ReCreating Europe

Copyright Content Moderation in the EU: Conclusions and Policy Recommendations

Authors

João Pedro Quintais, Christian Katzenbach, Sebastian Felix Schwemer, Daria Dergacheva, Thomas Riis, Péter Mezei, and István Harkai

Deliverable Title	D6.4 - Best Practices and Policy Recommendations Brief
Deliverable Lead:	UvA (IViR)
Partner(s) involved:	UvA (IViR), UCPH (CIIR), UBREMEN, SZG
Related Work Package:	WP6 - Intermediaries: Copyright Content Moderation and Removal at Scale in the Digital Single Market: What Impact on Access to Culture?
Related Task/Subtask:	T.6.5. Synthesize and Articulate Recommendations and Best Practices
Main Author(s):	João Pedro Quintais (UvA) Christian Katzenbach (UBREMEN) Sebastian Felix Schwemer (UCPH)
Other Author(s):	Daria Dergacheva (UBREMEN), Thomas Riis (UCPH), Péter Mezei (USZ), and István Harkai (USZ),
Dissemination Level:	Public
Due Delivery Date:	31.12.2022
Actual Delivery:	30.01.2023
Project ID:	870626
Instrument:	H2020-SC6-GOVERNANCE-2019
Start Date of Project:	01.01.2020
Duration:	39 months

TABLE OF CONTENTS

Abbreviation List 1
Figures 2
Executive Summary 3
1. Introduction 5
2. Mapping of Copyright Content Moderation Rules and Practices 6
2.1. Conceptual Framework 6
2.2. Copyright Content Moderation Rules at the EU Level 9
2.3. Copyright Content Moderation Rules at National Level 17
2.4. Private Regulation by Platforms: empirical research 20
3. Evaluation and Measuring: impact of moderation practices and technologies on access and diversity 23
3.1. Evaluating multi-level legal frameworks 25
3.1.1. Overlaps and interplay of existing legal frameworks 25
3.1.2. Benchmarks for normative assessment: “rough justice” and “quality” 28
3.1.3. Looking into the future: Context and bias in content moderation 32
3.2. Measuring the impact of moderation practices and technologies on access and diversity 34
3.2.1. Assessing transparency reports 37
3.2.2. Measuring content blocking and deletion on platforms, and its impact on diversity 38
3.2.3. Social media creators’ perspective on copyright content moderation in the EU 41
4. Recommendations for Future Policy Actions 43
References 47

ABBREVIATION LIST

AG	Advocate General
AI	Artificial Intelligence
AIA proposal	European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, COM/2021/206 final
CDSMD	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
Charter	Charter of Fundamental Rights of the European Union
CJEU	Court of Justice of the European Union
DSA	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 (Digital Services Act)
DSM	Digital Single Market
e-Commerce Directive	Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market [2000] OJ L178/1
EC	European Commission
E&Ls	Exceptions and /or limitations
EU	European Union
InfoSoc Directive	Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society OJ L 167, 22.6.2001, CFp. 10–19
IP	Intellectual Property
OCSSP	Online content-sharing service provider
T&Cs	Terms and Conditions
UGC	User-generated content
VLOP	Very Large Online Platform

FIGURES

Figure 1. Relationship substantive copyright rules and intermediary framework 26
Figure 2. Relationship between rules on intermediaries and industry practices 26
Figure 3. Empirical Research Design (Representation) 36

EXECUTIVE SUMMARY

This report describes and summarizes the results of the research carried out on the mapping of the EU legal framework and intermediaries' practices on copyright content moderation and removal. In particular, this report summarizes the results of our previous deliverables and tasks, namely: **(1)** D.6.2. Final Report on mapping of EU legal framework and intermediaries' practices on copyright content moderation and removal, which includes our research in the Tasks T.6.1.1 (EU Level Mapping); Task T.6.1.2 (Comparative National Level Mapping); Task T.6.1.3 (Private Regulations by Platforms: ToS, Community Guidelines); and **(2)** D.6.3 Final Evaluation and Measuring Report - impact of moderation practices and technologies on access and diversity, which includes our research in Task 6.3 (Evaluating Legal Frameworks on the Different Levels (EU vs. national, public vs. private) and Task 6.4 (Measuring the impact of moderation practices and technologies on access and diversity).

Our previous reports contain a detailed description of the legal and empirical methodology underpinning our research and findings. This report focuses on bringing together these findings in a concise format and advancing policy recommendations. After a brief introductory chapter, **Section 2** of the report summarizes the main conclusions and findings from our **mapping analysis** into content moderation of copyright-protected content on online platforms in the EU. This analysis covers our conceptual framework, copyright content moderation rules at EU and national level, and our empirical research on private regulation by platforms. Regarding the latter, we studied the copyright content moderation structures adopted by 15 social media platforms over time, with a focus on their terms and conditions and automated systems.

Section 3 then summarizes the main conclusions and findings from our **evaluation analysis**. This includes first a **legal and normative analysis on multi-level legal frameworks** regulating copyright content moderation, which covers an examination of the overlaps and interplay of existing legal frameworks, the development of benchmarks for normative assessment (focusing on concept of "rough justice" and "quality" of moderation), and, with a view to future regulation in this field, a reflection on context and bias in copyright content

moderation. The **empirical prong of our research addresses the challenging topic of measuring the impact of moderation practices and technologies on access and diversity**. To do so, we tackle **three dimensions** of this problem: **(1)** we investigate all the aggregated data on copyright moderation provided by the platforms themselves; **(2)** we analyse content level data of platforms with regard to changes and factors of cultural diversity on social media and streaming platforms, specifically YouTube; **(3)** we explore creators' understanding and experiences of copyright moderation in relation to their creative work and the labour of media production on social media platforms

Section 4 outlines our **policy recommendations for EU and national policymakers**. These recommendations touch upon the following topics: the definition of "online content-sharing service provider"; the recognition and operationalisation of user rights; the complementary nature of complaint and redress safeguards; the scope of permissible preventive filtering; the clarification of the relationship between art. 17 CDSMD and the DSA, including as regards the application of fundamental rights through terms and conditions; monetisation and restrictive content moderation actions; recommender systems and copyright content moderation; transparency and data access for researchers; trade secret protection and transparency of content moderation systems; the relationship between art. 17 CDSMD, the DSA and the AI Act Proposal respectively; and human competences in copyright content moderation.

1. INTRODUCTION

This research is part of the reCreating Europe project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870626. This report describes and summarizes the results of the research carried out in the context of Work Package (WP) 6 on the mapping of the EU legal framework and intermediaries' practices on copyright content moderation and removal. In particular, this report summarizes the results of our previous deliverables and tasks, namely:

- D.6.2. Final Report on mapping of EU legal framework and intermediaries' practices on copyright content moderation and removal, which includes our research in the Tasks T.6.1.1 (EU Level Mapping); Task T.6.1.2 (Comparative National Level Mapping); Task T.6.1.3 (Private Regulations by Platforms: ToS, Community Guidelines); and
- D.6.3 Final Evaluation and Measuring Report - impact of moderation practices and technologies on access and diversity, which includes our research in Task 6.3 (Evaluating Legal Frameworks on the Different Levels (EU vs. national, public vs. private) and Task 6.4 (Measuring the impact of moderation practices and technologies on access and diversity).

