to "achieve a well-functioning marketplace for copyright."[46]

The directive contains several controversial articles, but most important are those involving AI and copyright liability for social media sites.[47] Article four creates a copyright exception for text and data mining for scientific research, but depending on whether it acknowledges the public domain status of certain facts or info, this could increase or decrease restrictions for AI[48] that relies on text data mining ("TDM") for training.[49] Article four also allows copyright holder to opt out of the exemption for scientific research, meaning that AI depending on TDM is severely hampered from developing.[50] Article seventeen is especially

[43] The Copyright Term Directive ensures a single duration for copyright (70 years) and related rights (50 years) across the E.U. It also lists the methods of copyright restoration and unifies the treatment of photos and videos. Council Directive 2006/116/EC, 2006 O.J. (L. 372) 12.

[44] The Information Society Directive distinguishes between copyright and related rights, lists exceptions in cases of photo reproductions, reproductions of private works, and archival reproductions. It also requires member states to give "adequate legal protection" against intentional circumvention of "effective technological measures" designed to prevent or restrict acts of unapproved copying. Council Directive 2001/29/EC, 2001 O.J. (L. 167) 10.

[45] Council Directive 2019/790, 2019 O.J. (L. 130) 92.

[46] *Copyright and Related Rights in the Digital Single Market: Summary of the Directive on Copyright in the Digital Single Market*, EUR-LEX (last updated Dec. 7, 2019), https://eur-lex.europa.eu/legal content/EN/LSU/?uri=uriserv:OJ.L_.2019.130.01.0092.01.ENG.

[47] *See also* Council Directive 2019/790, art. 5, 2019 O.J. (L. 130) 92 (creates mandatory exception for copyrighted works as part of "digital and cross-border teaching activities." Article five has been criticized for its narrow definition of "educational establishments," which do not include cultural heritage institutions and requires teachers to ensure that there are no "adequate licenses" for the material available on the market. *See* Javiera Atenas, *Educators Ask for a Better Copyright*, Open Education Working Group (Jan. 16, 2018), https://web.archive.org/web/20180705150720/https:/education.okfn.org/educators-ask-for-a-better-copyright/.

[48] AI that requires TDM training can be employed in scanning websites for copyrighted text. Elena Riva, *The Copyright Directive is a Warning Signal for Europe's AI Ambitions*, INLINE (Apr. 15, 2020), https://www.inlinepolicy.com/blog/copyright-directive-warning-signal.

[49] Council Directive 2019/790, *supra* note 45 at art. 4.

[50] *Id*.

controversial, as it targets commercial web hosts who "store and give the public access to a large number of works or other subject-matter uploaded by its users which [they] organize and promote for profit-making purposes."[51] This means that social media sites such as Twitter, YouTube, and Facebook are liable for copyright infringement that occurs on their platform unless they can prove that they: (1) made their best efforts to obtain an authorization; (2) made their best efforts to ensure that the specific copyrighted work was unavailable, if the copyright holder provided them with relevant and necessary information, and; (3) upon receiving sufficiently substantiated notice from a copyright holder, acted expeditiously to disable access to or remove the content from their website and made best efforts to prevent said content from being uploaded again.[52] To determine if service providers adequately complied with Article 17's requirements, two factors should be accounted for: (1) the "type, audience, and size of service and type of works...uploaded by users of the service," and (2) the availability of effective and suitable means for the service providers to expeditiously remove the offending content.[53]

The balancing act of preventing the unauthorized distribution of copyrighted material while allowing it to be used for educational and artistic purposes is a difficult one, as neither the U.S. nor the E.U. have perfected it. Comparing the legality of AI TDM between U.S. and E.U. reveals the benefits and drawbacks of both forms of copyright law, especially regarding their potential relationship to AI being used to detect and stop the unauthorized sharing of unauthorized content. While the E.U. is attempting to create a "well-functioning market for copyright," it has instead created what European tech companies fear will be an unreasonable

---

[51] *Id*. at art. 17.
[52] Council Directive, *supra* note 45 at art. 17.
[53] *Id*.

amount of obligations and restrictions in moderating content on social media.[54] Some

commentors fear that the directive will lead to social media sites utilizing strict upload-filters for

content posted by E.U. citizens,[55] which would certainly not be in line with the E.U.'s stated

goals of having the directive carve out copyright exceptions and ensure a wide access to

content.[56] Article four's opt-out provision for copyright holders regarding TDM ML models also

harms the E.U., as it slows the development of technology that could be used to scan websites

and help prevent piracy, thus reducing the need for measures like upload-filters.[57] In comparison,

the U.S. Courts of Appeals have often found TDM to fall under fair use protection,[58] which

allows AI utilizing TDM to develop and potentially be used to fight digital piracy. However, AI

that is trained to detect piracy can often flag and remove content that was protected by fair use

laws, which courts emphasize must be applied in a flexible manner.[59]

Based on this application, it seems that governments must choose between two evils:

having an AI filter through massive amounts of content for copyrighted material at the risk of

suppressing material that is protected by fair use or other exceptions, or taking more labor-

intensive or restrictive measures, such as hiring employees to manually review and take action

against a tremendous amount of content or restricting exactly what kind of content can be

uploaded in the first place, which would either do little to stop the flood of copyrighted material

constantly posted to social media or suppress speech on popular and important platforms. The

E.U.'s opt-out provision for scientific works in the public domain unnecessarily stifles the

---

[54] Ally Boutelle and John Villasenor, *The European Copyright Directive: Potential Impacts on Free Expression and Privacy*, Tech Tank (Feb. 2, 2021), https://www.brookings.edu/blog/techtank/2021/02/02/the-european-copyright-directive-potential-impacts-on-free-expression-and-privacy/.
[55] *Id*.
[56] EUR-Lex, *supra* note 46.
[57] Council Directive, *supra* note 52.
[58] *See generally Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014); *A.V. v. iParadigms, LLC* (4th Cir. 2000); *Perfect 10 v. Amazon*, 508 F.3d 1146 (9th Cir. 2007), and; *Kelly v. Ariba Soft*, 336 F.3d 811 (9th Cir. 2003).
[59] Krista L. Cox, Issue Brief: Text and Data Mining and Fair Use in the United States, 2 (Jun. 5, 2015).

growth and development of AI in aiding human employees in their moderation of copyrighted or illegal content. It is unrealistic to expect social media sites with user counts in the billions to hire enough and train enough employees to effectively monitor an entire website for copyrighted or elicit content; AI will need to be used eventually, and it is best that it be trained on data that is publicly available. Unfortunately, it currently seems that AI cannot account for the numerous exceptions to copyright laws and regulations, especially concerning content hosted on privately-owned sites.

B. Free Speech, Social Media, and AI

United States citizens are guaranteed freedom of speech without abridgement by Congress via the First Amendment,[60] and the United States Supreme Court has held that any regulations which discriminate types of speech based on their content would be subject to a strict scrutiny analysis.[61] Things have changed drastically since the enactment of the First Amendment; a large portion of public speech and human interaction no longer occurs in public forums, or even in person, but on social media platforms, where ordinary people with internet access can contribute to the marketplace of ideas at their leisure.[62] The U.S. has extensively changed and updated its legislation concerning the internet to promote the exchange of free ideas online: in 2015, the Federal Communications Commission (FCC) classified ISPs as common carriers which prevented them from censoring content,[63] but this was rolled back in 2018, when FCC Chairman Ajit Pai rescinded the common carrier classification, allowing ISPs to slow traffic to sites as they saw fit, or to bundle internet packages and only allow their users to access

---

[60] U.S.C.S. CONST. AMEND. 1.
[61] *Police Dep't of Chi. v. Mosely*, 408 U.S. 92 (1972).
[62] Colby M. Everett, *Free Speech on Privately-Owned Fora: A Discussion on Speech Freedoms and Policy for Social Media*, 28 KAN. J.L. & PUB. POL'Y 113, 119 (2018).
[63] 30 FCC Rcd. at 5601

certain sites in said bundle.[64] In section 230 of the Communications Decency Act, Congress

established the internet as the country's main forum for speech, opportunity, and intellectual

progress, and immunized users and webhosts alike (i.e. Facebook and Twitter) from tortious

liability for the copyright infringement of their users.[65]

The strict parameters established by the First Amendment and the Supreme Court cases

interpreting it only apply to federal restrictions of free speech, but there has been a recent push[66]

to enact stricter regulations for popular social media sites such as Facebook and Twitter, which

not only facilitate the sharing of speech on a grand scale, but were allegedly used by Russian

agents to propagate false information during the 2016 United States elections.[67] Arguments in

favor of the federal regulation of privately-owned social media sites focus on the fact that these

important platforms are regulated not by professionals, but by profit-motivated companies who

set their own arbitrary guidelines.[68] The power social media sites and their staff have to delete

user-generated content can be and often times is used for removing unpopular views, it is argued,

amounting to "constitutionally-approved censorship."[69] Prominent social media site Facebook,

which as of 2020 boasts approximately 2.8 billion users, is moderated by a large team of

employees who manually review posts for content that goes against the companies terms and

---

[64] Everett, *supra* note 62 at 118.

