Radsch, Courtney C.

Working Paper

Weaponizing privacy and copyright law for censorship

CIGI Papers, No. 276

Provided in Cooperation with:
Centre for International Governance Innovation (CIGI), Waterloo, Ontario

Suggested Citation: Radsch, Courtney C. (2023) : Weaponizing privacy and copyright law for censorship, CIGI Papers, No. 276, Centre for International Governance Innovation (CIGI), Waterloo (Ontario)

This Version is available at:
https://hdl.handle.net/10419/299973

Centre for International
Governance Innovation

CIGI Papers No. 276 – May 2023

Weaponizing Privacy and Copyright Law for Censorship

Courtney C. Radsch

Centre for International
Governance Innovation

CIGI Papers No. 276 — May 2023

Weaponizing Privacy and Copyright Law for Censorship

Courtney C. Radsch

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

À propos du CIGI

Le Centre pour l'innovation dans la gouvernance internationale (CIGI) est un groupe de réflexion indépendant et non partisan dont les recherches évaluées par des pairs et les analyses fiables incitent les décideurs à innover. Grâce à son réseau mondial de chercheurs pluridisciplinaires et de partenariats stratégiques, le CIGI offre des solutions politiques adaptées à l'ère numérique dans le seul but d'améliorer la vie des gens du monde entier. Le CIGI, dont le siège se trouve à Waterloo, au Canada, bénéficie du soutien du gouvernement du Canada, du gouvernement de l'Ontario et de son fondateur, Jim Balsillie.

Credits

Managing Director of Digital Economy Robert Fay
Director, Program Management Dianna English
Project Manager Jenny Thiel
Publications Editor Susan Bubak
Publications Editor Lynn Schellenberg
Graphic Designer Brooklynn Schwartz

Printed in Canada on Forest Stewardship Council® certified paper containing 100% post-consumer fibre.

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

Table of Contents

About the Author

Courtney C. Radsch is a CIGI senior fellow whose work centres on the nexus of technology, media and rights. Her research focuses on technology policy, media sustainability and the future of journalism, governance and the geopolitics of technology, and power dynamics in information ecosystems. She is the author of *Cyberactivism and Citizen Journalism in Egypt: Digital Dissidence & Political Change* (Palgrave-Macmillan, 2016), and her articles and commentary have been published in peer-reviewed journals and leading publications around the world. Courtney has testified before Congress and participated as an expert in consultations at the United Nations, European Union, Organization for Security and Co-operation in Europe, and Organisation for Economic Co-operation and Development related to technology policy, freedom of expression, and safety of journalists on- and offline. She was appointed to the Multistakeholder Advisory Group of the Internet Governance Forum and currently serves on the boards of Tech Policy Press, the Dangerous Speech Project and Ranking Digital Rights as well as the Independent Advisory Committee of the Global Internet Forum to Counter Terrorism.

Courtney is involved in the responsible tech and platform accountability movement, and has worked on issues related to content moderation, countering violent extremism online, and digital rights as a founding member of the Christchurch Call Advisory Network and the International Panel on the Information Environment, an expert advisor to the World Economic Forum's Global Coalition for Digital Safety, and serves on the International Science Council's panel of experts on the Public Value of Science.

She is currently a post-doctoral fellow at the Institute for Technology, Law & Policy at the University of California, Los Angeles, where her research focuses on how technology policy impacts media. She is also a fellow at the Center for Democracy and Technology and the Journalism and Media Research Center. She holds a Ph.D. in international relations from American University.

Acronyms and Abbreviations

AI	artificial intelligence
CJEU	Court of Justice of the European Union
CUSMA	Canada-United States-Mexico Agreement
DMCA	Digital Millennium Copyright Act
DSA	Digital Services Act
ECtHR	European Court of Human Rights
GDPR	General Data Protection Regulation
IP	intellectual property
NSD	notice and stay down
NTD	notice and takedown
OSPs	online service providers
PR	public relations
RBSS	Raqaa is Being Slaughtered Silently
SLAPPs	strategic lawsuits against public participation
SNAPPs	strategic notices against public participation
TVEC	terrorist and violent extremist content
UGC	user-generated content
VLOPs	very large online platforms
WSJ	*Wall Street Journal*

Executive Summary

The deep structural inequalities in the information ecosystem are increasingly visible as states vie for their right to govern information flows and technology, with profound ramifications on journalism far beyond their borders. This paper analyzes how a constellation of globally influential technology policies aimed at enhancing individual privacy and intellectual property (IP) rights, addressing content moderation, and mitigating online harms have been wielded by powerful government and business officials as a weapon to censor independent news media and deter investigative reporting. It argues that US and European copyright and privacy laws shape the visibility and viability of news media globally, from their ability to claim fair use and conduct investigative reporting in the public interest to the resources they must deploy to navigate these techno-legal systems. These effects are particularly pronounced when it comes to investigative reporting and news media, particularly in countries where political leaders do not engage with independent media and where state-aligned media often provide the main source of government information. It introduces the concept of moderation mercenaries and the use of strategic notices against public participation (SNAPPs) as helpful concepts for making sense of how these laws are weaponized. Malign actors have weaponized copyright and privacy laws — and the technological infrastructure created by tech platforms to implement them — to claim that journalistic articles infringe on copyright or restrictions on personal data and collection, resulting in critical journalistic coverage being erased from the internet and news archives. Content farms have further drained digital advertising coffers by plagiarizing and monetizing original news reporting. A failure to grapple with this dual-pronged reality risks further undermining media freedom and the viability of public interest news media. This paper shows how poorly designed and implemented techno-legal regimes empower wealthy and powerful individuals to intimidate and coerce the media into removing coverage while becoming essential tools in the arsenal of the public relations (PR) and reputation management firms that conduct influence operations around the world, to the detriment of press freedom and the fight against disinformation and corruption.

Introduction

A constellation of globally influential American and European technology policies aimed at enhancing individual privacy and IP rights online, addressing content moderation and mitigating online harms is being wielded as a weapon to harass independent news media, deter investigative reporting and censor public interest journalism. In the digital age, copyright and privacy laws have given rise to techno-legal regimes that are inscribed into the platforms through their artificial intelligence (AI) systems and content moderation procedures and are embedded by governments in trade agreements with other countries.

This paper focuses on the US Digital Millennium Copyright Act (DMCA) and the EU Directive on Copyright in the Digital Single Market (Copyright Directive), and the so-called right to be forgotten and the EU General Data Protection Regulation (GDPR) because of their influence and importance as techno-legal regimes. They govern significant aspects of our digital communications ecosystem through a combination of legal and technological systems designed to regulate the use and distribution of digital content and protect personal data and IP rights. These governance frameworks have influenced legislation around the world, shaped the technological systems of the most important global platforms through their content moderation and enforcement mechanisms, and been identified by journalists and digital rights groups as enabling censorship. Furthermore, several common features and protocols found in these laws, or what Chris Riley and Susan Ness term “modules” (Riley and Ness 2022), are becoming standard in legislation aimed at addressing digital rights and obligations in the platform era, such as the Digital Services Act (DSA) in the European Union and laws prohibiting illegal and harmful content or data sharing.

News outlets and journalists around the world rely on platforms that host user-generated content (UGC) to reach their audience and, in many cases, to bypass government censorship. But the platforms they rely on are governed by US and European law (as well as domestic laws). One would not expect that national laws from one country could be used to censor media in another country, yet that is what is happening as powerful governments deputize platforms to enforce privacy,

copyright and content-related laws globally. Due to a range of factors addressed in greater detail throughout this paper, US and European copyright and privacy laws shape the visibility and viability of news media globally, from their ability to claim fair use and conduct investigative reporting in the public interest, to the content moderation systems they must contend with just to do their jobs.

Yet criminals, corrupt officials and a burgeoning industry devoted to influence operations, reputation management and information manipulation are weaponizing these techno-legal regimes to censor media with impunity, while content farms that traffic in plagiarized news face few repercussions. Governments do not pursue penalties for misusing copyright laws; platforms have not put in place protective or remedial systems; and many news media lack the legal expertise and resources to fight back against spurious privacy and copyright claims. This is especially true in countries where independent media already struggle to remain viable and confront hostile press freedom conditions, as in much of the Global South.

The censorial impact of the misuse of legal frameworks developed in the Global North on public interest news media in the Global South not only underscores the deep structural inequalities in the global information ecosystem but also undermines efforts to combat disinformation and propaganda, hold those in power to account and improve media sustainability. This paper suggests how to fix this exploitation and better protect public interest news outlets from the deliberate and systematic suppression of information by powerful actors. This means closing legal loopholes and framing the use of these techno-legal regimes as censorial efforts to reduce public participation in the public sphere. The challenge of protecting news outlets from being maliciously targeted also rests on the need to distinguish these outlets from other content producers amid the scale of content produced online and reported through complaint mechanisms, while providing a meaningful remedy that accounts for the temporal dimension of news and the limited resources at their disposal.

This paper focuses on copyright and privacy laws for several reasons. First and foremost, there has been insufficient attention paid to the way that they are used to harass and censor critical reporting and to the costs they exact on publishers. This paper rectifies that and provides a road map for policy makers to consider how to assess the potential impacts of platform-related regulation on news media. It scopes out the impacts on journalism in order to be clearer about the trade-offs involved, decide acceptable error rates, implement corrective remedies and measure proportionality. Since digital copyright and privacy laws have been in place for several years, there is empirical evidence to assess their impacts on journalism and suggest potential solutions to the balancing challenge.

Second, there is strong public interest in protecting IP and privacy rights, but figuring out how to do so amid the seemingly infinite amount of UGC online remains a challenge. However, to comply with international human rights standards on freedom of expression, limitations must be necessary and proportionate, and this paper underscores the disproportionality of current approaches, given the outsized impact on independent news media. While there are many important protections and justifications for these legal frameworks that are addressed elsewhere (Hauser 2008; Koberidze 2015; Ravn 1999; Albrecht 2016; Goddard 2017; Herrle and Hirsh 2019; Buttarelli 2016), the focus here is on how they are abused so that policy makers and tech platforms can understand how to close the loopholes that allow them to be weaponized.

As governments regulate platform responsibility for addressing specific types of content, and platforms develop new policies and practices to implement those requirements, they create new capabilities and expectations that are translated into future legal regulatory frameworks (Lessig 1999; Cohen 2017). The way these laws shape content moderation online is underappreciated, as is their impact on news media around the world. But considering such second-order impacts would enable law makers to design better techno-legal safeguards and prevent the replication of problematic technological mechanisms in new laws. Law makers in the Global North have been slow to address the ways that their laws are being wielded by corrupt public officials and private businessmen,[1] although the European Union's DSA represents what could be a promising exception.

This paper begins with an explanation of why some national laws become the de facto global standard when they are applied transnationally by platforms, are integrated into trade agreements

1 The author has yet to find a case instigated by a businesswoman.

and become influential global templates for other countries. When tech platforms are pushed by powerful governments such as the United States and the European Union to remove or prevent certain types of content on their services, they have responded by adjusting their algorithms to help reduce the visibility and virality of problematic content or even prevent its upload in the first place (Nicholas 2022; Radsch, forthcoming 2023). And platforms adjust their terms of service or community guidelines to allow or disallow categories of content or accounts, imposing various penalties for violating them, and thus affecting the visibility and viability of news media on their platforms. The paper then analyzes enforcement mechanisms, namely, upload filters, notice and takedown (NTD) and notice and stay down (NSD) regimes, hashing, and filtering, which are all part of the broader content moderation system.

Automated algorithmic decision making and enforcement (Perel and Elkin-Koren 2016) have given rise to technical solutions that shape content moderation and platform governance more broadly. Content moderation refers to "the interventions on content or behavior considered unacceptable by a platform intermediary, including the rules they impose, the technology they deploy, and the institutional mechanisms of enforcement" (Gillespie 2018, 2). Most enforcement is enacted algorithmically and, for the moment, with protections against intermediary liability for content moderation decisions (Klonick 2018; Grimmelmann 2015).

When, where and how the detection of illegal or harmful content takes place and the platform responses that such detection triggers are important to understand because they are reappearing in new legislation and can be improved to protect them from being weaponized against news media. The section titled "How the DMCA Automates Copyright Abuse" examines how malign actors have weaponized these mechanisms. Better understanding how various actors abuse the technological capabilities and precedents platforms have created to censor critical and investigative reporting can suggest what safeguards are needed to protect the dual public interest of ensuring the free flow of public interest journalism and preserving IP and privacy rights.

Censorship is not only about silencing or removing journalistic content (which should be protected in the public interest), but also about the harassment of journalists and news media through the abuse of the legal system. A common type of abuse is strategic lawsuits against public participation (SLAPPs), which refer to abusive civil lawsuits filed by powerful actors that are aimed at silencing public criticism on- and offline and deterring reporting about issues of public interest (Snow 2009). The manipulation of copyright and privacy regimes by powerful actors is similar in its abuse of legal regimes for censorial objectives. A failure to grapple with this dual-pronged reality risks further undermining media freedom and the viability of public interest news media (Simon 2023; Krishnamurthy et al. 2021; Radsch, forthcoming 2023).

Platform content moderation systems and enforcement mechanisms end up censoring reporting on newsworthy issues and have even been used to try to shut down investigative journalism. This creates additional burdens on news outlets that have limited resources and are already under strain, often struggling to evade censorship and repression, including media organizations that receive donor funding from the United States and Europe, meaning that one set of their policies is undermining another. Censorship is worst in countries that limit press freedom and restrict access to the airwaves, where social media and online platforms provide critical lifelines for independent media, and in those that can access private sector moderation mercenaries. These effects are particularly pronounced when it comes to investigative reporting and independent digital news media, and particularly in countries where political leaders do not engage with independent media and where state-aligned media often provide the main source of government information.

The following sections outline the threefold problem of the translation of legislation into technological systems, the lack of consistent and clear enforcement, and the deliberate misuse of techno-legal regimes by malign actors. By failing to remedy these problems, both states and the business sector are abdicating their responsibilities under the UN Framework and Guiding Principles on Business and Human Rights. These require that states protect against human rights abuses and provide a framework for holding companies accountable when their policies and practices restrict human rights, such as the ability to receive and impart information and protection for press freedom.