Our previous reports contain a detailed description of the legal and empirical methodology underpinning our research and findings. This report focuses on bringing together these findings in concise format and advancing policy recommendations to that basis. For that reason, we have also limited references and sources to the minimum necessary.

The report proceeds as follows. **Section 2** summarizes the main conclusions and findings from our mapping analysis into content moderation of copyright-protected content on online platforms in the EU. This mapping analysis provides a basis for our subsequent normative and evaluative research. **Section 3** then summarizes the main conclusions and findings from our evaluation analysis. On that basis, **Section 4** outlines our policy recommendations for EU and national policymakers.

2. MAPPING OF COPYRIGHT CONTENT MODERATION RULES AND PRACTICES[1]

The main research question of our extensive mapping analysis is as follows:

- How can we map the impact on access to culture in the Digital Single Market (DSM) of content moderation of copyright-protected content on online platforms?

We divide this question into multiple sub-research questions (**SQR**), listed below.

- **SQR(1):** How to conceptualise and approach from a methodological perspective the interdisciplinary analysis of content moderation of copyright-protected content on online platforms and its impact on access to culture in the DSM?
- **SQR(2):** How is the private and public regulatory framework for content moderation for online platforms structured?
- **SQR(3):** How do the various elements of that regulatory framework interact?
- **SQR(4):** How are copyright content moderation rules organized by platforms into public documents?
- **SQR(5):** Which copyright content moderation rules do different platforms employ to regulate copyright, and how have they changed over time?
- **SQR(6):** How do platforms' automated copyright content moderation systems work?

2.1. CONCEPTUAL FRAMEWORK[2]

Our analysis starts by addressing SQR(1): How to conceptualise and approach from a methodological perspective the interdisciplinary analysis of content moderation of copyright-protected content on online platforms and its impact on access to culture in the DSM?

To answer this question, we develop a conceptual framework and interdisciplinary methodological approach to examine copyright content moderation on online platforms and its potential impact on access to culture. The analysis clarifies our terminology, distinguishes between platform "governance" and "regulation", elucidates the concept of "online platform", and positions our research in the context of regulation "of", "by" and "on"

[1] This section of the report is based on João Pedro Quintais and others, 'Copyright Content Moderation in the EU: An Interdisciplinary Mapping Analysis' (2022) reCreating Europe Report <https://papers.ssrn.com/abstract=4210278> accessed 7 September 2022.
[2] This section summarizes the contents of Quintais and others (n 1) ch 2.

platforms. Our legal analysis focuses on the regulation "of" platforms, predominantly through EU and national law. This includes, to name the most relevant, the Charter of Fundamental Rights of the EU, the InfoSoc Directive (2001/29/EC), the CDSMD (2019/790), the e-Commerce Directive (2000/31/EC), and the DSA (Regulation (EU) 2022/2065).

Our empirical analysis focuses on a subset of the regulation "by" platforms. In this context, an effort is made to clarify the meaning of the structures of copyright content moderation that underpin our analysis, namely the resources that platforms create and employ to regulate copyright. The main structure we focus on relates to the rules set by platforms to moderate copyright-protected content, mainly their terms and conditions (**T&Cs**)[3], which we consider playing a dual role: normative and performative.[4]

A second structure we examine refers to the systems that platforms deploy to automatically moderate and enforce copyright through computational techniques, such as content recognition and filtering/blocking tools. Both structures are also examined later on from the perspective of EU law.

Building on the concept of "content moderation" in the Digital Services Act (**DSA**)[5], we advance a working definition of "copyright content moderation" as

the activities, automated or not, undertaken by providers of hosting services – either as consequence of a legal notice-and-action obligation or as voluntary activity – aimed in particular at detecting, identifying and addressing content or information that is illegal under EU copyright law and is incompatible with providers' T&Cs, provided by recipients of the service, including measures taken that affect the availability, visibility and accessibility of that illegal content or that information, such as demotion, demonetisation, disabling of access to,

[3] Our analysis adopts the definition of "Terms and Conditions" in art. 3(u) DSA. On this provision and its fundamental rights implications, see João Pedro Quintais, Naomi Appelman and Ronan Fahy, 'Using Terms and Conditions to Apply Fundamental Rights to Content Moderation' [2023] German Law Journal.

[4] In our view, the normative role of T&Cs stems from the fact that the very public codification of what counts as an acceptable conduct creates expectations of accountability that are potentially mutual, even if radically unequitable. Differently, the performative role results from the fact that that, by the virtue of being public, these rules are, inevitably, an organizational performance.

[5] See art. 3(t) DSA.

or removal thereof, or the recipients' ability to provide that information, such as the termination or suspension of a recipient's account.

This concept, when seen in light of our subsequent analysis, elucidates the fact that many content moderation activities are not explicitly regulated in EU copyright law. Hence, the regulation of such activities is mostly left to the complementary application of other instruments (e.g., the DSA), national legislators' margin of discretion, and – perhaps predominantly – private ordering by online platforms (e.g. through their T&Cs). Our research shows that EU copyright law mostly focuses on what could be understood as "hard-line" moderation of *content items*, namely certain measures aimed at addressing the availability or accessibility of content, such ex ante filtering, blocking or removal of content items. This results in regulatory gaps in the EU copyright law coverage of copyright content moderation activities. In particular, there are no explicit rules on measures: (1) affecting the visibility and monetisation of content; or (2) addressing a user's ability to provide information, e.g., relating to the termination or suspension of his account. As we note below in our recommendations, this regulatory gap should be further examined by policymakers, especially as regards monetisation activities.

Finally, in preparation of the evaluation of the results from the mapping analysis, we briefly outline a possible approach to define access to culture for purposes of content moderation, highlighting the descriptive and normative dimensions of the concept. The descriptive dimension posits that the "quality" of copyright content moderation is correlated to access to culture, because access to culture is considered embedded in the existing copyright framework. Since the existing framework is assumed to strike the appropriate balance between exclusivity in copyright protection and access to culture, any deviation from that balance – beyond the margin of interpretation allowed by law – will impact on access to culture. While obviously insufficient per se, this descriptive dimension is useful insofar as it provides a theoretical framework to compartmentalize the specific issues of copyright content moderation by online platforms. The focus of our approach is on the "downstream" issue of mitigation of errors in content moderation (i.e., false positives and false negatives).

This is particularly relevant in the context of EU copyright law, since the crux of the balance sought by the Advocate General (**AG**) and the Court of Justice in Case C-401/19[6] (on the validity of art. 17 CDSMD) is placed on whether ex-ante filtering measures can be deployed while avoiding the risks of over-blocking (and false positives) to platform users' right to freedom of expression.