[65] 47 U.S.C. §230 (2018).

[66] *See* Chris Fox, *Social Media: How Might it be Regulated?*, BBC NEWS (12 Nov. 2020), https://www.bbc.com/news/technology-54901083 (last visited Apr. 8, 2021).

[67] David Shepardson and Warren Strobel, *U.S. Accuses Russian Spies of 2016 Election Hacking as Summit Looms*, REUTERS (Jul. 13, 2018), https://www.reuters.com/article/us-usa-trump-russia-indictments/u-s-accuses-russian-spies-of-2016-election-hacking-as-summit-looms-idUSKBN1K32DJ (last visited Apr. 10, 2021).

[68] *See Terms of Service*, FACEBOOK, https://www.facebook.com/legal/terms (last visited Apr. 27, 2021) *and Terms of Service*, YouTube, https://www.youtube.com/static?template=terms (last visited Apr. 27, 2021).

[69] Everett, *supra* note 64 at 119-20.

services, with the rest being relegated to algorithms.[70] Facebook, of course, is not the only social media site to attract a wide user base and scorn for its content-regulating practices.

With more than two billion users a day[71] and very little competition - its' closest competitor, video-hosting site Bitchute, attracts little engagement and is notorious for hosting hate speech and anti-Semitic videos[72] - YouTube is the world's leading repository of user-generated videos, covering topics from cooking to gaming to commentary.[73] YouTube is not unique in its struggle to remove copyright-infringing content from its website,[74] but it is notable because its owner, Google, employs a ML model to detect, flag, and automatically delete or remove advertising from user-generated videos it deems to be infringing on copyright.[75] YouTube's ML model, Content ID, and its application to user-generated content provides a harrowing look into how private social media companies can use AI to suppress user speech and expression either out of fear of copyright claims or in order to earn money from large corporations at the expense of its userbase.[76]

YouTube launched Content ID in 2007 and states that it, instead of human employees, handles 98% of copyright issues on the site.[77] To have Content ID scour YouTube for their

---

[70] H. Tankovska, *Number of Monthly Active Facebook Users Worldwide as of 4th Quarter 2020*, Statista (Feb. 2, 2021), https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/.

[71] *YouTube for Press*, YOUTUBE ABOUT, https://www.youtube.com/intl/en-GB/about/press/ (last visited Apr. 27, 2021).

[72] *See* Milo Trujillo, et. al., *What is BitChute? Characterizing the "Free Speech" Alternative to YouTube*, arXiv:2004.01984, (2020).

[73] *Culture and Trends Report*, YOUTUBE CULTURE AND TRENDS, https://www.youtube.com/trends/ (last visited Apr. 26, 2021).

[74] Twitch.tv allows users to stream video game footage to a live, participating audience. Twitch users often complain that the platform's staff is quick to remove any content that appears to be infringing on copyright, even if it was protected by fair use. A particularly egregious example was when the rock band Metallica streamed themselves playing their own music for a live event, only for their livestream to have its audio muted, presumably over issues involving copyright infringement. *About Twitch*, Twitch, https://www.twitch.tv/p/en/about/ (last visited Apr. 26, 2021); Jack Morse, *Twitch Dubbed Over Metallica Stream, and the Internet Thinks That's a Shame*, Mashable (Feb. 20, 2021), https://mashable.com/article/metallica-twitch-livestream-dmca/.

[75] *How Google Fights Piracy*, Google, 1, 24 (Nov. 2018).

[76] *Id*.

[77] *Id*.

illegally-uploaded copyrighted materials, copyright holders (mostly corporations, though

YouTube users can utilize Content ID by applying for the YouTube Partner Program)[78] give

YouTube "reference files" containing audio, visual, and metadata of their protected work, select

what they would like the algorithm to do once it detects that a user-uploaded video matches the

work, and then permits Content ID to scan YouTube for the work.[79] The three actions that

Content ID can automatically take against a copyright-infringing video are: (1) Allow the

copyright holder to earn money off of the video, either by taking a percentage or all of the

creator's advertising revenue[80]; (2) allow the video to remain on the site and monitor its viewing

statistics, or; (3) block the video from YouTube altogether.[81] If YouTube determines that a

copyright claim is valid and removes an infringing user's video from the site, that user receives a

"strike" on her channel. Though a user can complete YouTube's online "Copyright School"

program to remove one of their channel strikes, if a user receives three strikes, their account is

suspended and all videos on said account are removed.[82]

YouTube claimed in 2018 that less than 1% of all copyright claims made against videos

on its site were disputed,[83] and that it would manually review and stop claimants who

misunderstood or flagrantly abused the Content ID system,[84] but many YouTube users have

expressed aggravation with the Content ID and copyright claimant processes as openly ignoring

---

[78] *YouTube Partner Program Overview and Eligibility*, YOUTUBE HELP (updated Nov. 2010),
https://support.google.com/youtube/answer/72851?hl=en.
[79] Google Piracy, *supra* note 75 at 25.
[80] YouTube users can choose to enable advertisements on their videos to receive a portion of the advertising revenue
received whenever their video is viewed. *See How to Earn Money on YouTube*, YOUTUBE HELP,
https://support.google.com/youtube/answer/72857?hl=en (last visited Apr. 26, 2021).
[81] Google Piracy, *supra* note 75 at 24-5.
[82] *Id*. at 29.
[83] *Id.* at 28.
[84] *Id*. at 30.

their right to fair use.[85] YouTube users also state that copyright claimants have a disproportionate amount of power over users; if a user disputes this claim as being false, the *claimant* will review it, often purposefully waiting weeks to do so, opening the system to abuse.[86] Multiple false copyright strikes can lead to the termination of a user's account, which causes the user to lose all of his previously-published videos and advertising revenue, even from non-infringing videos, and potentially a portion of his following.[87] If a copyright holder insists on claiming a user's video when the content in question was protected by fair use, the final recourse Google offers is a form requiring the user's personally identifiable information that said user can then use to file suit against the copyright holder.[88] Many users do not see this as a valid option, as litigation can be expensive and time consuming, and users who deliver controversial opinions or simply value anonymity and do not prefer to have their personal information tied to their account.[89]

YouTube's Content ID system is a prime example of AI, coupled with staff and systems that fail to consider copyright law and fair use on a popular online forum, stifling expression and monetary incentives for its users.[90] Users have gotten creative in circumventing these unfair copyright claims: users Ymfah and the Original Ace created videos teaching viewers to avoid losing revenue on their videos by applying for YouTube's Partner Program, submitting an

---

[85] Musician Gus Johnson describes his experience of his clearly non-copyright infringing video being manually claimed by music corporations who thereafter received the ad revenue from those videos. Gus Johnson, *YouTube's Content Claim System is Out of Control*, YOUTUBE (Dec. 17, 2018), https://www.youtube.com/watch?v=Tqj2csl933Q; *see also* Original Ace, *Abusing YouTube Copyright Claims (Tutorial)*, YOUTUBE (Nov. 24, 2019), https://www.youtube.com/watch?v=Mz14Ul-r63w *and* Adam Neely 2, *Warner Music Claimed My Video for Defending their Copyright in a Lawsuit they Lost the Copyright for*, YOUTUBE (Feb. 6, 2020), https://www.youtube.com/watch?v=KM6X2MEl7R8.

[86] Original Ace, *supra* note 85.

[87] *Id.*

[88] User Ian Corzine states that this form requires a user's full name, email, phone number, and physical address, and often only arrives days or weeks after waiting for a response from the copyright claimant. Ian Corzine, *How to FIGHT False Copyright Strikes!!!*, YOUTUBE (Jun. 25, 2019), https://www.youtube.com/watch?v=3G3f1-x-ZwM.