Platforms and Extraterritorial Legal Regimes

Google, with 5.6 billion searches per day and 500 hours of video uploaded to YouTube every minute, and Meta, with its 2.93 billion Facebook and 1.4 billion Instagram users, are among the top platforms for UGC, including news (Statista 2021; Skai 2019; Kemp 2022; Meta 2022b). News media are particularly reliant on these very large online platforms (VLOPs), as the European Union terms platforms with an active user base of more than 10 percent of the European population, since the vast majority of public interest media in developing countries and those with limited press freedom rely on Facebook, Google Search and YouTube to reach their audiences (Radsch, forthcoming 2023; Sembra Media 2021; Newman et al. 2022).[2] Seven of the top 10 global tech firms are American while the others are Chinese (Ponciano 2022).

Since most of these "big tech" companies are headquartered in the United States, they are de jure subject to US law and jurisdiction, and therefore their interpretation of their US legal obligations often amounts to de facto policies for the rest of the world that relies on their platforms. Legal obligations to comply with US digital copyright law, as well as protections from intermediary liability for platforms that host or moderate UGC, are explicitly exported through bilateral and multilateral treaties and trade agreements as well. The 2020 Canada-United States-Mexico Agreement (CUSMA), for example, includes provisions on copyright protection and intermediary liability that are closely modelled on the DMCA (Bagley 2020) and section 230 of the Communications Decency Act of 1996 (Krishnamurthy and Fjeld 2020), respectively, making them the norm for all North America. In 2020, Mexico passed a new copyright law based on the US system, which was criticized as a threat to the wide freedom of expression guaranteed in the Mexican constitution (Doctorow 2020).

Meanwhile, the European Union compels compliance by big tech firms because its policies govern access to the 30-member European Economic Area and are bolstered by a series of legal rulings as well as by the threat of mandatory legislation that hangs over voluntary efforts. The European Union's right to be forgotten and the GDPR have become globally influential and compelled enforcement beyond EU borders by creating standards that multinational businesses adopt in order to do business there and so end up creating systems that are deployed globally. These laws have also created a legislative template, especially for countries with linguistic or historic connections to Europe (Petrova 2019; Bradford 2020), with the EU Copyright Directive and the DSA likely to do the same. This so-called Brussels Effect has been well documented (Bradford 2020; Christakis 2020; Gunst and De Ville 2021) and has helped make European privacy protections a de facto global standard. A version of this privacy right has since spread to Argentina, Colombia, India, South Korea and elsewhere, and journalistic content has increasingly been seen as a legitimate target for erasure efforts (Docksey 2022). These privacy rights have been used by state and private actors to censor journalism, from forcing the closure of news media outlets and prompting news sites to cut off access to their articles, to fuelling the rise of moderation mercenaries and influence operations aimed at censoring public interest.

The question of whether technology policies *should* apply to a specific national or regional jurisdiction or be applied globally is moot when platforms implement technical solutions that affect users globally. The way platforms interpret their legal responsibilities has technological ramifications, as discussed in the next section, and therefore a substantial impact on the viability of public interest news media.

As regulators seek to address how platforms should deal with the vast amount of content created online, they are coalescing around a series of modules, "discrete mechanisms, protocols, and codes" (Riley and Ness 2022), such as notification requirements and responsiveness requirements that impose explicit or presumptive filtering obligations. Legal requirements compel platforms to figure out how to implement their obligations through their technology using AI systems. Understanding how these modules shape content moderation and trust and safety responses by platforms is therefore critical. VLOPs have the resources, and increasingly the mandate, to build the systems to remove content more quickly, although not

2 TikTok use by these same newsrooms remains relatively low.

necessarily more accurately, to address copyright and privacy complaints (Volokh 2022). Furthermore, the technical solutions created to address platform obligations create new capabilities and expectations that can give rise to path dependency.

Global Copyright Law: The DMCA and the EU Copyright Directive

When the US DMCA was first introduced in 1998, the intention was to protect copyright holders from having their work stolen or reproduced online without permission. It extended US copyright law to the internet through US tech platforms, which are subject to its jurisdiction worldwide, and thus de facto to much of the rest of the world. As such, it was one of the first transnational laws of the digital age and shaped content moderation policies that affect users around the world. It amounts to a global copyright regime that laid the groundwork for algorithmic governance and features that have now become standard in other laws.

Furthermore, section 512 of the DMCA grants statutory civil immunity to online service providers (OSPs) for copyright infringement by users but requires compliance with an NTD system and the expeditious removal or disabling of access to the infringing materials.[3] Companies only receive safe harbour from monetary liability for copyright infringement claims if they remove or disable access to the infringing material. To benefit from this protection given the amount of UGC uploaded to and shared on their platforms, tech companies instituted automated NTD regimes that have become a template for content moderation legislation around the world, as discussed further throughout this paper. These systems are largely implemented algorithmically, meaning platforms apply algorithms and machine learning "to perform qualitative determinations, including the discretion-based assessments of copyright infringement and fair use" (Perel and Elkin-Koren 2016, 477).

Unlike the DMCA, the EU Copyright Directive, which is in the process of being transposed into national law by member states, exposes for-profit platforms that host UGC to liability for allowing infringing content claimed by a rights holder to remain on their platform.[4] Yet the directive imposes no penalties on fraudulent or repeat bad faith claimants. This makes it likely that platforms will rely on the use of hash databases and algorithmic screening systems, including upload filters, and are therefore likely to err toward over-removal for both technical and legal reasons.

The scale of takedown requests has increased exponentially over the years, flooding platforms with copyright claims that give large platforms few alternatives to algorithmic automation (despite assertions on Google's website that copyright claims are "carefully reviewed") (Fuller, Grind and Palazzolo 2020; Tewari 2021). Google, for example, reportedly had about 100 reviewers dealing with one million requests per day (Fuller, Grind and Palazzolo 2020). The DMCA (as well as the more recent EU Copyright Directive) allows copyright owners to make an infringement claim to the OSP, triggering the platform to render the content unavailable or remove it unless a counter notice is received, and even then, this often does not result in protection from inaccurate removals. Section 512 ostensibly includes the ability to contest copyright infringement claims through a counter notice and carries penalties of perjury for both notices and counter notices, although in practice these have provided little protection or recourse to most news media organizations (Radsch, forthcoming 2023). The ways this NTD approach is abused and misused are illustrative of the trade-offs policy makers and platforms must grapple with when balancing the need to enforce copyright claims at scale with the risks posed by over-compliance/removal (Bar-Ziv and Elkin-Koren 2018). Yet international human rights and business standards require companies to adopt only necessary and proportionate restrictions on speech and to ensure that remedy is available to those whose rights are affected.

Furthermore, although the DMCA was created to crack down on digital piracy, the types of news outlets targeted by weaponized notices have not

3 *Digital Millennium Copyright Act,* 17 USC § 512 (1998) [DMCA].

4 *Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC,* PE/51/2019/REV/1, OJ L 130 [*EU Copyright Directive*], online: <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.

been able to use it to combat the content farms that plagiarize and recycle their content, further siphoning off what little revenue is available from digital advertising and undermining brand integrity, as discussed further below.

The digital copyright system globally has had little consideration for how approaches developed in the Global North operate at scale and in repressive contexts, much less how they impact independent journalism. Designed primarily with the creative industries in mind, the DMCA has proven to be a blunt tool for enforcing copyright that is increasingly weaponized by state-aligned actors and wealthy businessmen seeking to impede investigative reporting, silence critical commentary and retaliate against independent media in countries with all types of political systems. Journalists could ally with artists and creators who are dissatisfied with the current system because the burden of policing infringing material is placed on users, not platforms (Preston 2020; Henley 2020).

The Right to Be Forgotten and the GDPR

Nearly a decade ago, the European Court of Justice established a right to be forgotten in the digital age in recognition that personal information may become outdated, irrelevant or inaccurate over time, and that individuals should have the right to request its removal from search engines. It acknowledged that the widespread availability and accessibility of personal information online can have significant consequences, such as affecting employment opportunities, financial prospects and social relationships, and wanted to provide individuals with a way to protect their privacy and reputation. A few years later, the new right was strengthened through the European Union's GDPR, which codified comprehensive data protection across EU member states. The GDPR regulates how companies protect personal data and gives control to individuals over how their data is used. Both approaches sought to ensure proportionality by balancing the right to privacy with the public interest and press freedom through exceptions for journalism. But because EU law allows individuals to request that "data controllers," such as internet search engines, remove personal data that is "inadequate, irrelevant, or no longer relevant,"[5] in practice both frameworks have enabled those who wish to censor unfavourable coverage to do so, underscoring the need to better understand how and why such abuse occurs.

The 2014 *Google Spain v Costeja González* case that established the right to be forgotten[6] acknowledged that "even initially lawful processing of accurate data may, in the course of time, become incompatible with the directive where those data are no longer necessary in the light of the purposes for which they were collected or processed."[7] The court specifically exempted journalistic coverage in recognition that there are different interests at play for publishers versus search engines.

However, this important exception has been interpreted differently by national courts and the European Court of Human Rights (ECtHR), contravening the guidance that news content should be exempt. The ECtHR ruled that the right to be forgotten could be expanded to media archives when it upheld a ruling requiring the newspaper *Le Soir* to anonymize an article by removing the name of the subject of the article.[8] Courts in Spain and Germany imposed obligations directly on news publishers, although in these cases they focused on delisting and left the archives untouched.[9]

In both cases, courts required news publishers to use technological measures to make specific

5 *Google Spain SL, Google Inc. v Agencia Española de Protección de Datos (AEPD), Mario Costeja González*, [2014], Case C-131/12 [*Google Spain SL*], online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62012CJ0131>.

6 The right to be forgotten is also referred to as the right to delisting or de-indexing or a right to erasure, although there are different interpretations of these rights (Guadamuz 2017).

7 *Google Spain SL, supra* note 5.

8 *Hurbain v Belgium,* No 57292/16, [2021] ECHR.

9 See Spanish Supreme Tribunal, Civil Chamber, Judgment 545/2015, *B and A v Ediciones El Pais*, S.L, (15 October 2015), online: <https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2018/06/Spain-RTBF-2016-2096STC.pdf>; Headnotes to the Order of the First Senate of 6 November 2019, 1 BvR 16/13 (Right to Be Forgotten), online: <www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2019/11/rs20191106_1bvr001613en.html>; Headnotes to the Order of the First Senate of 6 November 2019, 1 BvR 276/17 (Right to Be Forgotten), online: <www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2019/11/rs20191106_1bvr027617en.html>. These are voluntary measures by news media organizations that are exploring how to address the very real privacy and socio-economic implications of networked permanency, which is quite different from government-mandated removal requirements.

content in newspapers and online publishers inaccessible to the public, even when it concerns public figures. Germany's constitutional court not only rejected arguments about journalistic privilege but even held that publishers of the original information bear an even greater responsibility as the originator of the information (Van Quathem and Shepherd 2019). Requiring media outlets, rather than search engines, to de-index their coverage or make archived content inaccessible goes beyond the original intent of the right to be forgotten and is incompatible with international human rights standards.[10] It also opens the door to even more nefarious press censorship.

As Google's senior privacy counsel phrased it, the right to be forgotten was a "landmark ruling" that immediately prompted action by US-based big tech companies (Carson 2015). Platforms responded by creating new mechanisms to accept requests for removal of an individual's name from their search engines and for removal of their personal data from their services and began acting on such content. Forms allowed individuals to request the removal of their personal information from search results, which in turn spawned bulk removal services and further fuelled reputation management and PR firms, as discussed further below.

Information of public interest was supposed to be exempt from removal by search engines. But such an assessment cannot be made algorithmically and thus is inherently prone to error and overreach (Keller 2018).[11] And how the public interest is defined shifts over time. News is a snapshot of the public interest at any given time and provides valuable documentation of how this interest shifts while simultaneously creating a record that may not become of public interest until, for example, a person runs for office or an executive is found embezzling.

The GDPR similarly contains an exception "for the processing of personal data carried out solely for journalistic purposes…in order to reconcile the right to the protection of personal data with the rules governing freedom of expression."[12] It also recognized that "the processing of personal data solely for journalistic purposes" should be exempted from certain provisions and "should apply in particular to the processing of personal data in the audio-visual field and in news archives and press libraries."[13] The GDPR has become a global template for privacy laws and the processing of personal data online around the world, while tech platforms have turned compliance into the default for countries beyond the European Union, as tech companies have opted to apply this standard globally (Houser and Voss 2018). But the failure of the European Union's Data Protection Board and other relevant bodies to hold national authorities responsible for abusing this legislation has allowed them to misuse the GDPR with impunity (Manancourt 2022), as outlined below.

Enforcement Mechanisms: Upload Filters, NTDs and Hash Databases

The DMCA, the right to be forgotten and the GDPR all impose content moderation obligations on platforms that host UGC. When and where the detection of content takes place, how and how quickly the assessment is done, the way enforcement takes place, and the access users have in order to contest these decisions or pursue remedy for inaccurate moderation are all decisions that have both technological and policy dimensions. As governments regulate platform responsibility for addressing specific types of content, and as platforms develop new policies and practices to implement those requirements, they create new capabilities and expectations that are translated into future legal regulatory frameworks.

Many countries are coalescing around a set of global standards on how to govern content moderation that includes a mix of automated and algorithmic enforcement mechanisms including NTDs or NSD provisions, filtering obligations, and coordination within and across platforms through the use of hash databases. These features are, de facto, creating a set of global modules (Riley and Ness

10 *Biancardi v Italy*, No 77419/16, [2021] ECHR.

11 Microsoft did include a question on its form about public figures.

12 *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, OJ L 119/1, art 85 [*GDPR*].

13 Ibid, art 85, Rec 153.

2022) for how to implement content moderation obligations. By creating technical capacities, collaborative approaches and practices, OSPs set precedents and path dependencies that shape future regulations, with important implications for the future of news media. As one digital rights organization put it, "if you build it, they will come," warning tech companies that governments will use the tools platforms build for their own purposes (McSherry and Trendacosta 2022).