The normative dimension, on the other hand, rejects the notion that the existing copyright framework strikes the optimal balance between exclusivity in copyright protection and access to culture. The model suggests that substantive law relevant in the field of copyright can be amended in a way that changes the balance with the result that it further increases access to culture by providing more freedoms to third parties to use and disseminate copyright-protected works, without encroaching on the legitimate interest of copyright holders. The actual practices of content moderation by platforms are affected by the state-enacted law (including case law) that platforms are subject to, which determines their "autonomy space" in defining such practices. In other words, the legal regulation "of" platforms determines the space available for regulation "by" platforms. Under this framework, adjustments to state-enacted law can affect the content moderation practices of platforms either by narrowing down their autonomy space (e.g., by broadening the scope of liability for platforms) or by raising the costs of acting outside the autonomy space (e.g., introducing more severe sanctions and more effective remedies). Both the descriptive and normative approach are useful to frame and understand EU copyright law's approach to regulating content moderation by platforms. We develop further this analysis in section 3 below.

2.2. COPYRIGHT CONTENT MODERATION RULES AT THE EU LEVEL[7]

Our mapping analysis then moves to aims to answer SQR(2) and SQR(3) from the perspective of EU law:

[6] AG Opinion in Case C-401/19, Republic of Poland v European Parliament, Council of the European Union, 15.07.2021, ECLI:EU:C:2021:613; Judgment of the Grand Chamber in Case C-401/19, Republic of Poland v European Parliament and Council of the European Union, 26.04.2022, ECLI:EU:C:2022:297.
[7] This section summarizes the contents of Quintais and others (n 1) ch 3.

- **SQR(2):** How is the private and public regulatory framework for content moderation for online platforms structured?
- **SQR(3):** How do the various elements of that regulatory framework interact?

For this purpose, we carry out a mapping of copyright content moderation by online platforms at secondary EU law level. The analysis starts with an exposition of the baseline regime from which art. 17 CDSMD departs from, which we call the pre-existing *acquis*. EU law has been subject to a high level of harmonization stemming from many directives on copyright and related rights, the interpretation of which is determined by the case law of the CJEU. In particular, the legal status of copyright content moderation by online platforms under this regime is mostly set by the Court's interpretation of arts. 3 and 8(3) InfoSoc Directive – on direct liability for communication to the public and injunctions against intermediaries – and arts. 14 and 15 e-Commerce Directive – on the hosting liability exemption and the prohibition on general monitoring obligations.[8] We explain this case law and its implications for platform liability and content moderation obligations up to the Court's Grand Chamber judgment in *YouTube and Cyando*[9], and how those developments contributed to the proposal and approval of art. 17 CDSMD.

Setting aside the political nature of legislative processes, from a systematic and historical perspective, art. 17 CDSMD and subsequently the DSA can be seen as the result of efforts in EU law and its interpretation by the Court for the last 20 years to adapt to technological developments and the changing role and impact of platforms on society. The result has been a push towards "enhanced" responsibility for platforms, characterised by additional liability and obligations regarding content they host and services they provide, as well as an increased role of fundamental rights – especially of users – in the legal framework.

[8] These provisions were replaced by arts. 4 to 10 DSA.

[9] Joined Cases C-682/18 and C-683/18, Frank Peterson v Google LLC, YouTube Inc., YouTube LLC, Google Germany GmbH (C-682/18), and Elsevier Inc. v Cyando AG (C-683/18), 22.06.2021, ECLI:EU:C:2021:503 (*Youtube and Cyando*). For a comment in the context of our research project, see João Quintais and Christina Angelopoulos, 'YouTube and Cyando, Joined Cases C-682/18 and C-683/18 (22 June 2021): Case Comment' [2022] Auteursrecht 46.

The heart of this part of the analysis is the complex legal regime of art. 17 CDSMD, which we carry out in light of existing scholarship, the Commission's Guidance on that provision[10], the AG Opinion and Court's Grand Chamber judgment in Case C-401/19. Our analysis sets out in detail the different components of this hybrid regime, including:

- The creation of the new legal category of "online content-sharing service providers" (OCSSPs), a sub-type of hosting service providers under the e-Commerce Directive, and "online platforms" under the DSA;
- The imposition of direct liability on OCSSPs for content they host and provide access to;
- The merged authorization regime for acts of OCSSPs and their uploading users, provided the user act does not generate significant revenue;
- The *lex specialis* nature of art. 17 CDSMD in relation to art. 3 InfoSoc Directive and art. 14 e-Commerce Directive, which is endorsed explicitly by the Commission's Guidance and AG the Opinion in C-401/19, and in our view implicitly by the Court in the same judgment;
- The relationship between the prohibition on general monitoring obligations in art. 15 e-Commerce and art. 17(8) CDSMD, where we argue that the latter may be understood as being of merely declaratory nature;
- The complex liability exemption mechanism comprised of best efforts obligations on OCSSPs (to obtain an authorization and to impose preventive and reactive measures) in art. 17(4); and
- The substantive and procedural safeguards in the form of exceptions or limitations (**E&Ls**) or "user rights" and in-/out-of-platform (complaint and) redress mechanisms in art. 17(7) and (9).

Our analysis addresses multiple points of uncertainty in this complex regime, some of which will no doubt be subject to litigation at the national level and likely the CJEU. The following aspects are worth highlighting, however, as they also reflect possible points of improvement of this regime from the perspective of copyright content moderation.

First, whether an online platform is subject to the pre-existing regime (as updated by the DSA) or the new regime in art. 17 CDSMD will depend on whether it qualifies as an **OCSSP**. Our research shows that there is significant legal uncertainty as regards this qualification, despite the Commission's Guidance. To be sure, certain large-scale platforms, especially with video-sharing features (e.g., YouTube, Meta/Facebook, Instagram), clearly qualify as OCSSPs. Others will also clearly be excluded from the scope of art. 17 because they are covered by the

[10] Communication from the Commission to the European Parliament and the Council Guidance on Article 17 of Directive 2019/790 on Copyright in the Digital Single Market, COM/2021/288 (final) (Guidance art. 17 CDSMD).

definitional carve-outs in art. 2(6) CDSMD.[11] Still, there remains a significant grey area, which affects both larger platforms and (especially) medium-sized and small platforms. The main reason is that the definition includes a number of open-ended concepts ("main purpose", "large amount", "profit-making purpose") that ultimately require a case-by-case assessment of what providers qualify as an OCSSP. Such assessment would partly take place In the context of the respective national Member State, which may lead to further uncertainty. Furthermore, even where it can be established that a platform falls within the scope of the legal definition, it might remain unclear to what extent it does. This is illustrated by the Guidance's statement that if a provider offers multiple services, then there is a need for service-by-service analysis to assess whether it qualifies as an OCSSP. This approach, although understandable, introduces complexity in determining relevant services and subsequent attribution of liability. The outcome might well be that the same provider is subject to art. 17 CDSMD for certain services and the pre-existing regime for others. Once we scale up this issue to numerous platforms hosting copyright protected content, each providing different services, the complexity of determining liability regimes and respective content moderation obligations -outside the most prominent and politically featured cases- becomes clear.

Second, a crucial part of our analysis of platforms' liability and copyright content moderation obligations refers to what we call the **normative hierarchy of art. 17 CDSMD**. We provide a critical analysis of how the Commission's Guidance has attempted to address this hierarchy and strike the balance between the competing rights and interests of rightsholders, platforms and users, drawing from the arguments of AG Opinion and CJEU judgment in C-401/19.