[89] *Id.*

[90] *See generally* Benjamin Boroughf, *The Next Great YouTube: Improving Content ID to Foster Creativity, Cooperation, and Fair Use Compensation*, 25 ALB. L.J. SCI. & TECH. 95 (2015).

original piece of music for Content ID to track, putting that music in their own video, and then requesting Content ID to flag their own video in order to collect the advertising revenue without having to worry about another individual or corporation falsely flagging it.[91]

While YouTube's mishandling of AI and fair use seems to create a strong case for the public regulation of privately-owned websites, the chilling effect that government regulation would have on users of social media sites cannot be understated. Users who suspect that they are being monitored by government employees will necessarily change their behaviors while online, which will lead to a stifling of new opinions and creative works. This unfortunately leaves little recourse for YouTube's content creators besides demanding that Google adopt a more balanced approach to copyright claim disputes, instead of giving copyright holders the ability to easily make false claims against those who should be protected by fair use. In the interim, YouTube creators seem content to engage in a kind of arms race against Content ID, "outsmarting" it in unique and creative ways.

III. Positive Piracy? Potential Social and Economic Benefits of Piracy

The word "piracy" tends to conjure images of criminals and suspicious thieves, but the act of pirating digital media is not always necessarily immoral. There seem to be few drawbacks to future advanced AI and ML models effectively destroying digital piracy, as copyright holders (and the U.S. economy) would stop losing revenue and the spread of malware via illegally streamed and shared materials would be significantly curbed.[92] Is digital piracy so straightforward an illegal act, however, that it has no benefit at all to society in the U.S. or internationally? Despite its illegality, end users who share information with each other,

---

[91] Ymfah, *How to Break YouTube (Copyright Claim Your Own Video)*, YOUTUBE (Jan. 16, 2020), https://www.youtube.com/watch?v=ieErnZAN5Eo; Original Ace, *supra* note 85.
[92] Alvaro Puig, *Malware from Illegal Video Streaming Apps: What to Know*, FEDERAL TRADE COMMISSION (May 2, 2019), https://www.consumer.ftc.gov/blog/2019/05/malware-illegal-video-streaming-apps-what-know.

copyrighted or otherwise, may benefit their society in three ways: distributing informative materials to demographics which would not normally have access to them, thus creating a more educated society; preserving art such as classical books, films, and video games, particularly when the copyright holders or owners of these materials no longer provide that art or make it unreasonably difficult to obtain, and; encouraging the growth of companies by using piracy to increase a product's publicity and future sales.

A. Distribution of Knowledge

It is not surprising to learn that attending college in the U.S. is expensive - the average price of tuition for both public and private universities is increasing,[93] and though scholarships can alleviate the pressure of such a large price tag, it does not account for the price of housing, food, and required materials for class such as textbooks. Despite predictions that 35% of American jobs would require at least a bachelor's degree in 2020,[94] college enrollment was found to have been declining by 3% at public colleges and 27% at for-profit institutions in 2018.[95] Individuals who are able to receive a college education must often take on debts to afford it; 75% of the U.S.'s massive $1.5 trillion student debt is borrowed by students attending a two or four-year college.[96] The pleas for federal student loan forgiveness grow stronger week by week, especially by those attending university during the COVID-19 pandemic, as the average

---

[93] In 2020-21, the tuition prices for all American colleges – private and public, in-state and out-of-state, two-year and four-year programs – increased by an overall average of about 1.5%. *Trends in College Pricing: Highlights*, COLLEGEBOARD, https://research.collegeboard.org/trends/college-pricing/highlights (last visited Apr. 11, 2021).

[94] Anthony P. Carnevale et al, *Recovery: Job Growth and Education Requirements Through 2020*, GEORGETOWN UNIVERSITY CENTER ON EDUCATION AND THE WORKFORCE 6 (2013).

[95] COLLEGEBOARD, *supra* note 93.

[96] Adam Looney, et. al., *Who Owes All That Student Debt? And Who'd Benefit if it Were Forgiven?*, BROOKINGS (Jan. 28, 2020), https://www.brookings.edu/policy2020/votervital/who-owes-all-that-student-debt-and-whod-benefit-if-it-were-forgiven/.

borrower will have to pay nearly $400 per month to satisfy their loans once the United States ends its pause on payments from federal student loan borrowers.[97]

In the face of massive debt and an uncertain job market, is it any wonder that students and nonstudents alike pirate textbooks to save some amount of money while receiving an education? According to the College Board, an average college student spent more than $1,200 on books and materials alone over the course of his education in 2018, which has only been exacerbated by classes requiring books bundled "online access codes" that expire at the end of the semester, severely diminishing the book's value.[98] An estimated 65% of students skipped buying required texts during at least one point in their college career because they could not afford them.[99] Pirating required textbook materials (either digitally or by scanning and retaining a copy of the textbook's pages) can be the only option for students who strive for high marks and a fulfilling education but are limited by budgetary constraints. Though a college education is not required for every job in the U.S., it has the additional benefit of giving young adults access to academic materials locked behind paywalls and exposes young adults to diverse viewpoints. While education remains expensive for American students, digital eBook piracy can be used to somewhat mitigate the cost and thereby contribute to a more educated society.

There have been several international movements advocating for the ease of access to knowledge by the world's population. The Access to Knowledge ("A2K") movement was created after the October 2004 Geneva declaration on the World Intellectual Property

---

[97] Annie Nova, *Pressure Mounts for Biden to Forgive Student Debt*, CNBC (Apr. 13, 2021), https://www.cnbc.com/2021/04/13/pressure-mounts-for-biden-to-forgive-student-debt-.html
[98] Kathy Kristof, *What's Behind the Soaring Cost of College Textbooks*, CBS NEWS (Jan. 26, 2018), https://www.cbsnews.com/news/whats-behind-the-soaring-cost-of-college-textbooks/.
[99] *Id.*

Organization, where there was a call for a treaty on access to knowledge and technology.[100]

Members of the A2K movement, governments and individuals from various countries, perceive an increasing imbalance between the "knowledge commons" (knowledge that is "owned" by the public) and "privatized knowledge"(knowledge that is controlled by an intellectual property rights holder), caused mainly by Northern governments pushing for broader and stronger intellectual property (IP) protection.[101] The "dramatic" increase in the duration of copyright protection has led to a public domain that "is only half as big...as the copyright regime of 80 years ago."[102] The A2K movement aims to make knowledge more freely available by increasing the availability of textbooks, scientific journals, medicines, and software while decreasing their prices, and promotes free communication over the internet.[103]

The A2K movement also presents a problem that digital piracy, specifically of educational texts and videos, seems poised to solve. While modern end users (in countries that do not censor or heavily restrict the internet for its citizens) are able to access much more information for free than ever before,[104] access to organized, informative research present in academic journals remains limited to university students and those who are able to afford it.[105] Elsevier, the world's largest publisher of academic journals, maintains a monopoly over academic journals, charging universities exorbitant subscription prices for its students to view

---

[100] Becky Hogge and Vera Franz, *The Rise of the Access to Knowledge Movement: An Interview with Vera Franz*, OPEN SOCIETY FOUNDATIONS (Feb. 1, 2011), https://www.opensocietyfoundations.org/voices/rise-access-knowledge-movement-interview-vera-franz.

[101] *Id*.

[102] *Id*.

[103] *Id*.

[104] Wikipedia, though slightly controversial because it permits any user to anonymously edit any article before moderators can review it, provides over 56 million articles on an extensive array of topics to any internet user for free. *Wikipedia*, WIKIPEDIA: THE FREE ENCYCLOPEDIA, https://en.wikipedia.org/wiki/Wikipedia (last visited Apr. 10, 2021).

[105] Brian Resnick and Julia Belluz, *The War to Free Science*, VOX (Jul.10, 2019), https://www.vox.com/the-highlight/2019/6/3/18271538/open-access-elsevier-california-sci-hub-academic-paywalls.

journals and charging academics to submit articles to its journals.[106] Digital piracy can be used to both spread knowledge to the wider public for free, eventually leading to a somewhat more educated general public, and put pressure on monopolist publishers to lower its prices so that it becomes more reasonably affordable for students and the curious-minded citizen alike.

B. Conservation of Art

The American Institute for Conservation describes its mission, art conservation, as "all those actions taken toward the long-term preservation of cultural heritage. . . .includ[ing] examination, documentation, and preventive care."[107] While this certainly refers to ancient, physical pieces of art, the same principle applies to the preservation of modern digital media. Books, video games, movies, television shows, and music all reflect the hopes, concerns, and character of the society and time that produced them. The internet must be used as an archival tool for digital media, as copyright holders who refuse to preserve their creations risk permanently losing them to the degradation[108] or destruction of current short-lived digital platforms and devices.[109] Where copyright holders fail, decentralized digital media piracy is the best archival tool available to citizens of the U.S.; ironically, it was the fear of digital that prevented the construction of robust digital archives by official cultural institutions.[110]

---

[106] Additionally, neither academic contributors nor peer reviewers are typically paid for their work. Resnick, *supra* note 105.