NTD and Intermediary Liability Regimes

Law makers around the world have pressured platforms to improve their content moderation systems to reduce online harms and improve detection of copyright-infringing content, asking them to do more to take down offending content and keep it off their platforms (Keller 2020; Department for Digital, Culture, Media & Sport et al. 2022). Some of the largest platforms, such as Google and Facebook, created automated systems that allow for the bulk submission of notices. The ease of automating notification has been used by creative industries, reputation management firms and copyright trolls to send massive numbers of notices to OSPs. These platforms, in turn, often deal with such notices through automated processing, algorithmic filtering, and blocking or removal of targeted content.

Notice and action procedures are increasingly embedded in laws governing IP and protecting OSPs from liability for illegal, harmful or objectionable UGC (Van Eecke 2011; Johnson and Castro 2021). NTD is a process where an online host disables access to illegal content on its platform upon receiving a notice from a rights holder or an order from a judicial authority. Section 512(c) of the DMCA is an example of such a system: platforms are offered safe harbour from liability for copyright infringement by users of the platform if, among other requirements, they take down infringing content when notified of that infringement by a rights holder.[14]

NSD is the additional requirement whereby a host, after complying with NTD, ensures that the same infringing content does not become available on their platform in the future, and involves the use of automated content filters similar to YouTube's Content ID system. Article 17 of the European Union's Copyright Directive is an example of an NST requirement: in addition to complying with an NTD system, hosts must make best efforts to prevent the reupload of taken down content in order to escape liability for infringing content uploaded by users.[15]

These obligations are embedded in the systems designed by technology platforms to comply with legal regulatory regimes. The ubiquity of NTD systems reflects the scale of UGC online and allows online intermediaries to algorithmically moderate enormous amounts of content without necessarily having to spend the resources to determine whether that content should be protected, since NTD modules are typically accompanied by some level of intermediary liability protections (Johnson and Castro 2021).

Once a notice is received, it creates "actual awareness" of the illegal or infringing content, meaning that the platform will be inclined to remove it to avoid liability rather than to assess its accuracy or legality. Given that platforms want safe harbour, they may decide, or even be required, to restrict access to the content immediately upon receipt of a notice without first establishing if it is legally valid or legitimate (Keller 2021). This type of safe harbour enables OSPs to host and process unprecedented amounts of content without having to review it in advance of publication and allows them to avoid the "moderator's dilemma" of the early internet in which proactively monitoring content or allowing users to report problematic content increases platform liability.[16] However, it also means that they are not forced to internalize the costs of more accurate or nuanced moderation nor the wider social costs created by these systems, such as the costs to independent media (Radsch 2023b).

US platforms have enjoyed the most robust safe harbour provisions for UGC and their content moderation practices. Section 230 of the Communications Decency Act immunizes online intermediaries from criminal liability for illegal or tortious UGC material while permitting them to engage in traditional publisher functions such as "deciding whether to publish, withdraw, postpone,

14 *DMCA*, *supra* note 3, 17 USC § 512(c)(1)(c).

15 *EU Copyright Directive*, *supra* note 4, art 17(4)(b).

16 *Cubby, Inc. v CompuServe Inc.*, 776 F Supp 135 (SDNY 1991); *Stratton Oakmont, Inc. v Prodigy Services Co.*, 23 Media L Rep 1794 (NY Sup Ct 1995).

or alter content."[17] This powerful protection has been exported in US trade deals, such as CUSMA as previously mentioned, and influences how US firms, such as Google and Facebook, have implemented content moderation practices around the world (Johnson and Castro 2021). The DMCA's statutory civil immunity, coupled with section 230 criminal liability protections, provides safe harbour if OSPs implement measures to address infringing material quickly and without having to adjudicate the merits of the claim. Daphne Keller's observation about the dangerous precedent set by an intermediary liability system giving one user "instantaneous veto power" over another user's expression (Keller 2015) is particularly acute when it comes to independent news media.

The scope and scale of automated notification make it virtually impossible for some tech platforms to effectively review or adjudicate allegations, so the burden ends up being placed on the target to comply by removing the offending content or to contest the notice (although this is often futile). As noted earlier, many media outlets are unaware or unable to file counter notices and thus journalists and others have found them largely unavailable as a form of remedy.

If we accept that safeguarding journalism is an important policy imperative, a more proportionate approach would require that safeguards be put in place to deter repeat infringers who fail to accurately and honestly file copyright or data protection claims, for example. Recognizing the similarities these SNAPP notices share with SLAPPs would provide a helpful framework for policy makers who already understand the chilling and censorial impacts that legal filings can have when deployed to deter public participation or journalistic oversight. What is needed is a more holistic approach that considers how platform governance policies impact the public and a system that explicitly considers the trade-offs between various human rights, IP and privacy rights.

17 *Zeran v America Online, Inc.*, 129 F (3d) 327 (4th Cir 1997).

Hashing and Filtering

The development of algorithmic filtering or screening systems involves training them to apply statistical knowledge to assess and classify the data input. Algorithmic content moderation deploys a set of tools that classify and label media (by medium, live-streamed content, keywords and so forth) and in some cases create digital fingerprints known as hashes that correspond to specific images, videos or audio. Hashing enables the same content to be identified in the future on the platform, or even across platforms. Hash databases are currently used internally by companies to identify and prevent problematic content from reappearing on their platforms, as well as to coordinate removal of harmful content such as child sexual abuse material or terrorist and violent extremist content (TVEC) within and across platforms (Farid 2021; O'Connell 2021; Radsch 2021; Global Internet Forum to Counter Terrorism 2021).

Hashes are used for algorithmic identification and in machine-learning systems to identify and take action on matching content and can be used in filtering systems to prevent upload of certain hashed content in the first place. Hash databases can thus be used to prevent the distribution of illegal and harmful content, and are an increasingly common technical solution for preventing the recirculation of a widening array of illegal or problematic content (Radsch 2020a; Turner Lee, Resnick and Barton 2019; Gorwa, Binns and Katzenbach 2020). However, the opacity of such databases and the inability to reconstruct the corresponding content raises concerns about accuracy, oversight and auditability of such approaches (Radsch 2020a; Farid 2021).

Screening systems that analyze and classify UGC at the point of upload are commonly referred to as pre-upload filters. These can then be deployed to identify content at the point of upload. This is the case with YouTube's Content ID system, a proprietary rights management system that "automatically scans all user uploads for infringement and generates claims on behalf of copyright owners… [solving] the logistical headache of monitoring content for infringement" (DeLisa 2016).

Platforms regularly boast that advances in machine learning have enabled detection of problematic content before it is flagged, seen or even uploaded. For example, most copyright claims

and removal requests on YouTube originate from its automatic detection tools Copyright Match and Content ID. According to its latest copyright transparency report, just one percent of copyright-related removal requests related to Content ID were disputed, and 60 percent of those removals were reinstated (YouTube Team 2021). Facebook claims it identifies and removes 98 percent of TVEC before it is ever seen (Meta 2022a).

Legislators from the European Union to Australia and New Zealand want platforms to prevent such content from being uploaded in the first place.[18] The European Union's DSA, for example, requires platforms to act expeditiously to remove or disable access to illegal content, while the United Kingdom's Online Safety Bill could permit the regulatory authority to impose such filters if a platform fails to appropriately comply.[19]

The EU Copyright Directive and the DSA both exemplify the turn toward algorithmic filters for making rapid, large-scale content moderation interventions. Although pre-upload filtering is not mandatory in either code, the imposition of legal liability on platforms for failing to act expeditiously is likely to give rise to a de facto expansion of the use of this technology by large platforms.

A challenge before the Court of Justice of the European Union (CJEU) on the EU Copyright Directive's compatibility with the Charter of Fundamental Rights of the European Union and freedom of expression, prompted the court to acknowledge the tension between the use of upload filters and freedom of expression,[20] but ultimately it dismissed the suit.[21] The CJEU stressed that article 17 imposes a de facto requirement to carry on *ex ante* review of uploaded content, and the liability regime under the directive (and "a limitation on the exercise of the right to freedom of expression and information of users of those content-sharing services") is justified and proportionate under EU law. To comply with such review, platforms should proactively implement safeguards that dissuade moderation mercenaries from manipulating this regime as they have the DMCA, and develop targeted solutions aimed at protecting news media from the types of offensive information operations that have become so prevalent under the DMCA.

Some civil society organizations and representatives of news publishers remain critical of the European Union's approach. Both the Digital Freedom Fund and the Electronic Frontier Foundation have argued that the CJEU's ruling virtually delegates the responsibility to control the respect of user rights to national governments by failing to establish clear parameters to help platforms assess whether to block content, and by missing the opportunity to evaluate whether automated tools per se are proportionate, as required by international human rights law (Reda 2022; Schmon, Lukás and McSherry 2022).

Filtering modules can be easily leveraged to deter journalistic coverage, as exemplified by the abuse by law enforcement or others who play music protected by copyright explicitly to prevent documentation from being uploaded to the internet. In the years since Rodney King's 1991 beating by police in California was captured on video, the convergence of smartphones and widespread community efforts to document and report on police actions, particularly during contentious moments, has led to the creation of entire sites and social media feeds devoted to documenting these videos (see, for example, Bair, n.d.).

Police in California and elsewhere have deployed the DMCA to restrict the circulation of such videos, using their knowledge of the automated filtering system to block videos of police action from being uploaded to social media by playing popular music intended to trigger copyright filters (Schiffer and Robertson 2021; Sung 2021; Cushing 2022). Law enforcement officers have thus added music to their arsenal of weapons, explicitly playing popular tunes from their squad cars, public address systems or on their phones when being filmed in the line of duty or approached by news media and activists to deter their ability to post their videos online. An Illinois police officer even mentioned in an incident report obtained via a Freedom of Information Act

18 See www.christchurchcall.com/about/christchurch-call-text; *Online Safety Act 2021 (No 76)* (Cth), 2021, online: <www.legislation.gov.au/Details/C2021A00076>; *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (DSA)*, OJ L 277, online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065>.

19 Under section 117, the United Kingdom's communications regulator might impose on a platform found in non-compliance the use of "proactive technology" (defined in section 187).

20 *Republic of Poland v European Parliament and Council of the European Union*, Judgment of the Court (Full Court) of 16 February 2022, C-157/21, online: <https://curia.europa.eu/juris/liste.jsf?num=C-157/21>.

21 Ibid.

request that he was "recently advised" to exercise this "copyright hacking" technique (Gault 2021).

News outlets and media activists have been forced to contend with algorithmic copyright enforcement when their coverage includes ambient music that is protected under copyright. Sennett Devermont, an Instagram live-streamer who covers protests and police interactions for his 300,000 Instagram followers, has had several interactions with police who began playing music in an apparent effort to deter his First Amendment right to film on-duty police officers (Thomas 2021a; 2021b). Unicorn Riot, a non-profit decentralized media outlet, was forced to delete an interview related to the Black Lives Matter protests because social media copyright filters flagged it for violating copyright. The outlet tweeted, "Facebook and YouTube have algorithmically interfered with our media coverage due to ambient copyrighted music. We are forced to delete interview audio overlaying background music" (Unicorn Riot 2020). Thus far, this technique does not appear to have been exported elsewhere, but if it is, platforms are ill-prepared to address it.

Reliance on upload filters increases the likelihood of over-removal given the inability of AI and algorithmic content moderation to identify context effectively and accurately, which has significant implications for journalism. False positives are a common problem of algorithmic copyright enforcement and content moderation. While algorithms may detect the presence of illegal, harmful or copyright-protected content, they cannot assess the larger context or differentiate a copyright violation from fair use. Filtering algorithms are not able to determine public interest. Algorithmic screening systems lack the ability to understand context and nuance, which means that they have trouble distinguishing between content that reports on an online harm, such as terrorism, from material glorifying or promoting terrorism, which is illegal under many laws, as well as prohibited by most platforms' terms of service.

Contextual clues, such as whether the author is a journalist or a verified news outlet, could help protect against erroneous removal, but would require classifier systems to be trained to recognize them or that additional ones be created, which requires not just technology but also politically fraught decisions about how to decide which accounts should receive such labels (Radsch 2020c; 2023b). Greater attention to labelling public service media could provide important contextual signals to content moderation and NTD systems. When paired with automated content moderation or NTD regimes, these systems have proved to be a blunt instrument that imposes high costs to inaccurately flagged journalistic content while alleviating platforms from having to engage in more meaningful due diligence or develop more accurate moderation. As Cory Doctorow writes, "the inability of Content ID to tell fair use from infringement is a feature, not a bug" (Doctorow 2019).

Automating Abuse: How Copyright and Privacy Are Abused for Censorship and Profit

State-affiliated media, governments and officials, as well as content farms and PR firms, regularly leverage copyright and privacy regimes to censor critical content and generate ill-gained revenue. A global multimillion-dollar industry devoted to influence operations, reputation management and information manipulation is weaponizing these techno-legal regimes to censor media, while content farms that traffic in plagiarized news continue to generate revenue despite growing awareness about their malfeasance (Forbidden Stories 2023). The next section addresses the weaponization of the DMCA, the GDPR and the right to be forgotten by state-aligned actors and moderation mercenaries, then discusses the limitations of copyright for addressing illegal use of journalistic materials by content farms. It reviews the specific techniques, such as cloning and back-dating, used by both types of actors, that constitute systematic abuse, arguing that tech platforms must mitigate these abuses to comply with their international human rights commitments.

How the DMCA Automates Copyright Abuse

The DMCA has been weaponized to shut down independent, internet-based media around the world and to restrict reporting on police violence in the United States. Abuse of the system globally has heightened opportunities for censorship, that is, the deliberate and systematic suppression of information by powerful actors, as well as for profiteering by financially motivated actors. A third of surveyed independent news organizations in the Global South or working from exile reported receiving DMCA takedown notices (Radsch, forthcoming 2023). Inconsistent interpretation and automation have left the concept of fair use in a precarious grey area that undermines the sustainability in media systems where state media are dominant or captured, especially where state-affiliated media regularly provide footage of officials and their activities.

When Nicaraguan journalists Miguel Mora and Lucía Pineda, founders of one of the country's few independent broadcast outlets, 100% Noticiás, recounted to the then US vice president Mike Pence how authorities had raided their station the year before, taken over their studios and hauled them off to prison, where they spent 172 days behind bars, they told him how important YouTube had become. They moved their broadcasting operations and archives to YouTube so they could keep reporting on the government's crackdown on protests out of reach of the ruling Ortega family (Committee to Protect Journalists 2019a).[22] Or so they thought.