The first important implication of the judgment is that the Court recognizes that art. 17(7) CDSMD includes an obligation of result. As such, Member States must ensure that these E&Ls are respected despite the preventive measures in paragraph (4), qualified as "best efforts" obligations. This point, already recognized by the AG and in the Commission's Guidance, is

[11] See art. 2(6) CDSMD, second paragraph: "Providers of services, such as not-for-profit online encyclopedias, not-for-profit educational and scientific repositories, open source software-developing and-sharing platforms, providers of electronic communications services as defined in Directive (EU) 2018/1972, online marketplaces, business-to-business cloud services and cloud services that allow users to upload content for their own use, are not 'online content-sharing service providers' within the meaning of this Directive."

reinforced by the Court's recognition that the mandatory E&Ls, coupled with the safeguards in paragraph (9), are "user rights", not just mere defences.[12]

The second and related main implication of the judgment is that the Court rejects the possibility of interpretations of art. 17 that rely solely on ex post complaint and redress mechanisms as a means to ensure the application of user rights. That was for instance the position defended by certain Member States during the hearing before the Court and in their national implementations. Instead, the judgment clarifies that Member States' laws must first and foremost limit the possibility of deployment of ex ante filtering measures; assuming that occurs, the additional application of ex post safeguards is an adequate means to address remaining over-blocking issues. This conclusion should be welcomed, especially in light of existing evidence that complaint and redress mechanisms are seldom used by users.

The third main implication of the judgment relates to the scope of permissible ex ante filtering by platforms. On this point, the Guidance states that automated filtering and blocking measures are "in principle" only admissible for "manifestly infringing" and "earmarked" content. However, the Court states unequivocally that only filtering/blocking systems that can distinguish lawful from unlawful content without the need for its "independent assessment" by OCSSPs are admissible. Only then will these measures not lead to the imposition of a prohibited general monitoring obligation under art. 17(8) CDSMD. Furthermore, these filters must be able to ensure the exercise of user rights to upload content that consists of quotation, criticism, review, caricature, parody, or pastiche.

On this point, it is noteworthy that the judgment endorses by reference the AG Opinion, which states inter alia that filters "must not have the objective or the effect of preventing such legitimate uses", and that providers must "consider the collateral effect of the filtering measures they implement", as well as "take into account, ex ante, respect for users' rights".[13]

[12] On this topic, see Sebastian Felix Schwemer and Jens Schovsbo, 'What Is Left of User Rights? – Algorithmic Copyright Enforcement and Free Speech in the Light of the Article 17 Regime', *Paul Torremans (ed), Intellectual Property Law and Human Rights* (4th edition, Wolters Kluwer 2020) <https://ssrn.com/abstract=3507542>.
[13] AG Opinion in Case C-401/19, Republic of Poland v European Parliament, Council of the European Union, 15.07.2021, ECLI:EU:C:2021:613, para 193.

In our view, considering the Court's statements in light of the previous case law and current market and technological reality, the logical conclusion is that only content that is "obviously" or "manifestly" infringing – or "equivalent" content – may be subject to ex ante filtering measures. Beyond those cases, for instance as regards purely "earmarked content", the deployment of ex ante content filtering tools appears to be inconsistent with the judgment's requirements.

It also remains to be seen whether this reasoning applies more broadly to other types of illegal content beyond copyright infringement. If it does, it might help to shape the scope of prohibited general monitoring obligations versus permissible "specific" monitoring, with relevance for future discussions on the DSA. In drawing these lines, caution should be taken in the application of the "equivalent" standard in *Glawischnig-Piesczek*[14], which likely requires a much stricter interpretation for filtering of audio-visual content in OCSSPs than textual defamatory posts on a social network.

Finally, we provide a brief analysis of the **interplay between art. 17 CDSMD and the potentially applicable provisions of the DSA proposal to OCSSPs**. On this topic, we refer readers to our parallel research, which offers an in-depth analysis.[15] With regard to copyright-protected material and online platforms, the DSA matters at two levels. First, because it replaces the e-Commerce Directive, the DSA and its rules on liability and due diligence obligations will apply to all providers that do not qualify as OCSSPs. Second, and less obvious, the direct application of the DSA to OCSSPs covered by the liability regime in art. 17 CDSMD. Both art. 17 CDSMD and multiple provisions of the DSA impose obligations on how online platforms deal with illegal information. Whereas art. 17 CDSMD targets copyright infringing content, the DSA proposal targets illegal content in general, including that which infringes copyright.

[14] Case C-18/18, Eva Glawischnig-Piesczek v Facebook Ireland Limited, 3.10.2019, ECLI:EU:C:2019:821.

[15] João Pedro Quintais and Sebastian Felix Schwemer, 'The Interplay between the Digital Services Act and Sector Regulation: How Special Is Copyright?' (2022) 13 European Journal of Risk Regulation 191. See also Alexander Peukert and others, 'European Copyright Society – Comment on Copyright and the Digital Services Act Proposal' (2022) 53 IIC - International Review of Intellectual Property and Competition Law 358.

Departing from the observation that a platform may qualify as an OCSSP under the CDSMD and an "online platform" (and "very large online platform") under the DSA, we conclude that the DSA will apply to OCSSPs insofar as it contains rules that regulate matters not covered by art. 17 CDSMD, as well as specific rules on matters where art. 17 leaves a margin of discretion to Member States. Importantly, we consider that such rules apply even where art. 17 CDSMD contains specific (but less precise) regulation on the matter. In our view, although there is significant legal uncertainty in this regard, such rules include both provisions in the DSA's liability framework and in its due diligence obligations (e.g., as regards the substance of notices, complaint and redress mechanisms, trusted flaggers, protection against misuse, risk assessment and mitigation, and data access and transparency).

In light of the above, one important conclusion from our analysis is the emergence of a bifurcated or multilevel legal framework for online platforms engaging in copyright content moderation. On the one hand, OCSSPs are subject to the regime of art. 17 CDSMD as regards liability and content moderation. On the other hand, non-OCSSPs are subject to the pre-existing regime under the InfoSoc and e-Commerce Directives (and now the DSA), as interpreted by the CJEU (most recently in *YouTube and Cyando*). Although the regimes have similarities – and can be approximated through the Court's interpretative activity – they are structurally different. This divergence may lead to further fragmentation, on top of the fragmentation that is to be expected by the national implementations of the complex mechanisms in art. 17 CDSMD. To this we must add the application of the horizontal rules on content moderation liability and due diligence obligations arising from the DSA. In sum, the multi-level and multi-layered EU legal landscape on copyright content moderation that emerges from our mapping analysis is extremely complex.

Relatedly, as anticipated above, certain copyright content moderation issues of relevance remain unregulated in the copyright *acquis*, namely rules on measures: affecting the visibility and monetisation of content; and addressing a user's ability to provide information, e.g., relating to the termination or suspension of his account. Although both categories are relevant, the issue of monetisation is in our perspective the most glaring regulatory gap, since

"monetisation" actions play a central and financial consequential role in platforms' content moderation practices.