[107] *What is Conservation?*, AMERICAN INSTITUTE FOR CONSERVATION, https://www.culturalheritage.org/about-conservation/what-is-conservation (last visited Apr. 13, 2021).

[108] Floppy disks store data, including software, via magnetic charges on a plastic disk that degrade over time, generally after about thirty years. This is troubling, as some historical software and media only exists on floppy disks. Benji Edwards, *Why History Needs Software Piracy*, TECHNOLOGIZER (Jan. 23, 2012), https://www.technologizer.com/2012/01/23/why-history-needs-software-piracy/.

[109] Abigail De Kosni,,*Piracy is the Future of Culture: Speculating About Media Preservation After Collapse*, 34 THIRD TEXT 1, 3 (2019).

[110] *Id*. at 6.

Video games are becoming more accepted as a unique form of art by scholars and the public alike,[111] but it has proven to be an art that is difficult to preserve in the current copyright landscape. Games that are created for or ported[112] to computers ("PC games") are simple to pirate, as individuals need only upload the games' files from a computer onto a P2P sharing service after circumventing any DRMs included in said game. Pirating video games that were released in the late 1980s, the 1990s, and the early 2000s ("retro" games)[113] or modern games exclusively released on video game consoles[114] is more difficult. If a person wanted to illegally share a console video game over a P2P network, he would need to acquire a legitimate copy of the game, copy the game's ROM[115] files onto his computer, and then share those files via a P2P network. Even if a ROM were already available via P2P file sharing, an individual would need to own the game's corresponding console and download the ROM onto the console's appropriate media storage device to play it. Fortunately for retro game enthusiasts, ROMs are not the only game-related software online; several websites provide game console emulators for free.[116] These emulators, computer software that mimics an actual console's Basic Input/Output system

---

[111] *See generally* Jeroen Bourgonjon et. al., *Perspectives on Video Games as Art*, 19 CLCWeb: Comparative Literature and Culture 1 (2017).

[112] "Porting" a video game generally refers to the process of editing the game's code so that it can be played on hardware (or "platforms") that it was not originally intended for. Normally, *See* Pawel Grabarczyk and Espen Aarseth, *Port or Conversion? An Ontological Framework for Classifying Game Versions*, PROCEEDINGS OF AUTHORS & DIGITAL GAMES RESEARCH ASSOCIATION 1 (2019).

[113] The term was allegedly coined by online video game store RetroGames in 1997, but how old a game must be to qualify as "retro" has never been officially quantified. *RetroGames*, RETROGAMES (Feb. 3, 1999), https://web.archive.org/web/19990922031733/http://www.geocities.com/Tokyo/1424/ (last visited Apr. 29, 2021).

[114] Video game consoles, or "consoles," are specialized computing devices that can be connected to a computer screen to play video games. A video game created for one business' console is typically not able to be played on a different business' console. Consoles are designed to read games off different types of hardware, such as cartridges, CD-ROMs, and Blue-ray discs. *See Games Console*, OXFORD ENGLISH DICTIONARY (3rd ed. 2006) *and Video Game Console*, PCMag, https://www.pcmag.com/encyclopedia/term/video-game-console (last visited Apr. 28, 2021).

[115] "ROM" refers to "Read-only memory," though it is commonly used to refer to the ROM dumps and ROM patches that mimic the game it was copied from. *ROM*, OXFORD ENGLISH DICTIONARY (3rd ed. 2006); James Conley, et. al., *Use of a Game Over: Emulation and the Video Game Industry, A White Paper*, 2 NW. J. TECH. & INTELL. PROP. 1, 5 (2004).

[116] *See generally* Jason Cohen, *The Best Emulators for Playing Retro Games on Modern Devices*, PCMag (Aug. 21, 2020), https://www.pcmag.com/how-to/the-best-emulators-for-playing-retro-games-on-modern-devices (lists sixteen different emulators and links to websites where readers can download them).

(BIOS) and allows the user to play console games on unintended hardware, are capable of

reading ROM files that are stored directly on a computer.[117] An individual with a P2P client, a

decent computer, access to the internet, and a VPN could effectively download hundreds of

ROMs and several emulators to play a litany of copyrighted games for free, and potentially at a

higher quality than the original consoles were capable of rendering.[118] Emulation is a divisive

topic among video game developers, corporations, and enthusiasts, but perfectly exemplifies the

ability of digital pirates to freely download and share games that are no longer supported by their

creators out of a sheer love of the art.

Video game giant Nintendo Co., Ltd. ("Nintendo")[119] is particularly infamous for seeking

out and removing online content that infringes on its copyright while failing to provide

consumers with reasonable means to legally acquire their games and consoles.[120] In 2019, the

company sent DMCA notices to various websites dedicated to providing free download links for

video game ROMs.[121] The host of RomUniverse stated that he "wasn't scared of Nintendo's

legal attack dogs" and would continue to host their games on his website.[122] Nintendo responded

with an immediate lawsuit against RomUniverse, demanding over $100 million in damages for

---

[117] Conley, *supra* note 115 at 4-5.
[118] *Id*. at 6.
[119] Nintendo was founded in 1889 as a business which produced traditional Japanese playing cards, but quickly grew to be a global powerhouse in the video games industry. Today, it is a global company with subsidiaries representing over thirty-seven countries and a net worth of approximately $85 billion in 2020. *History of Nintendo: Where did Nintendo Come From?*, BBC (Jun. 12, 2019), https://www.bbc.co.uk/newsround/48606526; *Nintendo Net Worth 2020*, Revenues and Profits (Apr. 7, 2020), https://revenuesandprofits.com/nintendo-net-worth-2019/.
[120] *See* Thomas Whitehead, *Talking Point: Nintendo and the Industry Needs to Get Serious About Game Preservation*, NINTENDOLIFE (Apr. 4, 2021), https://www.nintendolife.com/news/2021/04/talking_point_nintendo_and_the_industry_needs_to_get_serious_about_game_preservation (last visited Apr. 27, 2021) (discussing Nintendo's failure to preserve its older games and allegedly using an illegally-posted ROM of their game Super Mario Bros. to re-publish that game on their Wii console).
[121] Timothy Geigner, *Nintendo's ROM Site War Continues with Huge Lawsuit Against Site Despite not Sending DMCA Notices*, TECHDIRT (Sep. 16, 2019), https://www.techdirt.com/articles/20190912/10021542977/nintendos-rom-site-war-continues-with-huge-lawsuit-against-site-despite-not-sending-dmca-notices.shtml.
[122] *Id*.

copyright infringement, and fans again berated Nintendo for quickly resorting to legal action instead of preserving their older games.[123] Nintendo has made an official statement describing the use of Nintendo console emulators and game ROMs as illegal, even if the user owned a legally-obtained copy of the game before emulating it,[124] which directly contradicts the U.S.'s 9th Circuit Court of Appeals' ruling that the creation and downloading of emulators is legal.[125]

Nintendo has attempted to make more of its retro game catalogue available for purchase to mixed results. Several of its consoles offered the "Virtual Console," an emulation software that users could subscribe to or purchase (depending on what was offered for a console) to play a small collection of Nintendo's most popular retro games.[126] Nintendo also released the "NES Classic" and "Super NES Classic" in 2016 and 2017, respectively; these miniature consoles were modeled after the original Nintendo NES and Super NES consoles released in 1985 and 1991 in America.[127] Both consoles emulated a small selection of popular games released for their original system without the need to swap game cartridges, but would also not play original NES or SNES cartridges.[128] Fans of Nintendo's retro titles quickly discovered how to use the Windows Operating System to alter the SNES classic to play original NES ROMs that were not included on the console and created a free website that teaches visitors how to do so.[129] For example, Nintendo's 2001 game *Pokémon Stadium 2* ("PS2"), released exclusively for the company's

---

[123] Geigner, *supra* note 121.

[124] Nintendo has since removed this statement from their website, but an archive and several publications referencing it remain. *Legal Information (Copyrights, Emulators, ROMs, etc.)*, NINTENDO CORPORATE, https://web.archive.org/web/20181105062706/https://www.nintendo.com/corp/legal.jsp (last visited Apr. 27, 2021); *see also* Tola Onanuga, *All That's Wrong with Nintendo's Heavy-Handed ROM Crackdown*, Wired UK (Aug. 10, 2010), https://www.wired.co.uk/article/nintendo-roms-emulator-loveroms-loveretro-lawsuit.