A few months after their meeting at the White House, their YouTube account was frozen, their archives rendered inaccessible and their critical reporting silenced amid a "brutal crackdown" on the press in Nicaragua (Vílchez 2020). Media companies owned by the Nicaraguan president's family or allies are commonplace and enjoy privileged access to state events and government interviews, meaning independent media must rely on those media for footage of public officials and activities. When a state television station claimed copyright violation for using television footage of the president's speeches and other governmental coverage, the 100% Noticías channel and its archives were rendered inaccessible.[23] Another independent news outlet, Confidencial, was also threatened with copyright complaints made by state-affiliated media, despite the DMCA's fair-use provisions that should have protected them (ibid.).

Like so many independent outlets around the world, these Nicaraguan media were dependent on a US-based technology company to publish and broadcast the news in their hometown. And like so many journalists before and since, they were at the mercy of a law and technical response designed in the United States with little consideration for the unintended impact on independent journalism or press freedom, and without access to remedy in their own country.

The Nigerian American citizen journalism outlet Sahara Reporters is another case in point. Its coverage of corruption, political misconduct and human rights abuses has meant its journalists have faced imprisonment, legal and administrative cyberattacks, surveillance and other threats in retaliation for their reporting.[24] In 2019, amid reporting on theft at Nigeria's Central Bank and pro-democracy protests, Sahara Reporters' founder and several reporters were arrested, its bank accounts frozen and its website targeted in a series of distributed denial of service attacks (Committee to Protect Journalists 2019b). Yet they continued their reporting. They had designed their news operations to mitigate against transnational repression by Nigerian authorities, including by establishing a US presence and hosting the website on secret hidden servers, explained Sahara Reporters' CEO La Keisha Landrum in an interview.[25] Yet even these measures were no match for the power of the DMCA, which succeeded in censoring their critical reporting in a way that other attacks and even imprisonment had not.

22 The author was present at this meeting, which was held at the White House on November 1, 2019.

23 Mora and Pineda reached out to the Committee to Protect Journalists, where the author was working at the time. Together with the author's colleagues Nathalie Southwick and Dánae Vílchez, they were able to raise Mora and Pineda's case directly to Google and YouTube and help them seek a resolution, which required weeks of back-and-forth and provided unique insight into the logistics of fighting a takedown notice (Vílchez 2020).

24 See https://cpj.org/tags/saharareporters/.

25 Interview and communication with the author in 2022.

"When you talk about the impact, we reach about 20 million people a month across our platforms. Our journalism, it reaches out really across the entire globe, and all of a sudden, we went dark and instantly, we could not publish. We couldn't figure out why," said Landrum.[26] She added that Sahara Reporters did not even receive a copy of the notice and had to rely on a constellation of international journalism support groups to regain access.

The DMCA is meant to provide immunity to the platforms as well as remedy to their users, but it routinely fails to do the latter due to legal and linguistic challenges. The notices are legalistic and often include intimidating language and even implicit threats, and many media outlets are unaware that they can file a counter notice or that making a fraudulent DMCA claim is a crime. According to Harvard University's Jessica Fjeld, who worked on a 2020 report about DMCA abuse in Latin America (Cyberlaw Clinic 2020), many of the journalists they spoke to "were so intimidated by the presence of US legalese that they didn't recognize their options to respond."[27] Small media outlets have fewer resources with which to fight these types of attacks and defend themselves, often succumbing to the manipulation for lack of access to legal support. An analysis by the Colombian news outlet La Silla Vacía of Spanish-language news sites targeted by copyright abuse found that smaller or less well-known news sites and blogs were more likely to remove targeted content than larger outlets (Lewin 2020).

DMCA notices from Facebook, Google and other US platforms are sent in English, which makes it difficult for journalists or media outlets in non-English-speaking countries to understand them, according to interviews with several journalists.[28] And sometimes the notices are ignored because of lack of awareness, or they get identified as spam and never show up in an inbox, according to the author's interviews with journalists. Although the UN Guiding Principles' "Protect, Respect and Remedy" framework specifically requires that private companies respect human rights and provide individuals with a remedy to address perceived grievances, the way companies implement the DMCA clearly fails to do so.

La Silla Vacía's investigation found that several of the articles targeted with copyright claims had previously been flagged for removal by a reputation management firm based in Spain, Eliminalia, that had tried unsuccessfully to get them removed under the right to be forgotten. This new form of forum shopping (Keller 2016) is problematic and should be addressed by law makers.

Yet a recent congressional assessment of the DMCA done as part of an effort to reform the globally influential law did not grapple with the press freedom challenges or the unintended impact on news media working in some of the most challenging places in the world (United States Copyright Office 2020). The consultations and review by the United States Copyright Office overwhelmingly conveyed the perspective that platforms are not doing *enough* to protect against infringement.[29] There was no apparent consideration of the thousands of news media whose reporting has been censored by the laws of countries that say they support press freedom and media sustainability. Although bespoke non-governmental organization reports have tried to highlight the problem (Krapiva, Rodríguez and Menjivar 2020), awareness of the threat that the weaponization of the DMCA poses to independent media and the costs it imposes remains inadequate.

26 Ibid.

27 Communication with the author. Republished with permission.

28 Interviews conducted with journalists in Colombia, Indonesia, Myanmar, Nicaragua and Thailand.

29 See comments received by the United States Copyright Office at www.regulations.gov/document/COLC-2015-0013-0001/comment.

How the Right to Be Forgotten and the GDPR Automate Abuse of Privacy Protections

Is the *Forbes* list of richest people a data privacy violation? The wealthy owners of a Hungarian energy company thought so when they sued the local publisher. But rather than throw the lawsuit out on the grounds that journalism is protected under the GDPR, a Hungarian court seemed to agree with the plaintiffs and issued a preliminary injunction against *Forbes*' Hungarian edition, forcing the magazine to recall print issues from the newsstands and remove the list from its website (Committee to Protect Journalists 2020). The same company won another preliminary injunction against a local news outlet even before anything was published after receiving a set of questions from an investigative journalist working on a different story (ibid.). Although journalism is supposed to be exempt from the GDPR, in practice media outlets have been targeted by private and public entities that have turned the data protection law into a weapon to remove news, investigative journalism tools and other protected forms of speech.

Authorities in several European countries, including those with poor press freedom records, have weaponized the GDPR to censor independent investigative reporting and create barriers to collaborative journalism. Data protection authorities in Hungary, Lithuania and Romania have all sought to shut down or muzzle investigative reporting and dismantle the tools they use to conduct investigative reporting (Committee to Protect Journalists 2020; ARTICLE 19 2022a; Mong 2019). For example, Lithuania's data protection authority sought to take down a database used by investigative journalists to analyze public documents and statements under the guise that it violated the GDPR. Hosted on the same platform as the Panama Papers and funded by the European Commission, the database was created by journalists to help analyze relationships and uncover corruption to improve democratic governance. Despite a finding from another public authority that oversees journalism in the country, which found that the database qualifies as journalism, the journalists were compelled to testify at a hearing. The failure to dismiss the case outright underscores the fraught situation for media, and the way that local authorities can manipulate these legal frameworks for nefarious ends.

Spain's La Silla Vacía published an essay outlining their attempts to navigate the requests to remove published articles from people claiming their "right to be forgotten" (Lewin 2019). These stories were not inaccurate or defamatory but rather seen as detrimental to the aggrieved parties, who sought to get the stories deleted entirely, to remove their names or to de-index the story from search engines. The fact that there was an uptick in requests around the time of the 2019 election did little to assuage concerns about efforts to "delete the news" (ibid.).

In Italy, the editor-in-chief of an online newspaper was initially liable under civil law for having kept an article on his newspaper's website and for not de-indexing an article reporting the facts of a criminal case instituted against private individuals.[30] PrimaDaNoi, a local news outlet that generated about US$2,200 in digital advertising per month at its height, found itself beset by removal requests despite trying to direct those efforts toward the search engines that were supposed to be responsible for delisting (Satariano and Bubola 2019). Ultimately, the resources of the tiny outlet were no match for the hundreds of legal demands, scores of lawsuits and increasing number of right-to-be-forgotten requests that they were forced to contend with following the court's decision. Amid mounting legal challenges and facing more than US$50,000 in debt from legal fees and fines, the news site closed after 13 years of operation, depriving the town of Pescara of a local news outlet.

As a result of the way these privacy laws have been interpreted and implemented, articles covering alleged corruption, murder and pedophilia, and describing criminal proceedings against public figures involving serious crimes, have been virtually scrubbed from the internet

30 The Supreme Court of Cassation (Italian Supreme Court) accepted an appeal and ruled in 2020 that "the individual who is the subject of a news story, subject to the limits of its truth, will not be able to have it removed from the archives of an online newspaper by invoking the right to be forgotten." According to the Italian judicial system, the case goes back to the Court of Appeal, which will issue a decision in light of the high court's interpretation. Technically, the case is pending, but even if the newspaper editor wins, he still ended up bankrupt and had to close the paper in 2018 because of the hundreds of notifications he received after the first judgment allowing the removal.

by requiring media outlets themselves to take steps to render their coverage inaccessible to search engines, creating "memory holes" that remove information and knowledge from the public sphere (Goldman and Silbey 2020). And excavating these memory holes is big business.

Moderation Mercenaries: The Reputation Management Industry and Commercialization of Information Operations

The commercialization of information and influence operations has risen exponentially alongside the expansion of global copyright and privacy laws and automation of NTD systems, giving rise to what the author terms "moderation mercenaries" — firms or individuals who sell their manipulation skills to whoever can pay. The automation and aggregation of removal notices (some requests include hundreds or thousands of URLs), coupled with lax enforcement of penalties for fraud or perjury related to counterfeit copyright claims, have allowed these mercenaries to flourish (Tewari 2021). Many of them provide content manipulation services, promising to get coverage removed, which further obfuscates quality independent journalism and contributes to the erosion of trust in the media. Meanwhile, the use of moderation mercenaries by political candidates has become a standard part of electoral campaign repertoires around the world, and the NTD and filtering systems designed to implement legal obligations under copyright and privacy laws are welcome assists.

The GDPR has been a boon to the burgeoning reputation management industry, with state-sponsored harassment campaigns targeting independent media and journalists fuelling at least part of this growth (Radsch 2022). The industry had grown to at least US$68 million by the end of the last decade, although this is very likely a vast understatement (Bradshaw and Howard 2019). Documents uncovered by journalists in Latin America showed a single target could cost upwards of $33,000 for removal of 60 URLs (Lewin 2020).

An investigation by Rest of World based on a trove of documents from a single reputation management firm revealed 17,000 URLs, including media websites and news articles, that were apparently targeted for removal or de-indexing over a four-year period (Guest 2022). The firm reportedly charged thousands of dollars per link, and tens of thousands of dollars for some high-profile clients.[31] Leading news sites in Argentina, Germany, Israel, Mexico and elsewhere in Africa, Latin America and the Middle East were listed alongside names of business people and the politically connected who sought to control information about themselves online. A number of journalistic investigations in Mexico, the Western hemisphere's most deadly country for journalists (Southwick and Martínez de la Serna 2022), "were suddenly deleted with no explanation" in 2018 in what appeared to be a pattern of takedowns linked to such reputation management operations (Guest 2022).

Investigative journalism outlets are a particular target for these types of information operations, particularly those involved in transnational anti-corruption projects. Global collaborative journalism investigations have become more common in the wake of groundbreaking coverage resulting from leaks such as the so-called Panama Papers, a cache of documents that revealed how political and business elites took advantage of the offshore finance industry to hide crime, corruption and wrongdoing.[32] A global collaboration involving journalists from 107 media organizations in 80 countries revealed the scale and scope of malfeasance, leading to resignations (Chittum 2016), imprisonment (Alecci 2018), and the recovery of more than $1.2 billion in fines and back taxes (Dalby 2019). Efforts to erase this reporting through fraudulent copyright claims and erasure requests have significant implications not only for the public interest but also for public coffers. "Story Killers," an investigation by the Forbidden Stories collaborative that was released as this paper was going into production, revealed even more disturbing examples of what it terms the "global disinformation-for-hire industry" that includes suppressing independent

31 See www.peopleperhour.com/freelance-jobs/business/legal-services/reputation-management-firm-help-with-legal-letter-3148590.

32 See https://panamapapers.org/.

journalism, interfering with elections and manipulating targets (Forbidden Stories 2023).

Content Farms and the Limitations of Copyright for Independent Media

On the one hand, the DMCA is weaponized against publishers by those without legitimate copyright claims. On the other hand, news media outlets are often plagiarized by content farms, entities that mass produce low-quality and plagiarized content with large amounts of advertising that aim to manipulate search and recommendation algorithms to maximize advertising revenue by increasing traffic to their sites (Shores 2019). Although the DMCA holds those who issue materially false takedown notices liable for damages, it can be difficult, if not impossible, for affected news outlets to pursue such liability. The costs related to fighting these efforts are prohibitive, and even filing a counter notice requires legal expertise and familiarity with US law. Yet platforms are not incentivized to spend the resources or provide the independent oversight needed to ensure that there is less manipulation, in part because they are immune from liability.

Content farms are designed to monetize content by republishing and repackaging copy from legitimate news publishers. These "financially motivated spammers," as one Facebook official described them (Agranovich 2021), have proliferated around the world, particularly in countries where revenues generated on tech platforms are larger and steadier than other sources of income (Hao 2021). They create websites that feature plagiarized news, then register with the platform's monetization program to enable advertising,[33] which siphons ad dollars from the real publisher. The Global Disinformation Index has documented how news websites lose revenue "to click-bait ad farm sites that spread hyper-sensational, misleading, and sometimes outright false news" because those hyper-polemic stories received higher rates of traffic than reported news stories (Breland 2019; Global Disinformation Index 2019). One investigation found that page clusters run out of Vietnam and Cambodia used fake live videos (taken from a media outlet's YouTube channel and reposted to Facebook as a live video), which can include in-stream ads, to rapidly increase their follower numbers and lure them to join Facebook groups disguised as pro-democracy communities in order to increase monetization of the plagiarized content (Hao 2021).