This is clear, for instance, from examining YouTube's latest (at time of writing) copyright transparency report, containing data from the first semester of 2022.[16] As described therein, ContentID is one of three tools of YouTube's Copyright Management Suite, together with the webform and the Copyright Match tool. Contrary to the other tools, ContentID is only available to users with a "[d]emonstrated need of scaled tool, understanding of copyright, and resources to manage complex automated matching system…".[17] ContentID thus aims at serving the needs of users that are large copyright holders, so-called "enterprise partners" like "movie studios, record labels, and collecting societies".[18] ContentID is the only tool in YouTube's Copyright Management Suite that allows users the option to monetize matched content, in addition to tracking and blocking it.[19] Importantly, YouTube reports that rightsholders using the tool opted to monetize 90% of claims on ContentID during the period reported.[20] In other words, the vast majority of rightsholders claims on ContentID during this period (amounting to over 750 million claims) are aimed at monetization rather than preventing the availability of content.[21]

This topic should therefore be subject to further research and policy action in the near future.

Still as regards regulatory gaps, it is important to underscore the complexity of the legal determinations and judgments required to assess human and algorithmic copyright content moderation practices. This strongly suggests a need for better transparency and access to data from platforms. In these regards, both the pre-existing regime prior to the DSA and art.

[16] YouTube, 'YouTube Copyright Transparency Report H1 2022' (YouTube 2022) Copyright Transparency Report <https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-22_2022-1-1_2022-6-30_en_v1.pdf>. (noting the YouTube paid USD 7.5 Billion of Ad revenue "to rightsholders as of December 2021 from content claimed and monetized through Content ID").
[17] YouTube (n 16) 1.
[18] YouTube (n 16) 3.
[19] YouTube (n 16) 3.
[20] YouTube (n 16) 3.
[21] For additional research on this topic, see also João Pedro Quintais, Giovanni De Gregorio and João Carlos Magalhães, 'How Platforms Govern Users' Copyright-Protected Content: Exploring the Power of Private Ordering and Its Implications [Forthcoming]' [2023] Computer Law & Security Review.

17 CDSM offer very little. As such, this is an area where serious consideration must be given to the potential application to OCSSPs and other copyright platforms of the DSA's transparency provisions, as well as to national solutions that impose on OCSSPs and non-OCSSPs transparency and data access obligations. As regards the DSA, the data access and scrutiny obligations vis-à-vis researchers are of particular importance. As regards national law solutions, in our view, the German transposition law provides an interesting blueprint in Section 19(3) UrhDaG in relation to rights to information.

2.3. COPYRIGHT CONTENT MODERATION RULES AT NATIONAL LEVEL[22]

Chapter 4 follows up on the EU level analysis with the comparative legal research at national level. It aims to answer SQR(2) and SQR(3) from the perspective of *selected national laws*. The findings are based on legal questionnaires carried out in two phases with national experts in ten Member States, the first before the due date for implementation of the CDSMD and the second after that date. This corresponds to our work on Task T.6.1.2 (Comparative National Level Mapping).

The key findings of the first phase questionnaire are as follows. First, the majority of the Member States has conceptualized service providers that store and give the public access to a large amount of protected content uploaded by their users; but the direct liability of such service providers was far from uniform in the Member States. E-Commerce, criminal and civil law concepts are alternatively or complementarily applied; and such liability is altogether missing in some countries. The new regime in art. 17 CDSMD will therefore require the introduction of new mechanisms in the majority of the Member States, as suggested by the Commission in its Guidance.

Second, the questionnaire indicated the need for the transformation of the liability regime of OCSSPs in the Member States' laws. So far injunctions, secondary liability, safe harbour and content moderation practices were mainly present in the analysed countries, unlike complaint-and-redress mechanisms, which were regulated only in a small number of Member

[22] This section summarizes the contents of Quintais and others (n 1) ch 4.

States. Art. 17 CDSMD will require the implementation of all of these elements, and hence Member States will be required to amend their legal system to a greater extent.

Third, the analysis highlighted that the end-users might be directly liable for unauthorized uploading of protected subject matter to OCSSPs systems, but such liability is rarely enforced in the Member States. Art. 17 CDSMD will also tend to push OCSSPs to authorize online users, and Member States' practices regarding end-user activities won't need to be amended significantly. On the other hand, several Member States will need to make more significant changes related to end-user flexibilities (especially parody, caricature and pastiche) and complaint-and-redress mechanisms. Similarly, based on the national respondents' reactions, it is conceivable that the "user rights" approach of the CDSMD might require a conceptual change in the way copyright laws qualify end-users and their entitlements in many Member States.

The key findings of the second phase questionnaire – taking place *after the implementation deadline for the CDSMD* – are as follows. The implementation of art. 17 CDSMD (or the related legislative proposals) took place in nine of the analysed Member States with important differences. A significant number of the elements of secondary importance of the new regime were almost uniformly transplanted. To the contrary, the implementation of the primary building blocks of art. 17, i.e., the economic rights affected; the new liability regime; or the balancing of fundamental rights of stakeholders show a diverse picture. Such diversity suggests that the initial goal of the CDSMD to harmonize certain aspects of copyright in the digital single market might not be met, leaving instead a fragmented legal landscape.

The nine Member States that had implemented the CDSMD at the time of our analysis can be grouped into three tiers. In **tier one,** the German and the Swedish models show above average detail in the implementation of the new regime, with a special focus on the strengthened protection of user rights and detailed liability mechanisms. In **tier two**, the Estonian, French and the Dutch legislation contain a smaller number of individual solutions. In **tier three**, Denmark, Hungary, Ireland and Italy took a rather restrictive approach through an almost verbatim transplantation of art. 17 CDSMD. Importantly, this three-tier system is not meant

to convey any qualitative ranking among the countries. It is likely that most national legislative institutions shall reconsider their domestic rules to make their laws fully compatible with the CJEU ruling in C-401/19 or with subject CJEU or national case law.

Our comparative research also flagged certain **conflicting statements in the Commission's and the CJEU's view on the proper method of implementation and substance of the national laws**, as noted the findings above, which are consequential for national implementations. The CJEU's judgment requires that Member States implement art. 17 CDSMD in a fundamental rights compliant manner. At the time of our analysis, various national solutions seem to be rather limited in terms of e.g., the priority of user rights over content filtering. Despite that, it is important to note that there is still no consensus on scholarship on the proper transposition method of art. 17, namely as regards the question of whether it is preferable to follow a (near) verbatim vs sophisticated (or "gold-plating") implementation of the provision. With that being said, if one considers the Commission's Guidance, the AG Opinion and the CJEU judgment in case C-401/19, there are strong arguments that national implementations must go some way beyond quasi-verbatim transpositions.[23]

Our findings indicate that it is plausible that a number of preliminary references on different aspects of art. 17 CDSMD will find their way to the CJEU in the short to medium term. These references will most probably focus on: interpretation of the newly introduced autonomous concepts of the CDSMD; the consistency of national transpositions with the EU law, especially in a fundamental rights dimension; and the exact scope and implications of "user rights" and respective safeguards under art. 17(7) and (9).