[125] *See Sony Computer Entm't, Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 1999).

[126] Whitehead, *supra* note 120.

[127] *See NES Classic Edition*, Nintendo, https://www.nintendo.com/nes-classic/ (last visited Apr. 28, 2021) *and SNES Classic Edition*, Nintendo, https://www.nintendo.com/super-nes-classic/ (last visited Apr. 28, 2021).

[128] *Id*.

[129] *How-to Add More Games to SNES Classic Mini*, SNES Classic Mods, https://snesclassicmods.com/how-to-add-more-games-to-snes-classic-mini/ (last visited Apr. 26, 2021).

Nintendo 64 console, has currently not been re-released or ported to one of Nintendo's more recent consoles.[130] To legally play PS2, one would need to purchase the long-discontinued Nintendo 64 console, which sells for an average of $130 used at a decent quality and can cost upwards of $200 if the console is brand-new or refurbished.[131] Acquiring a legal copy of PS2 will cost just as much, if not more, as various sites dedicated to tracking and compiling the prices for re-sales of the game show that it can sell from $90 to $300 to even $800, depending on the physical quality of the game cartridge.[132]

C. Increasing a Product's Publicity and Future Sales

Some digital piracy researchers have made the surprising claim that digital piracy can benefit the entertainment industry as well as those who pirate content from them. Researchers for Indiana University have found that when digital goods are sold to customers via a retailer, a "moderate amount of piracy" can enhance consumer welfare while increasing the profits of the product's manufacturer and retailer.[133] The researchers described digital piracy as a kind of "invisible competition" which increases if manufacturers, such as HBO, raise prices too high.[134] HBO also conceded that it benefitted from pirates who would illegally acquire episodes of its television show, *Game of Thrones*, and speak with others about it, generating "additional buzz and consumer interest."[135]

---

[130] *Pokémon Stadium 2*, THE OFFICIAL POKÉMON WEBSITE, https://www.pokemon.com/us/pokemon-video-games/pokemon-stadium-2/ (last visited Apr. 28, 2021).
[131] Bobby Anhalt, *How Much is a Nintendo 64 Worth in 2021?*, RETRO GAME BUYER (Jan. 20, 2021), https://retrogamebuyer.com/how-much-is-a-nintendo-64-worth/#open.
[132] *Pokémon Stadium 2*, Game Value Now, https://gamevaluenow.com/nintendo-64/Pokemon-Stadium-2?gameid=195 (last visited Apr. 27, 2021); *Pokémon Stadium 2*, PRICE CHARTING, https://www.pricecharting.com/game/pal-nintendo-64/pokemon-stadium-2 (last visited Apr. 27, 2021).
[133] Karl Bode, *Online Piracy Can be Good for Business, Researchers Find*, VICE (Jan. 28, 2019), https://www.vice.com/en/article/vbwedx/online-piracy-can-be-good-for-business-researchers-find (last visited Apr. 12, 2021).
[134] *Id.*
[135] *Id.*

However, this argument is weak when compared to the previously discussed justifications for digital piracy. Two digital piracy researchers writing for the Harvard Business Review website argue that, while online marketing and word-of-mouth is important to increase revenue for movies, it is not worth backing anti-piracy enforcement measures to achieve positive word-of-mouth.[136] They provide four bases for this conclusion: (1) digital piracy reduces legal sales for a majority of products; (2) digital piracy researchers are coming to a consensus that anti-piracy regulations can reduce piracy consumption and increase sales; (3) piracy benefits rarely outweigh its harms, and; (4) piracy is not the only method by which companies can increase word-of-mouth online.[137] This sentiment is echoed by individuals directly involved in digital media markets. Gabe Newell, one of the main developers of Steam,[138] stated in a 2011 interview that while he did not consider piracy to be a major boon or issue in the "big picture," he understood it as a sign that a content creator was not creating an adequate service value for his customers, and that stated customers generally did not pirate when they felt satisfied by the quality of the service they were being asked to pay for.[139]

IV. "High-Seas" Weaponry – How AI Is and Can be Used to Detect Piracy

    A. Data Packet Inspection and AI

When considering how to prevent digital piracy, the natural first step would be examining the illegal files as they are being shared. When an end-user attempts to access a website or send a

---

[136] Michael D. Smith and Brett Danaher, *The Digital-Piracy Dilemma*, HARVARD BUSINESS REVIEW (Oct. 19, 2020), https://hbr.org/2020/10/the-digital-piracy-dilemma (last visited Apr. 13, 2021).

[137] *Id.*

[138] Steam achieved $4.3 billion in digital video game sales in 2017 (not including purchases of downloadable content or revenue generated by microtransactions), meaning that its developer, Valve Corp., owned at least 18% of the digital PC game sales market and has since continued to grow. Dustin Bailey, *With $4.3 Billion in Sales, 2017 Was Steam's Biggest Year Yet*, PCGAMESN (Mar. 22, 2018), https://www.pcgamesn.com/steam-revenue-2017 (last visited Apr. 11, 2021).

[139] Tom Francis, *We Ask Gabe Newell About Piracy, DRM and Episode Three*, PCGAMER (Sept. 15, 2020), https://www.pcgamer.com/we-ask-gabe-newell-about-piracy-drm-and-episode-three/ (last visited Apr. 11, 2021).

message through the internet, his computer facilitates this by sending a data (or "network")

packet, which is essentially a container holding the raw data to be sent (the "payload") along

with metadata and routing information (e.g. the "header," which states the IP address of origin

and IP address of destination).[140] ISPs could initially only scan and read the header of data

packets, as this is all that is relevant to moving a data packet across a network.[141] The creation of

deep packet inspection (DPI), however, provides ISPs with a somewhat invasive method of

detecting piracy, as it allows ISPs to "scan the payload of [data] packets" as well as the

header.[142] DPI systems additionally come equipped with the capability to "make decisions" on

what to do with a packet or stream of packets based on an expression or pattern in the payload –

essentially, if an ISP determines that certain expressions or patterns in a payload are similar to a

virus or illegal content, it can take action to block its transmission.[143] Though it applies this topic

to IoT services in the medical industry, MediGate argues that DPI provides more certainty in

determining anomalies or the details of devices connected to or communications sent over a

network, while AI requires "cross-referencing with additional data sources" and can falsely flag

packets as suspicious.[144]

　　While DPI seems to offer a convenient means of both detecting and regulating suspicious

content or high amounts of traffic in one package,[145] ISPs must deal with its drawbacks: serious

latency as user packets run through DPI inspection checkpoints, the length of time needed to

decrypt and inspect encrypted traffic, users opting to skip network perimeter protections by using

---

[140] *Data Packet*, TECHOPEDIA, https://www.techopedia.com/definition/6751/data-packet (last visited Apr. 10, 2021).
[141] Ralf Bendrath and Milton Mueller, *The End of the Net as We Know It? Deep Packet Inspection and Internet Governance*, 13 NEW MEDIA & SOC'Y 1142, 1144.
[142] *Id*.
[143] *Id*.
[144] *DPI vs AI: Certainty vs Probability: The Challenges with Security IoT Services with AI-based Solutions*, MEDIGATE, https://medigate.pathfactory.com/c/DPI-vs-AI?x=PzKYM0&xs=87820 (last visited Apr. 11, 2021).
[145] Bendrath, *supra* note 141.