Although content farms reuse copyright-protected content in violation of copyright protections, they are rarely held accountable. But while the film and music industries found at least some recourse to tamp the piracy that was rampant in the early days of the internet, the news industry and journalists have found little recourse to prevent plagiarism of their work by content farms and malign actors (Krapiva, Rodríguez and Menjivar 2020). The director of Mizzima, a Burmese online media outlet that posts several videos per day and has faced DMCA takedowns, said that each three-to-five-minute video costs around $150 to make, meaning that real journalists are subsidizing the costs for troll farms while struggling to avoid copyright filters from erroneously removing their own content.[34] Despite having more than 15 million followers, the outlet has been unable to monetize its own reporting on Facebook.

Content farms drain the already dry coffers of media struggling for commercial viability in the digital age, undermine their brand and drown them out in a sea of low-quality content while compounding the challenges of misinformation. Facebook is a primary vector for clickbait plagiarism given its popularity and availability — it has nearly three billion users (Meta 2022b) and widespread availability — with as much as 60 percent of engagement with Instant Articles taking place on scraped content (Allen 2019). In Myanmar, for example, six of the top 10 websites with the most Facebook engagement in 2015 were from legitimate media, but by 2018 this number had dropped to zero (Hao 2021). And although the companies have been pressured to remove monetization from sites linked to information operations like these when researchers bring it to their attention, no journalists the author spoke with had filed DMCA counterclaims, and it is

33 These include programs such as Facebook's Instant Articles and Audience Network, IGTV Monetization for Instagram, In-Stream Ads for Live Videos and Google's AdSense.

34 Author's interview with U Soe Myint, Mizzima, June 7, 2022.

not clear whether the pages were actioned for copyright violations or if they faced any enduring repercussions for their plagiarism. Meanwhile, journalism outlets struggle for eyeballs and scraps of digital revenue as they do the costly work of reporting, editing and video production that goes into producing journalism, with many independent outlets in the Global South lacking the legal support to pursue copyright claims against the real abusers.

Attention needs to be paid to the impacts of this important difference in reducing plagiarism of news media by content farms as well as to how the technical compliance systems are implemented and their impacts on news media.

The EU Copyright Directive shares many of the same features of the DMCA, although it is newer and thus it is yet unclear whether it will permit the same level of weaponization by bad actors with bad intentions. However, there is also a significant difference from the DMCA in terms of how it treats intermediary liability. The DMCA's section 512 protects intermediaries from liability for copyright infringement, whereas the EU Copyright Directive's article 17 holds online content-sharing service providers liable for unlicensed content displayed on their platforms by the users, creating a very different incentive structure for content moderation. Whether this will help combat the scourge of content farms remains to be seen.

The economic impact of fighting copyright-related takedowns, coupled with the likelihood that the offending account could be removed from the platform for repeated violations even as other content is plagiarized by content farms, poses an existential threat to the sustainability of affected media outlets. There is no accounting of the costs incurred to comply with an increasingly complex web of laws while defending against their weaponization by moderation mercenaries and malign actors, although the author's interviews with journalists and publishers indicate that they can be substantial. "When you are an independent media outlet, like Sahara Reporters, your resources are very, very limited. And you're running a very lean organization," said Landrum.[35] Although it is difficult to figure out the costs involved in fighting offensive information operations, media managers cite the strain on staff time, budgets and technical infrastructure.

Like many digital native media, Sahara Reporters relies on digital advertising for a significant part of its revenue. The site reaches about 20 million people around the world across all its platforms, but when it went dark for two days without notice amid turmoil in Nigeria, its audience dropped off and revenue plummeted, which threatened the very sustainability of the entire outlet. And when Sahara Reporters did get back up — with the help of the Electronic Frontier Foundation and other global advocacy groups with inroads to tech companies — the damage was in some ways already done in terms of reducing their audience, eating up scarce resources and forcing their journalism offline for a period of time when the news they were covering was most relevant.

As more legal frameworks adopt NTD regimes, SNAPPs are likely to be a recurring problem, unless perpetrators are held accountable for perjury and the incentive structure for platforms is adjusted to encourage more accurate moderation or provide more meaningful remedy to journalists and news organizations that have their content and accounts erroneously removed.

Systematic Abuse: Cloning, Backdating, Copyfraud and Copystrike

Investigations into these moderation mercenaries, as well as analysis of the DMCA notices submitted to the Lumen Database,[36] have been able to identify abusive submitters, although there is no indication that they have faced meaningful sanctions or liability for abusing the law. For example, Qurium reported that the company Eliminalia, registered in the European Union and the United States as well as in Ukraine, was behind several of the bogus notices. Servers linked to the registered director of that company were hosting nearly 300 fake newspapers that

35 Interview with the author, May 20, 2021.

36 A project of the Berkman Klein Center for Internet & Society at Harvard University, the Lumen Database collects and analyzes legal complaints and requests for removal of online materials. See www.lumendatabase.org/.

were used to clone existing websites in an attempt to, first, tamper with search results by de-indexing problematic content and, second, make spurious copyright claims by backdating plagiarized articles or using copyright registry services (Qurium Media Foundation 2021). Similar techniques were used against two anti-corruption investigative media outlets based in Africa, which were targeted by information operations linked to the company that sought to censor their hard-hitting reporting and specific coverage of officials allegedly involved in corruption.[37] In the case of Sahara Reporters, the attackers copied the offending article and reposted it online on a faux news site, backdating the publication date to one day prior to the real article. The perpetrator then sent a takedown request to the site's hosting service, which left no other option than to take down the story in order to restore the website.

Cloning, the digital version of plagiarism, and backdating are among the most common tactics used by moderation mercenaries to manufacture fraudulent copyright claims against news media (Tewari 2022; Fuller, Grind and Palazzolo 2020). Box 1 outlines a set of common tactics used. One recent study found that over a two-and-a-half-year period, nearly 34,000 DMCA notices sent to Google cited today-news.press as the original domain for the plagiarized "fake original" articles targeting more than 550 domain names, most of which appeared to be news related (Tewari 2022). There appeared to be a particular operation targeting Lithuanian, Russian and Ukrainian news sites covering allegations of misconduct, corruption, sexual harassment and the like "against the same set of individuals, making it quite plausible that these notices were all part of a systematic and organized attempt to remove critical news articles" (ibid.). Similarly, many of the Eliminalia-related domains were used in coordinated information operations targeting independent media outlets, including smear campaigns and takedown efforts that leveraged both the DMCA and the GDPR (Qurium Media Foundation 2021; Lewin 2020).

A *Wall Street Journal* (*WSJ*) investigation found hundreds of newsworthy articles that were erroneously de-indexed by Google after receiving fraudulent DMCA notices, including more than a dozen local news items that were spoofed or had their content cloned by faux news sites or by sites masquerading as real news sites (Fuller, Grind and Palazzolo 2020). Based on the *WSJ* findings, the company ended up reinstating thousands of de-indexed links and was able to trace suspicious removals to identify more than 100 abusive senders, although it was unclear what happened to those accounts or if they were ever referred for legal action. YouTube did not respond to a request for comment by the *WSJ* or this author.

Scammers also wield copyright claims as a tool of extortion, specifically on platforms that penalize repeated copyright offences (Maxwell 2019). By making fraudulent copyright claims ("copyfraud"), they can trigger suppression or even removal of journalistic content. YouTube's three strikes policy, for example, can result in account closure after three copyright violations, which can mean losing channels and videos without the right to appeal. Scammers make a bogus claim, then seek monetary compensation to remove their strikes. The technique of trying to trigger, or threatening to trigger, YouTube's automated violations rule is referred to as "copystrike." The prevalence of these tactics is exacerbated by the commercialization of services that automate and bulk-submit false claims that target legitimate content, and then

Box 1: Common Tactics

Plagiarizing: republishing and repackaging content.

Cloning: making and publishing a copy of an existing piece of content.

Live spoofing: republishing plagiarized videos as live videos.

Backdating: changing the date of publication of cloned content to a date prior to the original publication.

Copyfraud: fraudulent claims of copyright, often made by claiming copyright to cloned content.

Copystrike: a copyright violation on YouTube. Three strikes can result in automatic closure of the account.

Scary faux legal notices: fraudulent notices that appear to be actual legal documents.

37 See www.makaangola.org/about; www.theelephant.info.

collects the monetization from the affected content (Lizalek 2021). Like SNAPPs, these false-flag campaigns can take place at scale given the automation of account creation and how easy it is to republish content across platforms. This makes plagiarism at scale not just possible but profitable.

The proliferation of moderation mercenaries and industrialized information operations demands that better safeguards be put in place to defend legitimate copyright holders, rebalance platform incentives and require a more meaningful form of remedy for those who are fraudulently targeted. This is especially true given the deadly threats that some journalists face to bring their reporting to the world (see Box 2). OSPs benefit from strong protections against liability without having to mitigate the risks of misuse or abuse of the systems they designed. The fact that companies that provide services such as plagiarism, content spoofing, takedowns, de-indexing and the like are allowed to operate and profit with impunity in the United States and the European Union (which, meanwhile, spend millions of dollars to support these same independent media) is problematic. So too is the failure to press charges against persistent and pervasive abuses of the copyright system and the filing of automated notices.

The DSA: Considering News Media

The DSA seeks to create a common European framework for content moderation, platform management and transparency, and will have repercussions for media and platforms around the world. Although it has similar tensions with respect to the technical solutions that platforms will adopt to meet their obligations, this groundbreaking legislation seeks to address many of the factors that have plagued independent media outlets head-on through safeguard provisions and impact assessment requirements.

The DSA requires platforms to expeditiously remove or disable access to illegal content to avoid being held liable, and to put in place mechanisms to allow users to notify them of the presence of illegal content, similar to digital copyright laws.[38] However, it also specifies that these notices must be sufficiently precise and substantiated to allow for assessment and action, which may reflect an attempt to fix one of the significant criticisms of the DMCA. Since platforms can be held liable for failing to remove or make that content inaccessible, and because such notices are considered to give rise to actual knowledge or awareness, platforms are likely to adopt automated tools of enforcement. However, they also have an obligation to respond in a timely, diligent, non-arbitrary and objective manner. If enforced, this threshold for compliance could provide much-needed safeguards against abuse by mandating greater transparency, risk assessments and more effective access to remedy. Furthermore, OSPs must communicate their decisions to the notified and should include information about whether automated means of detection were used. This should help media and researchers gain a better understanding of how algorithmic intervention impacts news content. In a step toward greater accountability, platforms must conduct algorithmic risk assessments and adopt risk mitigation measures that are tailored to the specific systemic risks identified, such as "adapting content moderation processes" to facilitate the expeditious removal of, or disabling access to, the content notified, in particular for illegal hate speech.[39] These risk assessments should also include explicit consideration of how these systems impact journalism and independent news outlets. This provision could provide better safeguards against abuse on VLOPs, such as Facebook and Google, since, as this paper has established, there is systemic risk to independent digital media posed by automated algorithmic filtering and NTD systems and their susceptibility to misuse.

The DSA is the first major piece of legislation that has recognized the need to consider and address potentially negative impacts on news media and impose obligations to mitigate them. It suggests safeguards against "unjustified removal" and limitation on error rates, an obligation that applies to all users of a platform and thus also to media organizations. VLOPs also have an obligation to conduct impact assessments to identify any systemic risks that could stem from the functioning and use of their services, especially in relation to

38 See *DSA*, *supra* note 17, art 6(1)(b).

39 Ibid., art 22.

Box 2: Syrian Citizen Journalists Risk Their Lives Only to Be Censored by Algorithms

Unfortunately, journalists and independent media around the world live in fear of these laws and the technological capacities developed to automate their content moderation obligations. Abdel Aziz al-Hamza and his colleagues from the Syrian media collectives Raqaa is Being Slaughtered Silently (RBSS) and Eye on the Homeland risked their lives to get news out of Syria online and through social media; five of his fellow citizen journalists were murdered while trying (Committee to Protect Journalists 2015; Ayoub 2016; Greenslade 2015). RBSS was among scores of news outlets and citizen journalism collectives whose members risked their lives in Syria to upload footage from the ground, providing some of the only reporting from one of the world's most geopolitically significant conflicts after international journalists were effectively barred from the country. Human rights activists established the Syrian Archive to collect and preserve digital documentation of human rights violations in a war that generated more hours of social media content documenting the conflict than hours in the conflict itself.*

But the combination of algorithmic content moderation and increasingly aggressive legislation mandating better moderation ends up misidentifying news media and removing what should be protected journalistic content. This can be seen in how algorithmic identification of terrorist content led to the blocking and removal of hundreds of thousands of videos from the Syrian Archive (content from the Shaam News Network, the Qasioun News Agency, RBSS and the Idlib Media Center), caused them to lose followers and imposed additional costs on outlets that had already paid a heavy price for their reporting (Radsch 2018a; Syrian Archive 2022).**

Note: *According to the Syrian Archive; see https://syrianarchive.org/en/tech-advocacy.
**Some outlets were simultaneously receiving media development assistance from the United States and the European Union, meaning that their resources were going toward dealing with the fallout from an American law and poorly implemented technical tools.

"any actual or foreseeable negative effects for... freedom of expression and information, including the freedom and pluralism of the media."[40] This would seem to take aim at moderation mercenaries and offensive information operations, since one of the examples of such a risk includes intentional manipulation, inauthentic use and automated exploitation of the service, with an actual or foreseeable negative effect.[41] But it is not clear how these will be implemented, or that platforms will devote specific resources to focus on creating these safeguards specifically for news media.

There does seem to be some preliminary recognition that the impact on news media needs to be explicitly addressed, as media exemptions have been put forward in the UK Online Safety Bill and the DSA. But they are contentious and have split civil society and regulators amid concerns over the potential impact that news media exceptions could have in the fight against misinformation. For example, during the DSA negotiations, a minority group of EU parliamentarians tried to introduce a media exemption from the content moderation obligations imposed on platforms by amending a requirement that companies provide information about content moderation policies, procedures and tools (European Digital Media Observatory 2021). The amendment would have restricted the power of platforms to moderate media content by prohibiting them from being able to "remove, disable access to, suspend or otherwise interfere with such content or the related service or suspend or terminate the related account on the basis of the alleged incompatibility of such content with its terms and conditions, unless it is illegal content" (Krack 2021).