These findings remained valid since the analysis of the ten selected Member States' transposition practices were closed. Following that analysis and until closing of this report,

[23] On this point, see also Christina Angelopoulos, 'Articles 15 & 17 of the Directive on Copyright in the Digital Single Market Comparative National Implementation Report' (Coalition for Creativity (C4C); CIPIL 2022) <https://informationlabs.org/copyright/> accessed 15 December 2022. (published after our mapping analysis).

multiple other Member States (but not all) have implemented Article 17 CDMSD. These domestic variations show differences in the key components of the new liability regime.[24]

2.4. PRIVATE REGULATION BY PLATFORMS: EMPIRICAL RESEARCH

Finally, the empirical component of our mapping analysis focused on the following sub-research questions:

- **SQR(4):** How are copyright content moderation rules organized by platforms into public documents?
- **SQR(5):** Which copyright content moderation rules do different platforms employ to regulate copyright, and how have they changed over time?
- **SQR(6):** How do platforms' automated copyright content moderation systems work?

In this context, we studied the copyright content moderation structures adopted by 15 social media platforms over time, with a focus on their T&Cs (rules) and automated systems. These platforms are grouped into (i) mainstream –Facebook, YouTube, Twitter, Instagram and Sound Cloud; (ii) *alternative* – Diaspora, DTube, Mastodon, Pixelfed, Audius; and (iii) *specialised* - Twitch, Vimeo, FanFiction, Dribble and Pornhub.This corresponds to the empirical research carried out in the context of Task T.6.1.3 (Private Regulations by Platforms: ToS, Community Guidelines).

Our analysis suggests that two dual processes seem to explain these structures' development. The first is *complexification/opacification*. Our empirical work indicates that virtually all 15 platforms' T&Cs have become more intricate, in various ways and to different extents. Over time, more (kinds of) rules were introduced or made public, and these rules were communicated in increasingly more diverse sets of documents. These documents were changed and tweaked several times, producing sometimes a plethora of versions, often located in a dense web of URLs. We therefore conclude that the way platforms organize, articulate and present their T&Cs matters greatly. For one, under increasing public and policy

[24] Readers might track the implementation process via CREATe's resource page developed in partnership with the reCreating Europe project, available at https://www.create.ac.uk/cdsm-implementation-resource-page/ , as well as the COMMUNIA DSM Implementation tracker , available at https://www.notion.so/DSM-Directive-Implementation-Tracker-361cfae48e814440b353b32692bba879.

pressure, platforms have felt the need to express and explain their practices and rules of operation, and they have done so with complex and greatly varying documentation. For observers, although this provides more information about platforms, it nevertheless makes understanding the trajectory of platforms and their T&Cs extremely challenging. For example, with YouTube as a major actor when it comes to copyright, our database of their highly fragmented T&Cs has not resulted robust enough to allow for a precise longitudinal examination of their rules. In that way, the very organization and presentations of T&Cs should be understood as one element of platforms' governance of content.

Substantially, we demonstrated that complexification can be radically distinct, depending on which platforms one considers. Very large ones, such as Meta/Facebook, experienced an almost continuous and drastic transformation; smaller ones, such as Diaspora, have barely changed. Yet, when a change occurred, it made those sets of rules more difficult to comprehend. Whilst our analysis did not take a longitudinal take on automated copyright content moderation systems, their emergence and eventual transformation into a central governance tool for various platforms is, in itself, an important element of broader complexification processes. These systems work at a scale that is hard to comprehend, through computational operations that are technically intricate, and under largely unjustified and seemingly arbitrary protocols on, e.g., how to appeal decisions. In other words, they are remarkably opaque, as so many of the T&Cs we studied. Our analysis pointed out that while in some cases some complexification might be impossible to avoid, opacification is by no means necessary or necessarily justifiable. From this perspective, then, the imposition by law of rules on platforms' internal content moderation procedures and their transparency is sensible and should prove beneficial. It will be critical to ensure that these reporting obligations are rolled out in robust and detailed ways, so that they are instrumental to the clarification and understanding of such procedures and related decision making.

The second process is *platformisation/concentration*. By categorizing rules into what we termed "normative types", the chapter argued that various platforms in our sample altered their rules so as to give themselves more power over copyright content moderation, usually

by increasing the number of their obligations and rights, which were, in turn, largely aligned with their own interests, logics and technologies. We suggested that this could be interpreted as a particular example of the broader phenomenon of "platformisation".[25] Nonetheless, our analysis argued that this transformation was by no means unidirectional. For platformisation enhances not only platforms' power but also their responsibilities over content moderation. It was curious to note, therefore, that while emboldening their normative legitimacy to control copyright, platforms did not necessarily alter their discursive focus on users-oriented rules. As with complexification, platformisation has been experienced differently by different platforms and deepened by the rise of automated copyright content moderation systems, which may severely impair ordinary users' ability to participate in and challenge removal decisions. That platformisation centralises power in the hands of platforms might be a truism – but our research also suggests that this process might end up giving more power to large rightsholders, to the detriment of essentially smaller rightsholders and (users-)creators, as well as other users.[26] Nowhere this was clearer than in our study of Meta/Facebook's Rights Manager, which does not appear to be accessible for small creators, for instance, a non-algorithmic bottleneck that has been rarely studied from an empirical perspective.[27]

[25] See e.g. Thomas Poell, David Nieborg and José van Dijck, 'Platformisation' (2019) 8 Internet Policy Review <https://policyreview.info/concepts/platformisation> accessed 18 February 2022; José van Dijck, Thomas Poell and Martijn de Waal, *The Platform Society* (Oxford University Press 2018) <https://oxford.universitypressscholarship.com/10.1093/oso/9780190889760.001.0001/oso-9780190889760> accessed 20 February 2022.

[26] Making a similar argument in relation to platforms' control of users' copyright-protected content and its monetisation, see Quintais, Gregorio and Magalhães (n 21).

[27] See also Quintais, Gregorio and Magalhães (n 21). suggesting similar problems with ContentID, based on data from YouTube's Copyright Transparency Reports.

3. EVALUATION AND MEASURING: IMPACT OF MODERATION PRACTICES AND TECHNOLOGIES ON ACCESS AND DIVERSITY[28]

Building on our mapping work, the evaluative part of the analysis centres on a normative examination of the existing public and private legal frameworks with regard to intermediaries and cultural diversity, and on the actual impact on intermediaries' content moderation on diversity.[29] The evaluation analysis pursues two main objectives.

- To *explain and evaluate the existing legal frameworks* (both public and private, existing and proposed) that shape the role of intermediaries in organising the circulation of culture and creative works in Europe, including in copyright content moderation.
- *To explain, critically examine and evaluate the existing practices and technologies* that intermediaries deploy to organise the circulation of culture and creative works in Europe, including in copyright content moderation.

Each objective corresponds to two main components of our analysis.