VPNs, and the increased workforce required to handle these issues.[146] Some companies, like Secucloud, offer DPI services that also utilize AI in order to make the process more efficient.[147] However, the significance of the previously-mentioned drawbacks, as well as the fact that users can subvert DPI checkpoints at all, has led some industry experts to conclude that DPI is an inefficient relic of the past compared to what AI can independently achieve when applied to network analytics.[148]

DPI can be a useful tool when combating digital piracy, as it can be used to detect the unauthorized sharing of multiple forms of copyrighted material (i.e. videos, texts, video games) in real time. AI and ML can be applied to incoming network traffic to more quickly and thoroughly classify certain packets as suspicious, though the methods by which an AI would be trained to do so could be costly or difficult to implement for an ISP. The unfortunate truth for DPI as a piracy-fighting tool, however, is that it is hampered by users sending compressed data and rendered completely useless by network encryption, which is available not only through VPNs but is also sometimes included directly in P2P clients.[149] This is especially devastating when considering that the international VPN market is projected to increase to a worth of $31.1 billion in 2021 as more and more adults utilize VPNs to protect their workplace and personal internet privacy.[150]

---

[146] Ericka Chickowski, *Deep Packet Inspection Explained*, AT&T BUSINESS (Oct. 2, 2020), https://cybersecurity.att.com/blogs/security-essentials/what-is-deep-packet-inspection (last visited Apr. 10, 2021).
[147] Secucloud claims to use machine learning to "rapidly classify domains, URLs, Ips and files with great speed" as well as "determine via proprietary AI how deeply a site should be scanned." *Application Control – Deep Packet Inspection*, SECUCLOUD, https://www.secucloud.com/dpi-filter/ (last visited Apr. 11, 2021).
[148] John Edwards, *Applying AI to Network Analytics*, NETWORK COMPUTING (Aug. 2, 2018), https://www.networkcomputing.com/networking/applying-ai-network-analytics (last visited Apr. 10, 2021).
[149] Alexandre M. Mateus and Jon M. Peha, *Dimensions of P2P and Digital Piracy in a University Campus*, TELECOMMUNICATIONS POLICY RESEARCH CONFERENCE 1, 14 (2008).
[150] Ivana Vojinovic, *VPN Statistics for 2021 – Keeping Your Browsing Habits Private*, DATAPROT (Mar. 21, 2021), https://dataprot.net/statistics/vpn-statistics/ (last visited Apr. 10, 2021).

B. Detecting Video Piracy

Digital video piracy is one of the biggest sources of lost revenue in the United States[151] and commonly occurs in two forms: accessing copyrighted materials through illegal Video on Demand (VOD) platforms and illegally streaming copyrighted materials to other users.[152] Illegally streamed video is especially difficult to catch and prevent; pirated streams use the same technologies and protocols as legal streams, which makes it difficult to detect without utilizing DPI, but even with DPI, the multi-tenant hosts, multiple IP addresses, and content delivery methods involved makes it difficult to identify exactly where the illegal stream is originating from.[153] This difficulty in detecting illegally-shared video, coupled with its devastating financial impact made the application of ML to networks in order to detect video piracy almost an inevitability; ML provides "the most direct way" to create a piracy detector.[154]

Researchers Matthew Tooley and Thomas Belford demonstrated a method by which ML could detect pirated livestreams of copyrighted materials merely by noting the sizes and flow usage of pirated content and flagging similar packets instead of directly examining packet payloads. To do so, they compared packet sizes and flow data features of popular video services (i.e. YouTube and Netflix), pirated video traffic, and a "collection of video traffic," namely short traffic including accessing cloud storage, email, and web browsing.[155] When compared to both ordinary internet traffic and legal streaming traffic, pirated streaming traffic had unique characteristics that made it immediately stand out from the other two.[156] The researchers

---

[151] Blackburn, s*upra* note 17.
[152] Matthrew Tooley and Thomas Belford, *Detecting Video Piracy with Machine Learning*, 2019 FALL TECHNICAL FORUM 4 (2019).
[153] *Id*.
[154] *Id*. at 9.
[155] Tooley, *supra* note 152 at 7-8.
[156] The histograms included in Tooley and Belford's paper demonstrated that pirated streams tend to have larger packet sizes and take up a larger percentage of all flows between hosts. *Id*.

compiled this data and hypothesized that the best algorithm to use in determining pirated streams was logistic regression, a "predictive analysis classification algorithm based on the concept of probability."[157] The flow data examined was extracted from "an open source flow data feature extractor called Joy."[158] The ML model was trained with over 50,000 flows of benign traffic (web browsing, Twitch streaming, etc.) and over 90,000 flows of traffic containing pirated streams, then tested on 14,000 flows of benign traffic and approximately 3,600 flows of traffic containing pirated streams.[159] The test results demonstrated that the random forest algorithm,[160] not the logistic regression algorithm, was the best in accurately detecting pirated streams, with a 97% accuracy rating and a false-positive rating of only 0.19%.[161] The researchers further tested their ML model on two residential cable operator broadband networks that provided a sample NetFlow feed; while the ML model detected some pirated streaming on both networks, the researchers noted that the model was trained to find one form of piracy streaming only, and because it was not trained to label certain gaming and music streaming sites as benign, it produced several false-positives.[162]

Despite the issues experienced when training a ML model with limited data, the results of Tooley and Belford's are encouraging for any business which streams copyrighted material for revenue, especially since "for-profit streaming piracy services" cost the U.S. almost $30 billion annually.[163] This issue has only been exacerbated by the COVID-19 pandemic, as both the

---

[157] Tooley, *supra* note 152 at 10.
[158] *Id*. at 10-11.
[159] *Id*. at 12.
[160] A random forest classification is ensemble-based, using multiple decision trees to "compute majority votes in the terminal leaf nodes when making a decision." *Id*. at 11.
[161] *Id*. at 15.
[162] *Id*. at 19-20.
[163] Jordan Vallinsky, *10 Years in Prison for Illegal Streaming? It's in the Covid-19 Relief Bill*, CNN BUSINESS (Dec. 22, 2020), https://www.cnn.com/2020/12/22/tech/illegal-streaming-felony-covid-relief-bill/index.html (last visited Apr. 10, 2021).

United States[164] and the United Kingdom[165] have reported an increase in pirated livestreams of movies. The ML model proposed is especially exciting because it does not depend on scrutinizing data packet payloads, which eases the burden on ISPs and better protects the privacy of end users. Though ML models such as this will require more training – network traffic is so broad and diverse that it will take researchers a long time before ML models stop falsely flagging benign data packets – at least some variation of this model will likely be one of the most efficient means by which copyright holders in conjunction with ISPs can detect illegal streaming of their protected content.

Not every pirated video is livestreamed, however, and copyright holders must combat both illegally pirated streams and illegally downloaded and hosted videos if they hope to prevent a serious loss of revenue. To that end, Dutch cybersecurity company Irdeto has created a new version of its Piracy Control software that can be used to detect illegally streamed and hosted copyrighted videos.[166] Instead of investigating data packets for suspicious patterns, Irdeto has created and trained a convolutional neural network (CNN) to "trawl the internet" for illegally-hosted or streamed content.[167] CNNs consist of several neurons, each of which receive an input through an input layer, the outputs of those that are connected to local regions are determined by the convolutional layer, sent to the pooling layer to "[reduce] the number of parameters within

---

[164] A piracy tracking firm found a 43% increase in traffic to pirated movie-streaming websites during COVID. Sarah Whitten, *More People are Pirating Movies During the Coronavirus Lockdown*, CNBC (Apr. 27, 2020), https://www.cnbc.com/2020/04/27/more-people-are-pirating-movies-during-the-coronavirus-lockdown.html (last visited Apr. 10, 2021).

[165] The UK's intellectual property protection organization found that illegal film streaming had tripled during COVID. *Coronavirus: Film Piracy Streaming Trebles in Lockdown*, BBC NEWS (May 15, 2020), https://www.bbc.com/news/technology-52676254 (last visited Apr. 11, 2021).

[166] Irdeto assists large content provider such as Comcast and Twentieth Century Fox in detecting and removing illegally obtained videos and streams of copyrighted materials. Joe Lemire, *Irdeto Incorporates AI into Anti-Piracy Software to Combat Illegal Streaming*, SPORTTECHIE (Jan. 29, 2018), https://www.sporttechie.com/irdeto-artificial-intelligence-piracy-control-sports-streaming/ (last visited Apr. 11, 2021).