France's Ministry of Culture had initially highlighted the need to introduce safeguards for freedom of the press on online platforms, fearing

40 Ibid, art 34(1)(b).

41 Ibid, art 34.

increased takedowns of lawful media content. A news media exemption would require determining which outlets qualify as protected journalism sites and the implementation of a labelling system that would enable algorithms to identify news content and filter it to prevent it from being removed by NTDs. Ultimately, however, the exemption was excluded from the adopted text after a public campaign opposing the exemption (People vs Bigtech 2022; EU Disinfo Lab 2022). But at least legislators specifically considered the impacts on news media and gave due consideration to journalism-specific implications of the law. In the absence of a news media exception, more explicit mitigation mechanisms against erroneous removals are critical to avoid creating yet another censorial mechanism that can be easily manipulated and weaponized. These could include positive requirements to improve identification and labelling of news media and to increase access to meaningful remedy for news outlets targeted by erroneous takedown efforts, or to enhance penalties for failure to mitigate NTD regimes or repeat violators from abusing these systems.

Conclusions

The interest by governments around the world in quickly and efficiently moderating UGC and ensuring that prohibited or protected content is prevented from circulating has only grown over the past several years as the amount of information circulating online has exploded alongside advances in AI and machine learning that have enabled more complex and comprehensive content moderation. Privacy and copyright frameworks that lack safeguards against abuse by moderation mercenaries and allow fake news farms to thrive are in opposition to other efforts to combat disinformation and protect quality information. Requirements that tech platforms identify problematic content, remove it within a very short time frame and prevent it from spreading are increasingly common features of legislation aimed at combatting online harms ranging from terrorism to hate speech to piracy. It is therefore essential that the technology incorporates safeguards to protect public interest news media, even if these simply involve flagging affected content for human review. Closing legal loopholes and reconceptualizing weaponized takedowns as a form of censorship aimed at deterring public interest reporting are also needed.

News Integrity and Trust Indicators as Part of the Solution to Algorithmic Enforcement of Content Moderation

The challenge of protecting news outlets from being maliciously targeted by information operations and moderation mercenaries and caught up in algorithmic content moderation systems rests on the need to distinguish these outlets from other content producers. Amid the scale of content produced and reported through NTD/NSD and complaint mechanisms, and the failure of most platforms to provide a meaningful remedy for news outlets, addressing information integrity not only is critical to ensure the viability of public interest news online but also will likely become a boon to platforms amid the wave of propaganda and disinformation expected to emerge amid the generative AI revolution.

The technical solutions devised thus far cannot effectively identify fair use, public interest, problematic but legal speech, satire, or news coverage, and there are few incentives for companies to invest in better solutions, meaning regulatory intervention may be needed to realign platform priorities.

Better labelling, hashing and other improvements to algorithmic automation by platforms could identify and protect journalistic coverage. Contextual clues, such as whether the author is a journalist or a verified news outlet, not only require classifier systems to be trained to recognize protected speech but also are politically fraught determinations, as illustrated by the controversy over how platforms label state-affiliated media (Radsch 2020c). In 2020, Twitter and Facebook started labelling some state-affiliated accounts, joining YouTube, which had started labelling "state-funded" media outlets in 2018 (Radsch 2018b; 2020b). Each company used a different term and definition, drawn from different sources and with varying levels of transparency. YouTube relied on Wikipedia designations, while Facebook relied on a semi-private group of experts, even delaying its initial timeline for release amid lobbying by some

media organizations seeking to avoid receiving a state-affiliated label (Radsch 2020c; 2020b; 2018b).[42]

In order to implement better technological solutions, each platform would first need to determine how to decide what qualifies as public interest media. To this end, drawing on civil society-led, multistakeholder initiatives that have sought to create frameworks to identify credible or quality news media would be preferable to platforms making this determination unilaterally. A plethora of efforts aimed at identifying quality or trustworthy news media on digital platforms could be a partial solution to algorithmic enforcement by providing machine-readable indicators that could be used to filter automated requests and targeted takedown efforts, at the very least so that a human being could provide oversight and review.

Many of these efforts are non-profit, industry-led self-regulatory initiatives with varying levels of comprehensiveness. Some, such as the NewsGuard and the Trust Project, are primarily US and Europe focused, while others, such as the Journalism Trust Initiative and Ads for News, are more global in scope and explicitly seek to include news media in the Global South. Although many small digital native publications are destined to be left out of these current initiatives because they lack the formalized standards and procedures needed to qualify for inclusion,[43] they nonetheless offer a jumping-off point for trying something innovative and solution oriented. NewsGuard and Trust.txt have provided proof of concept for how to translate integrity indicators into signals the platforms can use, but uptake by platforms and media outlets remains limited.

Another option would be for platforms to defer to the lists of trusted local news media outlets curated for advertisers, such as Ads for News. Deemed a "World Changing Idea" by *Fast Company* (Internews 2021), Ads for News has curated 10,100 trusted local news websites from 53 countries and offers brands and agencies the ability to embed the list for free into their campaign management systems and programmatic advertising platforms so that they can continue to run their ads on outlets that have been deemed "brand safe." Tech platforms could similarly use such curated lists to adapt algorithmic signals of news into their content moderation systems. Ideally, a range of recognized and broadly accepted professional and ethical bodies would identify the parameters for media labels, which affect visibility, monetization and other types of algorithmic intermediation.

SNAPPs

Policy makers in the United States and Europe must address the abuse of their frameworks to censor and intimidate independent journalism around the world, on par with the efforts to combat SLAPPs. The abuse by powerful actors of unmoderated notification modules (NTDs and NSDs), some of which are legally mandated, which are aimed at silencing public criticism on- and offline and at deterring reporting about issues of public interest, are akin to SLAPPs (Snow 2009). The author proposes categorizing these types of censorial efforts as SNAPPs. Reconceptualizing fraudulent copyright infringement notices as SNAPPs – that have the same problematic implications for press freedom as SLAPPs – could help frame the problem and raise awareness about the weaponization of copyright and privacy laws through automated NTDs and algorithmic enforcement.

Politicians, public figures and corporations across the Americas (Business & Human Rights Resource Centre 2022; Vining and Matthews, n.d.), Europe (ARTICLE 19 2022b) and Asia (ARTICLE 19 2021) use SLAPPs to silence independent journalism, reflecting many of the same dynamics as the abuse of copyright and privacy techno-legal regimes. The European Union recognized the dangers of these types of vexatious lawsuits being used against journalists and human rights defenders with a proposed directive and commission recommendation earlier this year,[44] and 32 US states have adopted anti-SLAPP laws (Vining and Matthews, n.d.). Concerns over the use of SLAPPs to censor critical academics, journalists, activists and other civil society actors have been raised by the UN Working Group on Business and Human Rights as well as by the United Nations Educational, Scientific and Cultural Organization (United Nations

42 The author was consulted by all three companies on definitions and which terminology to use and spoke with representatives of media organizations; she also conducted her own assessment of the accuracy and global scope of YouTube's labelling efforts in 2018.

43 The author would like to thank Janine Warner, executive director of Sembra Media, for this insight.

44 *Proposal for a Regulation of the European Parliament and of the Council establishing a common framework for media services in the internal market (European Media Freedom Act) and amending Directive 2010/13/EU*, COM/2022/457 final, online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0457>.

Human Rights Office of the High Commissioner 2022; Soraide 2018). Weaponized NTDs and NSDs are a form of censorship, and both states and tech platforms have a responsibility to protect human rights online by ensuring that restrictions on access to information and press freedom needed to balance the rights of privacy and copyright are necessary and proportionate and to put in place better protection and remedy options.

As filtering and NTD regimes are embedded into transnational legal regulatory requirements and become more commonly used, they pose a growing threat to the sustainability of internet-based independent news outlets. Failing to address this issue could also impact a media outlet's ability to stay online even if it does not rely on social media platforms, as web-hosting providers may refuse to host outlets unless the offending content is removed, which is why the DSA's protective provisions are a welcome improvement. The failure by intermediaries to address the negative externalities that algorithmic enforcement of the DMCA, the GDPR, the right to be forgotten and other legal regulatory measures have on legitimate news content must be rectified voluntarily or through regulatory requirements.

More Transparency and Better Enforcement Needed

Technical solutions do not do enough to protect against malicious and fraudulent claims of copyright, while meaningful civil or criminal penalties for abuse of legal process, needed to provide a deterrent effect, are lacking. Law enforcement must do a better job of enforcing legal provisions that criminalize the filing of knowingly false copyright claims and ensure that penalties are imposed on firms engaging in these types of information operations. If abusive behaviour is not deterred, then moderation mercenaries will simply continue to thrive amid the expansion of filtering and NTD provisions in copyright and other laws. The failure to effectively address the abuse of copyright and privacy laws not only means that independent media face costly content removals, account closures and legal challenges, but also that the infrastructure of information operations is left intact and free to profit.

Tech platforms should also report, in particular, on how news media outlets are affected by these specific content moderation efforts. To do so would also require platforms being able to identify news outlets online, underscoring the utility of improving labelling and classification efforts. While some platforms release transparency reports about copyright and GDPR content moderation, only a few contribute DMCA takedown notices to the Lumen Database, which collects notices requesting removal of allegedly infringing content based on legal grounds such as copyright and privacy. All social media companies, and ideally all OSPs, should join Google in voluntarily forwarding copies of all DMCA notices to the Lumen Database so that there is a central repository for researchers that can be cross-referenced with the transparency reports from platforms. Access to this data enables journalists and researchers to discover important information with implications for human rights and freedom of expression.[45] Law makers should require that companies report takedown requests to a central, independent research database and that they report on this type of content moderation in their transparency reports.

Ensuring the sustainability of journalism and news media should be a central concern of policy makers seeking to shape the information ecosystem, protect their citizens' privacy and uphold IP rights in the digital age. Countries with relatively strong press freedom records and independent, financially viable media are also those that have the greatest influence over transnational technology policies; thus it is incumbent that these countries assess the risks their policies pose in countries with poor press freedom records, or where media sustainability is limited, and seek to mitigate them.

As this paper has shown, the visibility and viability of independent news media sit in the crosshairs of how we regulate copyright, privacy and content moderation; the protections platforms enjoy and the responsibilities they incur as they seek to moderate content on their services; and the trade offs made between different policy goals. Legislation and voluntary codes developed in the United States and the European Union have had an outsized impact on the journalism field, particularly on scrappy investigative journalism outlets and independent digital outlets in countries with poor press freedom records and small advertising markets, by providing tools of repression that contradict broader geopolitical and foreign policy goals. As outlined above, these techno-legal regimes contain modules that have become the building

45 See www.lumendatabase.org/media_mentions/search.

blocks for other techno-legal approaches such as the DSA, making it imperative that we fix the problems before they further proliferate. The failure to revise techno-legal regimes that are weaponized to censor independent reporting not only threatens the sustainability of public interest news media but also detracts from efforts to combat disinformation and improve public accountability. As "wicked problems," there are conflicting, iterative values, and solutions that can turn out worse for other parts of the system, which are all interconnected. Therefore, bringing in the perspective of journalism and news media sustainability is essential.

Works Cited

Agranovich, David. 2021. "1/ A few quick thoughts on reporting that equates clickbait farms with foreign troll farms seeking to manipulate public debate ahead of an election. The pages referenced here, based on our own 2019 research, are financially motivated spammers, not overt influence ops." (Twitter thread). Twitter, September 17, 8:16 p.m. https://twitter.com/DavidAgranovich/status/1439020401946816516.

Albrecht, Jan Philipp. 2016. "How the GDPR Will Change the World." *European Data Protection Law Review* 2 (3): 287–89. https://edpl.lexxion.eu/article/edpl/2016/3/4.

Alecci, Scilla. 2018. "Former Pakistan PM Sharif Sentenced To 10 Years Over Panama Papers." International Consortium of Investigative Journalists, July 6. www.icij.org/investigations/panama-papers/former-pakistan-pm-sharif-sentenced-to-10-years-over-panama-papers/.

Allen, Jeff. 2019. "How Communities Are Exploited On Our Platforms: A Final Look At The 'Troll Farm' Pages." Document Cloud, October 4. https://s3.documentcloud.org/documents/21063547/oct-2019-facebook-troll-farms-report.pdf.

ARTICLE 19. 2021. *The Global Expression Report 2021: The state of freedom of expression around the world.* ARTICLE 19. July. www.article19.org/wp-content/uploads/2021/07/A19-GxR-2021-FINAL.pdf.

———. 2022a. "Lithuania: Stop harassment of the Karštos Pėdos journalist platform." ARTICLE 19, January 27. www.article19.org/resources/lithuania-stop-harassment-of-the-karstos-pedos-journalist-platform/.

———. 2022b. *SLAPPS against journalists across Europe.* Media Freedom Rapid Response. March. www.article19.org/wp-content/uploads/2022/03/A19-SLAPPs-against-journalists-across-Europe-Regional-Report.pdf.

Ayoub, Joey. 2016. "How Syrian Activists in Raqqa Are Resisting ISIS." *Global Voices* (blog), February 16. https://globalvoices.org/2016/02/16/how-syrian-activists-in-raqqa-are-resisting-isis/.

Bagley, Ross. 2020. "USMCA Set To Export U.S. Copyright Law to North American Neighbors." IPWatchdog, January 29. https://ipwatchdog.com/2020/01/29/usmca-set-export-u-s-copyright-law-north-american-neighbors/id=118269/.

Bair, Madeleine. n.d. "Caught on Camera: Police Abuse in the U.S." WITNESS Media Lab. https://lab.witness.org/caught-on-camera-police-abuse-in-the-u-s/.

Bar-Ziv, Sharon and Niva Elkin-Koren. 2018. "Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown." *Connecticut Law Review* 50 (2): 339–85. https://opencommons.uconn.edu/cgi/viewcontent.cgi?article=1395&context=law_review.

Bradford, Anu. 2020. *The Brussels Effect: How the European Union Rules the World*. New York, NY: Oxford University Press.