The *first main component* deals with the evaluation of legal frameworks on the different levels. In this context, we first expand on the assessment of regulatory environment and revisit the starting point for access to culture and the creation of cultural value. In doing so, we introduce a concept of "Rough Justice", which acknowledges the difficulties and differences vis-à-vis a full "fair trial" setup and proposes conceptualization in the context of procedural rules, substantive rules and competences. A second starting point for the legal evaluation is provided in analysing and evaluating the framework for *quality* of automated copyright content moderation as put forward in the CDSMD and the DSA in light of erroneous decisions. It is suggested that decision quality should be a decisive factor that is to be seen as a separate from ex post mitigation mechanisms. We also examine the benchmark put forward in the sector-specific CDSMD and the horizontal DSA. A third perspective relates to the

[28] Making a similar argument in relation to platforms' control of users' copyright-protected content and its monetisation, see Quintais, Gregorio and Magalhães (n 21).

[29] In lieu of a comprehensive report, like that carried out for our mapping analysis, the evaluation analysis is based on a series of draft articles based on our research, which are attached to the Evaluation Report. These draft articles are identified at the start of the corresponding sub-sections below.

realisation that copyright content moderation increasingly requires an understanding of contextual use and whether the potential risk of "bias carry-over" from datasets to content moderation is sufficiently addressed in the current framework. It is suggested that the question of bias mitigation and access to copyright data should increasingly be addressed.

The *second main component* of our analysis is an attempt to measure the impact of copyright content moderation on access and diversity. We start by presenting existing research in the field and by discussing options to investigate these complex questions. On these grounds, we explain our research design consisting of three empirical sub-studies, and then present the results of this work. In the first sub study we investigate aggregated data on copyright and content moderation published by platforms themselves, often in the form of transparency reports; secondly, we analyse content level data with regard to the sustaining availability and the diversity of content on social media platforms; and thirdly we present results from in-depth interviews with cultural creators with regard to their experiences with copyright content moderation. Overall, the results indicate a strong impact of copyright regulation and content moderation on diversity, and potentially an impact that leads to a decrease in diversity of content. Yet, the research has also shown that these interpretations cannot be fully verified based on the limited data that is available to researchers and the public.

The following subsections provide a summary of the findings and conclusions of each of these main components of our evaluation analysis, namely the evaluation of the existing legal frameworks (3.1) and of existing practices and technologies (3.2). A common theme we highlight and return to in our recommendations is the need for further research on issues of diversity and access on social media platforms, given its high relevance for European societies, and at the same time its complex nature, specifically in the context of contemporary fragmented media landscapes. Consequently, we conclude with a strong call for robust mandatory data access clauses in future regulations.

3.1. EVALUATING MULTI-LEVEL LEGAL FRAMEWORKS[30]

The evaluation of legal frameworks we have carried out involves a normative assessment of how legal rules and contractual terms on the moderation and removal of copyright content on large-scale user-generated content (**UGC**) platforms affect digital access to culture and the creation of cultural value. We assess how such rules and terms shape the design of removal and moderation by UGC platforms, the activities of creators and users, and the role of fundamental rights and freedoms – namely the freedom of expression, freedom of the arts and freedom to conduct a business – in shaping these rules and terms. It also evaluates how the state-enacted rules in the DSM shape the emergence of private models for content moderation and removal, examining how the production of law is shaped by the intrinsic characteristics and needs of the actors on the DSM within the legal framework conditions. Our research shows that the existing legal framework has increasingly focused on how it shapes the role of intermediaries in organising the circulation of culture and creative works in Europe, including copyright content moderation.

3.1.1. OVERLAPS AND INTERPLAY OF EXISTING LEGAL FRAMEWORKS

The *assessment of the existing legal frameworks* that shape the role of online platforms in organising the circulation of culture and creative works in Europe through content moderation has shown the complex landscape of interacting rules in this field.

For instance, the relevant substantive copyright rules are contained in national copyright legislation, partly based on harmonising instruments such as the InfoSoc Directive. The relevant rules regarding intermediary or platform regulation, are contained in art. 17 CDSMD (and its national implementations), the e-Commerce Directive's framework for intermediary liability exemptions in arts. 12-15, replaced and amended by the DSA.

[30] This section of the report is based on D.6.3. Final Evaluation and Measuring Report - impact of moderation practices and technologies on access and diversity" (2023).

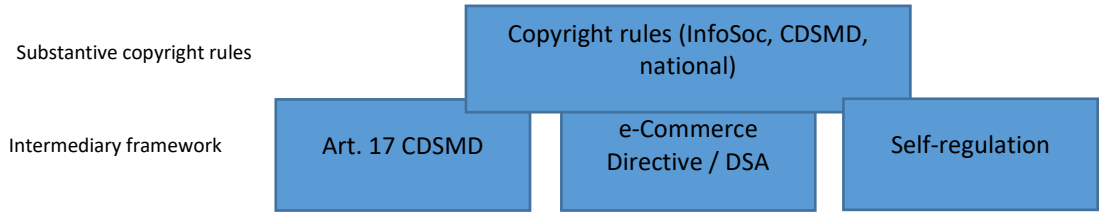

Figure 1. Relationship substantive copyright rules and intermediary framework

In order to understand the regulatory, i.e., both law and self-regulatory, environment surrounding the moderation of online content, it is necessary to recall that art. 14 e-Commerce Directive sets forth the horizontal basic rules for an intermediary's mandated response to illegal content, including copyright- infringing works. These rules are now replaced by the corresponding provision in the DSA.[31]

Notably, the e-Commerce Directive refrained from further specifying the notice-and-action regime. In this void (or more positively: freedom of operation) industry-practices have merged. These, in, turn, now appear to at least partly codified in arts. 17 CDSMD with regards to OCSSPs, and in the DSA with regards to other online platforms (or non-copyright services of the same platforms) that fall outside the scope of art. 17 CDSMD.[32]

Figure 2. Relationship between rules on intermediaries and industry practices

[31] See for an in-depth comparison Sebastian Felix Schwemer, 'Digital Services Act: A Reform of the E-Commerce Directive and Much More' in A Savin (ed), *Research Handbook of EU Internet Law [Forthcoming]* (Edward Elgar 2022).

[32] And complementarily to those falling within its scope for matters not dealt with in Art. 17 CDSMD.

One issue related to the regulatory framework regards its complexity and potential overlaps and interplay. This is specifically relevant in the context of online platforms and copyright, where both art. 17 CDSMD and the DSA specify and adjust platforms' room of operation for content moderation and which we have previously explored.[33] Further complexity is added with the specific national implementations of art. 17 CDSMD as previously analysed.[34]

Besides this overlap, there are notable other areas where rules interact. Since content moderation often also involves the processing of personal data, for example, future research should look into the interplay between the General Data Protection Regulation (GDPR) and the sector specific CDSMD framework as well as the horizontal rules in the DSA. Since content moderation is – as explored earlier – regularly performed or supported by algorithmic means, furthermore, also the potential intersection with the Artificial Intelligence Act (**AIA**)[35], a Regulation which was proposed on 21 April 2021, is of interest.[36] The AIA introduces "rules regulating the placing on the market and putting into service of certain AI systems"[37] and focusses on the regulation of the provider as well as the user of such AI system. In the context of copyright content moderation, the AIA is of interest given the broad and generic definition of AI system in Art. 3(1) AIA, which means "software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with".[38] In our view, content moderation

[33] Quintais, J., & Schwemer, S. (2022). The Interplay between the Digital Services Act and Sector Regulation: How Special Is Copyright? *European Journal of Risk Regulation, 13*(2), 191-217. doi:10.1017/err.2022.1

[34] See supra at 2.2.