[167] Thomas McMullan, *Running with the Pirate Hunters: How AI is Creating an Online Piracy Arms Race*, ALPHR (Jan. 19, 2018), https://www.alphr.com/artificial-intelligence/1008242/pirate-hunters-ai-sport-illegal-streams-piracy (last visited Apr. 12, 2021).

that activation," and finally the fully-connected layers attempt to utilize the preceding information for class scores to be used for classifications.[168] CNNs are particularly apt at spotting the most minute of details in an image (the "classifications"), which Irdeto utilized by giving its AI "more than three million samples" of possible TV channel logos as training data to compare with videos of illegally streamed or hosted content such as soccer games.[169] If the CNN recognizes a specific broadcaster logo present on an illegal stream or video, it can flag the site hosting said video or stream so that Irdeto and the video's copyright holder can be notified.[170] The challenges involved in training this CNN include it accidentally flagging billboards or signs in a video as a broadcaster logo and the fact that pirated copies of videos and streams can be shown at different resolutions, levels of zoom, or aspect ratios, which can hamper the CNN's ability to detect the broadcaster logo.[171] Rory O'Connor, the senior vice president of cybersecurity services for Irdeto, stated that the company planned on training the CNN to recognize images other than company logos, such as the faces of specific boxing competitors or specific team uniform patterns, in order to increase the effectiveness of its CNN.[172]

Today, Irdeto's website boasts the utilization of both "automation and callable technology" with "human piracy expert oversight" to detect and remove its clients' illegally livestreamed materials as well as recordings on VOD sites, direct download (DDL) sites, social media pages, and even partners with YouTube and Facebook to use their copyright identification systems to flag and remove infringing content.[173] Irdeto has also apparently improved its approach to detecting pirated video and streaming through the creation of Irdeto TraceMark for

---

[168] Keiron O'Shea and Ryan Nash, *An Introduction to Convolutional Neural Networks* 4-5 (Nov. 26, 2015).
[169] McMullan, *supra* note 167.
[170] *Id.*
[171] *Id.*
[172] *Id.*
[173] *Online Piracy Detection*, IRDETO, https://irdeto.com/video-entertainment/online-piracy-detection/ (last visited Apr. 11, 2021).

Distribution, a cloud-based solution that embeds an "invisible, unique watermark" to a video while it is downloaded.[174] Irdeto's AI can then be used to search the internet for any illegally-streamed or uploaded versions of its clients videos by ordering the AI to look for that unique watermark, which can in turn be used to trace leaked content to its source and help its clients begin to create and implement countermeasures.[175] This seems to be a natural and more effective evolution and application of Irdeto's ML model, and though it likely will not deter pirates forever, it will at least make video piracy more difficult for those who choose to steal or illegally stream watermarked videos.

C. How Pirates Fight Back – Circumventing Anti-Piracy Measures

If the increasing competency of AI in detecting piracy is a certainty, then the increasing efforts of pirates in evading that detection is also a certainty. Whether motivated by greed, necessity, or a sense of indignation,[176] digital pirates will not be deterred from finding and exploiting weaknesses in anti-piracy practices and technology.

Both deep packet inspection and digital rights management tools have been ineffective in preventing digital piracy. As discussed previously, DPI can be useful for determining if a packet's payload is illegally transferring copyrighted materials, but it is invasive and can be rendered virtually useless if a pirate utilizes a VPN or proxy service.[177]

DRM tools are not as obviously circumvented, but can still be surprisingly easy to bypass, considering the widespread nature of its use. Anti-piracy measures that rely on tokens,

---

[174] *Irdeto Cleans the Table at VideoTech Innovation Awards 2020*, IRDETO (Dec. 9, 2020) https://irdeto.com/news/irdeto-cleans-the-table-at-videotech-innovation-awards-2020/ (last visited Apr. 11, 2021).
[175] *Id*.
[176] For example, there is a large movement of individuals who vehemently oppose DRM tools, as they can hinder a legitimate customer's enjoyment of or ability to use a product they paid for. Not every member of this movement pirates content containing DRM tools, of course, but many push to eliminate DRM tools entirely. *DRM*, ELECTRONIC FRONTIER FOUNDATION, https://www.eff.org/issues/drm (last visited Apr. 12, 2021).
[177] Mateus, *supra* note 149.

watermarks, or tags tend to be consistently undermined by technologies designed to remove these preventative measures.[178] DRM tools are particularly common on Valve's Steam platform, a popular digital PC game distributor[179] – certain games (particularly those that feature online gameplay with other players) even include "always-on DRM," or DRM tools that require a consumer to remain connected to a server so that a publisher can confirm that the user's copy of the software is authentic.[180] Not only are DRM tools often quickly "cracked" (bypassed),[181] but legitimate customers are often discouraged by DRM's limitations (i.e. authentication and regional settings) and DRM systems can potentially inflate company costs.[182]

Digital video pirates circumvent ML models and CNNs in a much simpler manner: altering the audio and visuals of a video or stream so that the content is still comprehensible to the average viewer while being more difficult to decipher for an ML model. Irdeto's Rory O'Connor was all-too-aware of the possible ways to circumvent the company's anti-piracy CNN, noting that once digital pirates became aware that the AI detected pirated streams and videos based on the presence of a broadcasting logo, they began using editing software to "blank out" the logos or cheekily replace the broadcasting logo with a completely unrelated, different broadcasting logo.[183] O'Connor aptly refers to this process as an "arm's race": when a CNN searches videos for the presence of broadcasting logos, pirates blank them out or crop the video;

---

[178] Hasshi Sudler, *Effectiveness of Anti-Piracy Technology: Finding Appropriate Solutions for Evolving Online Piracy*, 56 BUS. HORIZONS 149, 153 (2013).

[179] Rick Lane, *What's the State of DRM in 2020?*, ROCK PAPER SHOTGUN (May 28, 2020), https://www.rockpapershotgun.com/whats-the-state-of-drm-in-2020 (last visited Apr. 11, 2021).

[180] Logan Westbrook, *Ubisoft DRM Authentication Servers Go Down*, THE ESCAPIST (Mar. 7, 2010), https://v1.escapistmagazine.com/news/view/98927-Ubisoft-DRM-Authentication-Servers-Go-Down (last visited Apr. 11, 2021).

[181] Activision's video game *Crash Bandicoot 4: It's About Time* was released with controversial always-on DRM tools, but said DRM was cracked by pirates merely one day after the game's release. Callum Williams, *Crash Bandicoot 4's Online-Only DRM Was Cracked One Day After it Launched*, GAMERANT (Apr. 5, 2021), https://gamerant.com/crash-bandicoot-4-pc-online-only-drm-cracked-launch/ (last visited Apr. 10, 2021).

[182] Sudler, *supra* note 178.

[183] McMullan, *supra* note 167.

when a company applies watermarks to videos, invisible or not, pirates will take versions of the video form different sources and splice it together to distort them; when distorting playback speed, filters, or adding transparent animations fails to throw off a CNN, pirates will resort to swapping audio to further confuse detection systems.[184]

V. <u>Privacy and Consumer Piracy Surveillance</u>

The appeal of privacy cannot be better summarized than it was in the first United States publication to truly advocate for it: piracy is the "right to be left alone."[185] Alan Westin would later describe the important psychological benefits that human beings can only achieve when they are able to spend time free of surveillance by others: personal autonomy, or the ability to protect one's true beliefs behind a "societal mask" to experiment with views and opinions; emotional release from not having to fret over adhering to social roles; self-evaluation, and; the ability to choose what to say, to whom, and when.[186] The means by which individuals interact with their society has changed after the advent of the internet, but the philosophies behind these writings remain relevant. In fact, data privacy[187] is paramount in allowing individuals to explore their interests online without fear of being tracked (or "watched") by a third party, but the use of AI in scanning networks and websites can provide businesses and ISPs with a tool that can more efficiently invade a user's privacy than it already is.

Despite these and more publications promoting the importance of privacy in the U.S., the United States Constitution never mentions the word "privacy," and modern data privacy laws

---

[184] *Id*.
[185] Samuel Warren and Louis Brandeis, *The Right to Privacy*, 4 HARVARD L. R. 193 (Dec. 15, 1890).
[186] ALAN F. WESTIN, PRIVACY AND FREEDOM 26-7 (2015).
[187] Data privacy concerns the proper protection and handling of personal data, or data that can be used to track the identity of an individual alone or when combined with other identifying information. *What is Data Privacy?*, STORAGE NETWORKING INDUSTRY ASSOCIATION, https://www.snia.org/education/what-is-data-privacy#_ftn1 (last visited Apr. 27, 2021).

differ vastly depending on which state a United States citizen resides in.[188] Several federal

regulations protect against unauthorized access to certain types of electronic communications

conducted by a citizen,[189] but these often fail to address a significant issue in American data

privacy law: most businesses track and sell the information of individuals who visit their

websites, and the burden is often placed on said individuals to "opt-out" of such an

arrangement.[190][191] Conversely, the California Consumer Privacy Act (CCPA),[192] signed into law

in 2018, places the burden more on businesses that are attempting to use or sell an individual's

personal information.

The CCPA offers some of the most robust data privacy protections for an individual in

the United States.[193] The personal information[194] protected by the CCPA ranges from biometric

to educational information.[195] The CCPA explicitly protects information relating to "internet or

other electronic network activity information, including but not limited to browsing history,

search history, and information regarding a consumer's interaction with a website."[196] The

CCPA offers Californian residents four rights to: know about the personal information a

---

[188] No single federal law regulates online privacy – a "patchwork" of federal and state laws apply. *Internet Privacy Laws Revealed – How Your Personal Information is Protected Online*, THOMSON REUTERS, https://legal.thomsonreuters.com/en/insights/articles/how-your-personal-information-is-protected-online (Apr. 10, 2021).