Bradshaw, Samantha and Philip N. Howard. 2019. "The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation." Working Paper 2019.2. Oxford, UK: Project on Computational Propaganda. https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2019/09/CyberTroop-Report19.pdf.

Breland, Ali. 2019. "Fake News Is Getting a Big Boost From Real Companies." *Mother Jones*, May 9. www.motherjones.com/politics/2019/05/fake-news-advertising-networks/.

Business & Human Rights Resource Centre. 2022. "SLAPPs in Latin America: Strategic lawsuits against public participation in the context of business and human rights." Business & Human Rights Resource Centre, February 9. www.business-humanrights.org/en/from-us/briefings/slapps-in-latin-america/.

Buttarelli, Giovanni. 2016. "The EU GDPR as a clarion call for a new global digital gold standard." *International Data Privacy Law* 6 (2): 77–78. https://doi.org/10.1093/idpl/ipw006.

Carson, Angelique. 2015. "The Responsibility of Operationalizing the Right To Be Forgotten." International Association of Privacy Professionals, March 12. https://iapp.org/news/a/the-responsibility-of-operationalizing-the-right-to-be-forgotten/.

Chittum, Ryan. 2016. "Iceland Prime Minister Tenders Resignation Following Panama Papers Revelations." International Consortium of Investigative Journalists, April 5. www.icij.org/investigations/panama-papers/20160405-iceland-pm-resignation/.

Christakis, Theodore. 2020. "'European Digital Sovereignty': Successfully Navigating Between the 'Brussels Effect' and Europe's Quest for Strategic Autonomy." SSRN. https://doi.org/10.2139/ssrn.3748098.

Cohen, Julie E. 2017. "Law for the Platform Economy." *UC Davis Law Review* 51: 133–204. https://lawreview.law.ucdavis.edu/issues/51/1/symposium/51-1_Cohen.pdf.

Committee to Protect Journalists. 2015. "CPJ calls for investigation into murder of Syrian journalists in Turkey." Committee to Protect Journalists, October 30. https://cpj.org/2015/10/cpj-calls-for-investigation-into-murder-of-syrian/.

———. 2019a. "CPJ raises press freedom concerns in meeting with U.S. Vice President Pence." Committee to Protect Journalists, November 18. https://cpj.org/2019/11/cpj-raises-press-freedom-concerns-in-meeting-with/.

———. 2019b. "Police in Nigeria assault, arrest journalists covering #RevolutionNow protests." Committee to Protect Journalists, August 26. https://cpj.org/2019/08/police-in-nigeria-assault-arrest-journalists-cover/.

———. 2020. "Hungarian court gags investigative report, citing EU data protection law." Committee to Protect Journalists, October 21. https://cpj.org/2020/10/hungarian-court-gags-investigative-report-citing-eu-data-protection-law/.

Cushing, Tim. 2022. "Cops Are Still Playing Copyrighted Music To Thwart Citizens Recording Their Actions." Techdirt, April 18. www.techdirt.com/2022/04/18/cops-are-still-playing-copyrighted-music-to-thwart-citizens-recording-their-actions/.

Cyberlaw Clinic. 2020. "Access Denied: New White Paper on How US Copyright Policy Negatively Impacts Free Expression Worldwide." Cyberlaw Clinic, December 3. https://clinic.cyber.harvard.edu/2020/12/03/access-denied-new-white-paper-on-how-us-copyright-policy-negatively-impacts-free-expression-worldwide/.

Dalby, Douglas. 2019. "Panama Papers helps recover more than $1.2 billion around the world." International Consortium of Investigative Journalists, April 3. www.icij.org/investigations/panama-papers/panama-papers-helps-recover-more-than-1-2-billion-around-the-world/.

DeLisa, Nicholas Thomas. 2016. "You(Tube), Me, and Content ID: Paving the Way for Compulsory Synchronization Licensing on User-Generated Content Platforms." *Brooklyn Law Review* 81 (3): 1275–1318. https://brooklynworks.brooklaw.edu/blr/vol81/iss3/8.

Department for Digital, Culture, Media & Sport, Home Office, Nadine Dorries and Priti Patel. 2022. "Online safety law to be strengthened to stamp out illegal content." Press release, February 4. www.gov.uk/government/news/online-safety-law-to-be-strengthened-to-stamp-out-illegal-content.

Docksey, Christopher. 2022. "Journalism on trial and the right to be forgotten." *Verfassungsblog* (blog), March 9. https://verfassungsblog.de/journalism-rtbf/.

Doctorow, Cory. 2019. "Clever hack that will end badly: playing copyrighted music during Nazis rallies so they can't be posted to Youtube." Boing Boing, July 23. https://boingboing.net/2019/07/23/double-edged-swords-r-us.html.

———. 2020. "Mexico's New Copyright Law: Copying and Pasting USA's Flawed Copyright System Is a Human-Rights Catastrophe in the Making." Electronic Frontier Foundation. July. www.eff.org/files/2020/07/31/mexicos_new_copyright_law.pdf.

EU Disinfo Lab. 2022. "DSA: The Proposed Amendments to Article 12 and Recital 38 Should Be Rejected." EU Disinfo Lab, January 17. www.disinfo.eu/wp-content/uploads/2022/01/20220117_ARTICLE12RECITAL38.pdf.

European Digital Media Observatory. 2021. "EDMO Workshop — 'Media exemption' in the DSA: protecting editorial independence or a loophole for disinformation?" EDMO, November 22. https://edmo.eu/2021/11/22/edmo-workshop-media-exemption-in-the-dsa-protecting-editorial-independence-or-a-loophole-for-disinformation/.

Farid, Hany. 2021. "An Overview of Perceptual Hashing." *Journal of Online Trust and Safety* 1 (1): 1–22. https://doi.org/10.54501/jots.v1i1.24.

Forbidden Stories. 2023. "Story Killers: inside the deadly disinformation-for-hire industry." Forbidden Stories, February. https://forbiddenstories.org/case/story-killers/.

Fuller, Andrea, Kirsten Grind and Joe Palazzolo. 2020. "Google Hides News, Tricked by Fake Claims." *The Wall Street Journal,* May 15. www.wsj.com/articles/google-dmca-copyright-claims-takedown-online-reputation-11589557001.

Gault, Matthew. 2021. "Cop Was Instructed to Use Music to Disrupt Filming." *Vice,* September 9. www.vice.com/en/article/93y77y/cop-was-instructed-to-use-music-to-disrupt-filming.

Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* New Haven, CT: Yale University Press.

Global Disinformation Index. 2019. *The Quarter Billion Dollar Question: How is Disinformation Gaming Ad Tech?* Global Disinformation Index. September. www.disinformationindex.org/files/gdi_ad-tech_report_screen_aw16.pdf.

Global Internet Forum to Counter Terrorism. 2021. *Broadening the GIFCT Hash-Sharing Database Taxonomy: An Assessment and Recommended Next Steps.* Global Internet Forum to Counter Terrorism. July. https://gifct.org/wp-content/uploads/2021/07/GIFCT-TaxonomyReport-2021.pdf.

Goddard, Michelle. 2017. "The EU General Data Protection Regulation (GDPR): European Regulation that has a Global Impact." *International Journal of Market Research* 59 (6): 703–5. https://doi.org/10.2501/IJMR-2017-050.

Goldman, Eric and Jessica Silbey. 2020. "Copyright's Memory Hole." *BYU Law Review* 2019 (4): 929–96. https://digitalcommons.law.byu.edu/lawreview/vol2019/iss4/6.

Gorwa, Robert, Reuben Binns and Christian Katzenbach. 2020. "Algorithmic content moderation: Technical and political challenges in the automation of platform governance." *Big Data & Society* 7 (1): 205395171989794. https://doi.org/10.1177/2053951719897945.

Greenslade, Roy. 2015. "Syrian journalist who reported on Isis crimes in Raqqa murdered." *The Guardian*, December 17. www.theguardian.com/media/greenslade/2015/dec/17/syrian-journalist-who-reported-on-isis-crimes-in-raqqa-murdered.

Grimmelmann, James. 2015. "The Virtues of Moderation." *Yale Journal of Law & Technology* 17: 42–109. https://doi.org/10.31228/osf.io/qwxf5.

Guadamuz, Andres. 2017. "Developing a Right to be Forgotten." In *EU Internet Law: Regulation and Enforcement*, edited by Tatiana-Eleni Synodinou, Philippe Jougleux, Christiana Markou and Thalia Prastitou, 59–76. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-64955-9_3.

Guest, Peter. 2022. "Exposed documents reveal how the powerful clean up their digital past using a reputation laundering firm." Rest of World, February 3. https://restofworld.org/2022/documents-reputation-laundering-firm-eliminalia/.

Gunst, Simon and Ferdi De Ville. 2021. "The Brussels Effect: How the GDPR Conquered Silicon Valley." *European Foreign Affairs Review* 26 (3): 437–58. https://doi.org/10.54648/eerr2021036.

Hao, Karen. 2021. "How Facebook and Google fund global misinformation." *MIT Technology Review*, November 20. www.technologyreview.com/2021/11/20/1039076/facebook-google-disinformation-clickbait/.

Hauser, Dave. 2008. "The DMCA and the Privatization of Copyright." *Hastings Communications and Entertainment Law Journal* 30 (2): 339–56. https://repository.uchastings.edu/cgi/viewcontent.cgi?article=1649&context=hastings_comm_ent_law_journal.

Henley, Don. 2020. "Oral Statement of Don Henley before the Committee on the Judiciary, United States Senate Subcommittee on Intellectual Property on Section 512 of the Digital Millennium Copyright Act." United States Senate Subcommittee on Intellectual Property, June 2. www.judiciary.senate.gov/imo/media/doc/Henley%20Testimony.pdf.

Herrle, Jeanette and Jesse Hirsh. 2019. "The Peril and Potential of the GDPR." Opinion, Centre for International Governance Innovation, July 9. www.cigionline.org/articles/peril-and-potential-gdpr/.

Houser, Kimberly A. and W. Gregory Voss. 2018. "GDPR: The End of Google and Facebook or a New Paradigm in Data Privacy?" *Richmond Journal of Law and Technology* 25 (1): 1–109. https://jolt.richmond.edu/gdpr-the-end-of-google-and-facebook-or-a-new-paradigm-in-data-privacy/

Internews. 2021. "Internews-led Ads for News Selected as Finalist in Fast Company's 2021 World Changing Ideas Awards." Internews, May 3. https://internews.org/internews-led-ads-for-news-selected-as-finalist-in-fast-companys-2021-world-changing-ideas-awards/.

Johnson, Ashley and Daniel Castro. 2021. *How Other Countries Have Dealt With Intermediary Liability*. Information Technology & Innovation Foundation. February. www2.itif.org/2021-section-230-report-4.pdf.

Keller, Daphne. 2015. "Notice and Takedown Under the GDPR: An Operational Overview." *The Center for Internet and Society* (blog), October 29. https://cyberlaw.stanford.edu/blog/2015/10/notice-and-takedown-under-gdpr-operational-overview.

———. 2016. "Global Right to Be Forgotten Delisting: Why CNIL Is Wrong." *The Center for Internet and Society* (blog), November 18. https://cyberlaw.stanford.edu/blog/2016/11/global-right-be-forgotten-delisting-why-cnil-wrong.

———. 2018. "The Right Tools: Europe's Intermediary Liability Laws and the EU 2016 General Data Protection Regulation." *Berkeley Technology Law Journal* 33: 287–364. https://doi.org/10.15779/Z38639K53J.

———. 2020. "Systemic Duties of Care and Intermediary Liability." *The Center for Internet and Society* (blog), May 28. https://cyberlaw.stanford.edu/blog/2020/05/systemic-duties-care-and-intermediary-liability.

———. 2021. "Empirical Evidence of Over-Removal by Internet Companies Under Intermediary Liability Laws: An Updated List." *The Center for Internet and Society* (blog), February 8. https://cyberlaw.stanford.edu/blog/2021/02/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws.

Kemp, Simon. 2022. "The Latest Instagram Statistics: Everything You Need to Know." DataReportal. https://datareportal.com/essential-instagram-stats.

Klonick, Kate. 2018. "The New Governors: The People, Rules, and Processes Governing Online Speech." *Harvard Law Review* 131 (6): 1598–1670. https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf.

Koberidze, Maryna. 2015. "The DMCA Rulemaking Mechanism: Fail or Safe?" *Washington Journal of Law, Technology & Arts* 11 (3): 211–84. www.cs.yale.edu/homes/jf/Koberidze.pdf.

Krack, Noémie. 2021. "DSA proposal and disinformation — Should 'traditional media' be exempted from platform content moderation?" *KU Leuven Centre for IT & IP Law* (blog), December 7. www.law.kuleuven.be/citip/blog/.dsa-proposal-and-disinformation-should-traditional-media-be-exempted-from-platform-content-moderation/.

Krapiva, Natalia, Rodrigo Rodríguez and Alejandro Menjivar. 2020. "Warning: repressive regimes are using DMCA takedown demands to censor activists." Access Now, October 22. www.accessnow.org/dmca-takedown-demands-censor-activists/.

Krishnamurthy, Vivek and Jessica Fjeld. 2020. "CDA 230 Goes North American? Examining the Impacts of the USMCA's Intermediary Liability Provisions in Canada and the United States." SSRN. https://doi.org/10.2139/ssrn.3645462.

Krishnamurthy, Vivek, Mark Latonero, Rachel Kuchma, Elif Nur Kumru and Geneviève Plumptre. 2021. "Media Freedom and Technological Change." Carr Center Discussion Paper Series.

Lessig, Lawrence. 1999. *Code and Other Laws of Cyberspace*. New York, NY: Basic Books.

Lewin, Juan Esteban. 2019. "Nuestra Posición Sobre El 'Derecho al Olvido.'" La Silla Vacía, June 19. www.lasillavacia.com/la-silla-vacia/opinion/articulos-columna/nuestra-posici%C3%B3n-sobre-el-derecho-al-olvido-/.

———. 2020. "Eliminalia miente bajo juramento para acallar La Silla y medios." La Silla Vacía, January 30. www.lasillavacia.com/historias/silla-nacional/eliminalia-miente-bajo-juramento-para-acallar-la-silla-y-medios/.

Lizalek, Justin. 2021. "Flipping the DMCA and its Progeny on Their Heads: Content Creators Reclaiming Revenue From Improper Copyright Claims." *UIC Law Review* 54 (3): 757–96. https://repository.law.uic.edu/cgi/viewcontent.cgi?article=2854&context=lawreview.

Manancourt, Vincent. 2022. "EU privacy chief bashes lack of GDPR enforcement against Big Tech." *Politico*, June 17. www.politico.eu/article/gdpr-europe-wojciech-wiewiorowski-privacy-chief-lack-enforcement-big-tech/.

Maxwell, Andy. 2019. "YouTube Strikes Now Being Used as Scammers' Extortion Tool." TorrentFreak, January 30. https://torrentfreak.com/youtube-strikes-now-being-used-as-scammers-extortion-tool/.

McSherry, Corynne and Katharine Trendacosta. 2022. "What Companies Can Do Now to Protect Digital Rights In A Post-Roe World." Electronic Frontier Foundation, May 10. www.eff.org/deeplinks/2022/05/what-companies-can-do-now-protect-digital-rights-post-roe-world.

Meta. 2022a. "Dangerous Organizations: Terrorism and Organized Hate." Transparency Center, September 15. https://perma.cc/3KL4-DM2F.

———. 2022b. "Meta Reports Second Quarter 2022 Results." Meta, July 27. https://s21.q4cdn.com/399680738/files/doc_financials/2022/q2/Meta-06.30.2022-Exhibit-99.1-Final.pdf.

Mong, Attila. 2019. "In Romania, EU data protection law used to try to muzzle Rise Project." Committee to Protect Journalists, January 16. https://cpj.org/2019/01/in-romania-eu-data-protection-law-used-to-try-to-m/.

Newman, Nic, Richard Fletcher, Craig T. Robertson, Kirsten Eddy and Rasmus Kleis Nielsen. 2022. *Reuters Institute Digital News Report 2022*. Oxford, UK: Reuters Institute for the Study of Journalism, University of Oxford. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-06/Digital_News-Report_2022.pdf.

Nicholas, Gabriel. 2022. "Shedding Light on Shadowbanning." The Center for Democracy & Technology. April. https://doi.org/10.31219/osf.io/xcz2t.

O'Connell, Tyler. 2021. "Two Models of The Fourth Amendment and Hashing to Investigate Child Sexual Abuse Material." *University of the Pacific Law Review* 53 (1): 293–327. https://scholarlycommons.pacific.edu/uoplawreview/vol53/iss1/16.

People vs Bigtech. 2022. "MEPs Must Reject 'Media Exemption' Loopholes in the DSA." People vs Bigtech, January 16. https://peoplevsbig.tech/meps-must-reject-media-exemption-loopholes-in-the-dsa.

Perel, Maayan and Niva Elkin-Koren. 2016. "Accountability in Algorithmic Copyright Enforcement." *Stanford Technology Law Review* 19: 473–533. https://dx.doi.org/10.2139/ssrn.2607910.

Petrova, Anastasia. 2019. "The impact of the GDPR outside the EU." Lexology, September 17. www.lexology.com/library/detail.aspx?g=872b3db5-45d3-4ba3-bda4-3166a075d02f.

Ponciano, Jonathan. 2022. "The World's Largest Tech Companies In 2022: Apple Still Dominates As Brutal Market Selloff Wipes Trillions In Market Value." *Forbes*, December 5. www.forbes.com/sites/jonathanponciano/2022/05/12/the-worlds-largest-technology-companies-in-2022-apple-still-dominates-as-brutal-market-selloff-wipes-trillions-in-market-value/?sh=377b58a73448.

Preston, Douglas J. 2020. "Section 512 Hearing: Is the DMCA's Notice-and-Takedown System Working in the 21st Century?" United States Senate Committee on the Judiciary Subcommittee on Intellectual Property, June 2. www.judiciary.senate.gov/imo/media/doc/Preston%20Testimony.pdf.

Qurium Media Foundation. 2021. "Dark Ops Undercovered: Episode I — Eliminalia." Qurium Media Foundation, April 12. www.qurium.org/forensics/dark-ops-undercovered-episode-i-eliminalia/.

Radsch, Courtney. 2018a. "Tweaking a global source of news." *Columbia Journalism Review*. Winter. www.cjr.org/special_report/internet-intermediary-news.php/.

———. 2018b. "YouTube labels on public broadcasters draw ire in US, Russia." Committee to Protect Journalists, March 15. https://cpj.org/2018/03/youtube-labels-on-public-broadcasters-draw-ire-in/.

———. 2020a. "GIFCT: Possibly the Most Important Acronym You've Never Heard Of." Just Security, September 30. www.justsecurity.org/72603/gifct-possibly-the-most-important-acronym-youve-never-heard-of/.

———. 2020b. "Tech platforms struggle to label state-controlled media." Committee to Protect Journalists, August 12. https://cpj.org/2020/08/tech-platforms-struggle-to-label-state-controlled-media/.

———. 2020c. "The Politics of Labels: How Tech Platforms Regulate State Media." In *2020 Annual Report: Dynamic Coalition on the Sustainability of Journalism and News Media*, edited by Daniel O'Maley, Hesbon Hansen Owilla and Courtney C. Radsch, 37–49. https://gfmd.info/h-content/uploads/2021/11/DC-Sustainability-Annual-Report-2020-FINAL-gfmd.pdf.

———. 2021. "Hash/Hash Database." In *Glossary of Platform Law and Policy Terms*, edited by Luca Belli, Nicolo Zingales and Yasmin Curzi, 157–58. Rio de Janeiro: FGV Direito Rio. https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/31365/0.%20MIOLO_Glossary%20of%20Platform%20Law_digital.pdf?sequence=1&isAllowed=y.

———. 2022. "AI and Disinformation: State-Aligned Information Operations and the Distortion of the Public Sphere." Organization for Security and Co-operation in Europe. July. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4192038.

———. 2023a. "Platformization and Media Capture: A Framework for Regulatory Analysis of Media-Related Platform Regulations." *UCLA Journal of Law & Technology, Special Issue: Platforms and the Press* 28 (2): 175–223. https://uclajolt.com/platformatization-and-media-capture-a-framework-for-regulatory-analysis-of-media-related-platform-regulations/.

———. 2023b. *Urgent: Understanding and Responding to Global Emerging News Threats*. Washington, DC: Internews. March. https://internews.org/wp-content/uploads/2023/03/29Mar-URGENT-Report-Final.pdf.

———. Forthcoming 2023. "On the Frontlines of the Information Wars: How Algorithmic Gatekeepers and National Security Impact Journalism." In *National Security, Journalism and Law in the Age of Information Warfare*, edited by Marc Ambinder, Jennifer Henrichson and Connie Rosati. CERL. Oxford University Press.

Ravn, Michelle A. 1999. "Navigating Terra Incognita: Why the Digital Millennium Copyright Act Was Needed to Chart the Course of Online Service Provider Liability for Copyright Infringement." *Ohio State Law Journal* 60 (2): 755–98. https://kb.osu.edu/bitstream/handle/1811/65001/OSLJ_V60N2_0755.pdf.

Reda, Felix. 2022. "CJEU says upload filters must respect user rights — but what if they don't?" *Digital Freedom Fund* (blog), May 11. https://digitalfreedomfund.org/cjeu-says-upload-filters-must-respect-user-rights-but-what-if-they-dont/.

Riley, Chris and Susan Ness. 2022. "Modularity for International Internet Governance." *Lawfare* (blog), July 19. www.lawfareblog.com/modularity-international-internet-governance.

Satariano, Adam and Emma Bubola. 2019. "One Brother Stabbed the Other. The Journalist Who Wrote About It Paid a Price." *The New York Times*, September 23. www.nytimes.com/2019/09/23/technology/right-to-be-forgotten-law-europe.html.

Schiffer, Zoe and Adi Robertson. 2021. "Watch a police officer admit to playing Taylor Swift to keep a video off YouTube." The Verge, July 1. www.theverge.com/2021/7/1/22558292/police-officer-video-taylor-swift-youtube-copyright.

Schmon, Christoph, Filip Lukáš and Corynne McSherry. 2022. "The EU's Copyright Directive Is Still About Filters, But EU's Top Court Limits Its Use." Electronic Frontier Foundation, May 4. www.eff.org/deeplinks/2022/05/eus-copyright-directive-still-about-filters-eus-top-court-limits-its-use.

Sembra Media. 2021. *Inflection Point International: A study of the impact, innovation, threats, and sustainability of digital media entrepreneurs in Latin America, Southeast Asia, and Africa.* Inflection Point International. https://data2021.sembramedia.org/wp-content/uploads/2021/11/Inflection-point-ENG-Nov3-2021-2.pdf.

Shores, Mark L. 2019. "Internet Reviews: The Rise of Content Farms." *Kentucky Libraries* 75 (3): 14–15. June. http://sc.lib.miamioh.edu/handle/2374.MIA/6433.

Simon, Joel. 2023. "Press Freedom Community: Prioritize the Defense of Journalism that Serves the Public Interest." *Nieman Reports,* February 7. https://niemanreports.org/articles/press-freedom-public-interest/.

Skai. 2019. "How Many Google Searches Per Day Are There? Useful Search Metrics for Marketers." Skai, February 25. https://skai.io/monday-morning-metrics-daily-searches-on-google-and-other-google-facts/.

Snow, Brandi M. 2009. "SLAPP Suits." The First Amendment Encyclopedia. www.mtsu.edu/first-amendment/article/1019/slapp-suits.

Soraide, Rosario. 2018. "The 'misuse' of the judicial system to attack freedom of expression: Trends, Challenges and Responses." Issue Brief, World Trends in Freedom of Expression and Media Development. CI-2022/WTR/4. https://unesdoc.unesco.org/ark:/48223/pf0000383832.

Southwick, Natalie and Carlos Martínez de la Serna. 2022. "In 2022, journalist killings continue unabated in Mexico amid a climate of impunity." Committee to Protect Journalists, August 30. https://cpj.org/2022/08/in-2022-journalist-killings-continue-unabated-in-mexico-amid-a-climate-of-impunity/.

Statista. 2021. "YouTube: Hours of Video Uploaded Every Minute 2020." Statista. www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/.

Sung, Morgan. 2021. "Cops are playing music during filmed encounters to game YouTube's copyright striking." Mashable, July 1. https://mashable.com/article/police-playing-music-copyright-youtube-recording.

Syrian Archive. 2022. "Lost and Found: Syrian Archive's work on content taken down from social media platforms." https://syrianarchive.org/en/lost-found.

Tewari, Shreya. 2021. "Evolution of DMCA Notices: Trends and a Timeline." Lumen Database, July 2. https://lumendatabase-org.medium.com/evolution-of-dmca-notices-trends-and-a-timeline-32636581af01.

———. 2022. "Over thirty thousand DMCA notices reveal an organized attempt to abuse copyright law." *Lumen* (blog), April 22. www.lumendatabase.org/blog_entries/over-thirty-thousand-dmca-notices-reveal-an-organized-attempt-to-abuse-copyright-law.

Thomas, Dexter. 2021a. "Is This Beverly Hills Cop Playing Sublime's 'Santeria' to Avoid Being Live-Streamed?" Vice, February 9. www.vice.com/en/article/bvxb94/is-this-beverly-hills-cop-playing-sublimes-santeria-to-avoid-being-livestreamed.

———. 2021b. "New Video Shows Beverly Hills Cops Playing Beatles to Trigger Instagram Copyright Filter." Vice, February 11. www.vice.com/en/article/bvxa7q/new-video-shows-beverly-hills-cops-playing-beatles-to-trigger-instagram-copyright-filter.

Turner Lee, Nicol, Paul Resnick and Genie Barton. 2019. *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms.* Brookings, May 22. www.brookings edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/.

Unicorn Riot. 2020. "Facebook and YouTube have algorithmically interfered with our media coverage due to ambient copyrighted music. We are forced to delete interview audio overlaying background music 😶😶😶... standby." (Twitter thread). Twitter, June 2, 10:17 p.m. https://twitter.com/UR_Ninja/status/1268003754445602816.

United Nations Human Rights Office of the High Commissioner. 2022. "UN experts concerned by systematic use of SLAPP cases against human rights defenders by businesses." Press release, December 16. www.ohchr.org/en/press-releases/2022/12/un-experts-concerned-systematic-use-slapp-cases-against-human-rights.

United States Copyright Office. 2020. *Section 512 of Title 17: A Report of the Register of Copyrights.* United States Copyright Office. May. www.copyright.gov/policy/section512/section-512-full-report.pdf.

Van Eecke, Patrick. 2011. "Online service providers and liability: A plea for a balanced approach." *Common Market Law Review* 48 (5): 1455–1502. https://kluwerlawonline.com/journalarticle/Common+Market+Law+Review/48.5/COLA2011058.

Van Quathem, Kristof and Nicholas Shepherd. 2019. "German Constitutional Court Reshapes 'Right to be Forgotten' and Expands Its Oversight of Human Rights Violations." Inside Privacy, December 3. www.insideprivacy.com/data-privacy/german-constitutional-court-reshapes-right-to-be-forgotten-and-expands-its-oversight-of-human-rights-violations/.

Vílchez, Dánae. 2020. "YouTube censors independent Nicaraguan news outlets after copyright complaints from Ortega-owned media." Committee to Protect Journalists, May 6. https://cpj.org/2020/05/youtube-censor-nicaragua-outlets-100-noticias-confidencial-ortega/.

Vining, Austin and Sarah Matthews. n.d. "Overview of Anti-SLAPP Laws." Reporters Committee for Freedom of the Press. www.rcfp.org/introduction-anti-slapp-guide/.

Volokh, Eugene. 2022. "The Reverse Spider-Man Principle: With Great Responsibility Comes Great Power." The Volokh Conspiracy, April 29. https://reason.com/volokh/2022/04/29/the-reverse-spider-man-principle-with-great-responsibility-comes-great-power/.

YouTube Team. 2021. "Access for all, a balanced ecosystem, and powerful tools." *YouTube Official Blog*, December 6. https://blog.youtube/news-and-events/access-all-balanced-ecosystem-and-powerful-tools/.

Centre for International
Governance Innovation

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

@cigionline