[189] *See generally* Electronic Communications Privacy Act, 18 U.S.C. §§2510-2523 (1986) *and* Computer Fraud and Abuse Act (CFAA) Pub. L. No. 99-474, 100 Stat. 1213 (Oct. 16, 1986).

[190] The opt-out standard is often used by businesses that send unsolicited emails to potential customers. Some lawmakers argue that by including a link in said email that unsubscribes the recipient from the business' mailing list, a business gives recipient adequate control over the information they receive. Donna Gillin, *The Opt-in vs. Opt-out Debate*, INSIGHTS ASSOC. (Feb 28, 2001), https://www.insightsassociation.org/article/opt-vs-opt-out-debate (last visited Apr. 27, 2021).

[191] *Supra* note 79.

[192] California Consumer Privacy Act, 2018 Cal ALS 55.

[193] In fact, the website cited contains a link near the bottom of the page that offers California residents the ability to opt out of their information being sold by Thomson Reuters to third parties. *Supra* note 79.

[194] Information that can reasonably be linked directly or indirectly to a person or household. *Supra* note 83 at §1789.140(o)(1).

[195] *Id*. at §1798.140(o)(1)(A)-(K).

[196] *Id*. at §1798.140(o)(1)(F).

business[197] collects about them, including how it is used and shared; delete personal information collected from them; to opt-out of the sale of their personal information, and; to non-discrimination by a business because the individual exercised his CCPA rights.[198] Any business that uses its website to collect or sell a person's personal information must give Californian users both a "do not sell" link, which allows the user to opt-out of having their information sold, and a "notice at collection" that lists the categories of information collected, what the information is used for, and if it has sold or will sell the data to a third party.[199] The CCPA additionally grants California residents the right to request that businesses delete their personal data or stop selling it to third parties, after which the business cannot request to sell the data for another twelve months.[200] Businesses that violate a CCPA provision are subject to a civil penalty up to $7,500 for each violation, 80% of which goes to the jurisdiction in which the action leading to the penalty was brought and 20% to the Consumer Privacy Fund.[201]

The United States may have a lengthy pro-consumer data privacy statute for one of its states, but the European Union boasts exhaustive, detailed protections for the entirety of its member states and the European Economic Area in the form of the General Data Protection Regulation (GDPR).[202] Styled the "toughest privacy and security law in the world," the official GDPR webpage states its purpose as imposing obligations on any organization that targets or

---

[197] A "business" for CCPA refers to a for-profit business that act in California, have a gross revenue annual revenue of over $25 million, derive 50% or more of their annual revenue from selling California residents' personal information, or buy, sell, or receive the personal information of 50,000 or more California residents. California Consumer Privacy Act (CCPA), State of California Department of Justice, https://www.oag.ca.gov/privacy/ccpa (last visited Apr. 27, 2021).
[198] *Id.*
[199] *Id*.
[200] Businesses can refuse to delete or reveal its use for personal information under limited exceptions. *Id*.
[201] CCPA, s*upra* note 192 at §1789.155(a)-(c).
[202] Council Regulation 2016/679, 2018 O.J. (L. 127) 1.

collects data related to citizens ("data subjects") of the E.U.[203] Like the CCPA, the GDPR

considers personal data to be any information that is related to a data subject and can be used to

directly or indirectly identify them, such as email addresses, names, and biometric data.[204] The

GDPR extends personal information to include beliefs, web cookies, and even pseudonymous

data (if it can be used to easily identify a data subject).[205] Unlike the CCPA, however, the GDPR

boasts extra-territorial jurisdiction over any business that sells products or services to E.U.

citizens, hefty fines to businesses that breach the regulation,[206] and, most importantly, strict

restrictions on the use of personal data.[207] Under the GDPR, data can only be collected and

processed as absolutely necessary for specified purposes, which often prevents businesses from

selling it to third parties,[208] and a data subject must opt-in with unambiguous, specific, informed,

and freely-given consent before a business can process their data.[209]

  Copyright holders have been and still are at an impasse with the privacy of internet users,

and the development of AI has only exasperated this conflict.[210] As copyright protections

expanded to digital media, courts initially failed to recognize that expanding control over

copyrighted material online caused tradeoffs in other areas of consumer protection, particularly

with issues of privacy.[211] The use of AI in tracking network data flows or scanning websites is

one that raises privacy concerns under the GDPR and the CCPA alike – though it may not seem

possible to identify an individual merely by the amount of data they use, it may soon become

---

[203] Ben Wolford, *What is the GDPR, the EU's New Data Protection Law?*, https://gdpr.eu/what-is-gdpr/ (last visited Apr. 27, 2021).
[204] *Id*.
[205] *Id*.
[206] A business that breaches the GDPR can face fines that amount to four percent of its annual gross revenues. *Id*.
[207] *Id*.
[208] *Id*.
[209] As in the CCPA, there are limited exceptions to this rule, such as when the data processing is necessary to save someone's life. *Supra* note 93 at Art. 6.
[210] Sonia K. Katyal, *Privacy vs. Piracy*, 7 YALE J.L. & TECH 223, 335 (2005)
[211] *Id*.

possible, if it is not already. Perhaps an individual's streaming habits, reflected in a data flow, could be linked back to him? Regardless, internet users must already remain vigilant if they wish to protect their privacy online, especially if they are not protected by either the GDPR or CCPA, and it does not bode well to have to worry about yet another aspect of one's internet use being monitored.

VI. <u>Conclusion</u>

The law should not incentivize the use of AI in detecting copyright infringement. Piracy is no longer something that copyright holders can ignore, especially if they hope to make a profit from selling or streaming their copyrighted materials. While businesses in developed economies struggle to curb media piracy, licit media remains a luxury item in "most parts of the world," where piracy rates soar.[212] The number of pirates that copyright holders must contend with will only increase as the rest of the world's population gains access to the internet.[213] The drawbacks of digital media piracy, however, do not outweigh the benefits of utilizing AI to detect and prevent it. While copyright holders and the economy may lose revenue as a result, the potential for creative speech and expression to be suppressed, art to be permanently lost, privacy to be lost, and educational information to remain out of the general public's reach is too devastating to society to warrant the incentivization of AI detecting copyright infringement.

It is easy to fear the implications of advanced ML algorithms being used to detect piracy with increasing accuracy – the already average privacy protections that the U.S. offers it citizens would be diminished, copyright holders could more efficiently remove any variant of its content from video-hosting sites like YouTube, and the sharing of information meant to promote

---

[212] KARAGANIS, *supra* note 3 at 399.
[213] In 2016, the number of global internet users increased from 413 million in 2000 to over 3.4 billion. Max Roser, et. al., *Internet*, OUR WORLD IN DATA (2016), https://ourworldindata.org/internet (last visited Apr. 11, 2021).

knowledge among ordinary people would further be curtailed. The monopoly present in academic journals does little to assuage the fear that piracy is an important tool in driving businesses to adopt reasonable prices and practices, if not the only way for some people to obtain materials necessary for their education.

Admittedly, not every pirate is motivated by a desire to preserve art, combat monopolies, or share knowledge for knowledge's sake. Many are simply drawn to the tempting prospect of obtaining an entertaining game or movie for free. The sheer amount of money that the U.S. economy loses to digital piracy is too much to ignore,[214] especially when the tools needed for piracy are easy to find and can often be obtained for free. Ideally, AI and ML models should be applied to networks and websites with great consideration for user privacy – the network ML model developed by Tooley and Belford[215] appears to be less-invasive than CNNs that directly scan and compare video and streaming content with copyrighted materials.

For now, the "arms race" between pirates and companies utilizing anti-piracy measures continues, and private content-hosting sites like YouTube still struggle with balancing the takedown of illegally copyrighted materials with respecting the fair use rights of its content creators. AI has the incredible potential of changing how the internet users, ISPs, and site owners interact without the need for excessive human oversight, but it is not a cure-all for the social and legal issues that have given rise to the current wave digital piracy. Only after these issues have been addressed can AI truly begin to flourish as a tool that will make the internet an even greater repository of knowledge, an archive for art, and a forum for sharing and exploring new ideas.

---

[214] Blackburn, *supra* note 17; Siwek, *supra* note 18.
[215] Tooley, *supra* note 152.