[35] European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, COM/2021/206 final.

[36] See Thomas Margoni, João Pedro Quintais and Sebastian Felix Schwemer, 'Algorithmic Propagation: Do Property Rights in Data Increase Bias in Content Moderation? – Part II' (*Kluwer Copyright Blog*, 9 June 2022) <https://copyrightblog.kluweriplaw.com/2022/06/09/algorithmic-propagation-do-property-rights-in-data-increase-bias-in-content-moderation-part-ii/> accessed 24 January 2023. See also generally on the topic Philipp Hacker, Andreas Engel and Theresa List, 'Understanding and Regulating ChatGPT, and Other Large Generative AI Models: With input from ChatGPT' [2023] Verfassungsblog <https://verfassungsblog.de/chatgpt/> accessed 24 January 2023.

[37] Recital 4 AIA proposal.

[38] This definition, is complemented by Annex I, which contains a detailed list of approaches and techniques for the development of AI.

technology likely falls within the scope of this definition. Furthermore, the scope the proposed Regulation focuses on risks inter alia to the protection of fundamental rights of natural persons concerned.[39] Copyright content moderation might come with risks for inter alia freedom of expression or the arts. The AIA differentiates between four types of risk: AI systems that come with unacceptable risks are prohibited; AI systems with high-risk are permitted but subject to specific obligations; AI systems with limited risk are subject to certain transparency obligations. Neither, however, seems to encompass copyright content moderation at this stage.

3.1.2. BENCHMARKS FOR NORMATIVE ASSESSMENT: "ROUGH JUSTICE" AND "QUALITY"[40]

3.1.2.1. A MODEL OF ROUGH JUSTICE FOR CONTENT MODERATION

Our research attempts to develop a model that can be used to say something meaningful about the quality of the legal framework that shapes the actual content moderation practices. It tries to evaluate the legal framework for the purpose of posing normative statements on how to improve the legal framework. In order to do that a value-based measuring scale is needed. Common values in rights-enforcement and human rights can be used in such a measuring scale. One place to look for common values is in the traditional legal perception of fair trial that includes values such as predictability, contradiction, production and presentation of evidence etc. However, in relation to platforms' content moderation practices, for all practical purposes, it is not possible to ensure the relatively high level of due process known from traditional civil procedure.

[39] see, e.g., recitals 1, 13, 27, 32, Arts. 7(1)(b), 65 AIA proposal.

[40] This section of the report is based on D.6.3. Final Evaluation and Measuring Report - impact of moderation practices and technologies on access and diversity" (2023), and in particular in the following working papers attached to that report: Thomas Riis: "A theory of rough justice for internet intermediaries from the perspective of EU copyright law" (forthcoming 2023); Sebastian Felix Schwemer: "Quality of Automated Content Moderation: Regulatory Routes for Mitigating Errors" (forthcoming 2023).

The level of justice in traditional civil procedure cannot be adopted one-t-one to platforms' content moderation practices because it will simply be too burdensome and resource intensive. Therefore, there is a need to modify the traditional conception of justice in the context of internet platforms. In our research, such a modification is called "rough justice". A model of rough justice does not presume to provide full justice but is significantly better than no justice.[41] In short, rather than envisioning a private-regulatory copy of a "full trial" setup, a different conceptual approach of "rough justice" is suggested to address the copyright content moderation by platforms.

Departing from the human rights concept of fair trial and the fundamental right to an effective remedy (art. 47 Charter), we argue that a conception of rough justice on internet platforms must address two major issues and be guided by three general objectives. The two issues are: (1) the accuracy of moderation practices as regards content that is illegal or contrary to a platform's T&Cs; (2) the inherent privatization of justice, which results from enforcement of rights being left to a private party with a risk of distortion of the balancing of interests in substantive law.[42] The three general objectives, which rights-enforcement systems must consider are: (1) "efficacy", in the sense of effective and affordable access to justice; (2) "fair trial", meaning consistency, predictability and proportionality in rights-enforcement; and (3) balanced use of resources, including costs of enforcement.

In searching for the operationalization of these principles, our research critically examines not only the relevant provisions in the DSA but also in three important attempts to establish codes or guidelines for fair trial on the internet, namely: (1) The Santa Clara Principles 2.0; (2) The Aequitas Principles on Online Due Process; and (3) The Council of Europe's recommendation on the roles and responsibilities of internet intermediaries (CM/Rec(2018)2). Our critical examination is made on the basis of the human rights approached to justice with a specific

[41] Peter Linzer, Rough Justice: A Theory of Restitution and Reliance, Contracts and Torts, 2001 Wis. L. REV. 695-775 (2001), p. 766.

[42] We argue that privatization of justice is problematic insofar as private parties substitute public rules with private rules. Whereas public rules pursue societal objectives and values, private rules must be assumed to pursue private objectives and values.

view to: (1) a substantial human rights norm to prevent over-enforcement; (2) Transparency; and (3) Fair trial.

On this basis, we develop a model on rough justice divided into three different parts, including associated recommendations: (1) Procedural rules, (2) Substantive rules and (3) Competences.

In respect of **(1) procedural rules**, we argue that there is a need for more transparency into how content moderation works, as this will improve the explainability of decision making, error and bias correction, and quality assurance. Transparency should cover the functioning of algorithms and the logic behind and working conditions of human moderators involved, if any. In our view, in light of potential trade secrecy protection of many of these aspects, achieving meaningful transparency require legislative intervention that exempts algorithms for content moderation from trade secrets protection.

As for **(2) substantive rules**, their purpose is to create a counter-weight to online platforms' tendency to over-enforce and to reduce moderation of content that is legal but incompatible with T&Cs. In our view, substantial rules based on human rights would be an important means to align platforms' T&Cs with societal objectives and public values, thereby counteracting the adverse effect of privatization of justice. International human rights law is binding on states only, not on individuals or companies. Therefore, it is recommended that an obligation to fully respect human rights are imposed on platforms, for instance by making international human rights directly applicable to platforms that moderate content. As some authors argue, the DSA may already go some way in this direction with its provision on T&Cs in Article 14.[43]

Finally, as regards **(3) competences** of human moderators, we note that such human competences directly impact the quality of the content moderation system. This much is recognized in the DSA, CDSMD and the codes we reviewed, which require human review in the appeal process, partly as a means to mitigate the risks of automated content moderation. From our viewpoint, a certain level of human involvement should also be required to reduce

[43] Quintais, Appelman and Fahy (n 3).

biases and errors and ensure accuracy in the first stage of automated moderation. One way to achieve this would be to mandate random tests of accuracy by human intervention. Furthermore, human competences must be ensured by adequate training and working conditions. More important than setting up precise standards for qualifications and working conditions, is to impose an obligation on platforms to inform on the internal criteria for appropriate qualifications and working conditions (transparency), so the users of the platform themselves are able to assess the legitimacy of the content moderation process.

3.1.2.2. QUALITY OF COPYRIGHT CONTENT MODERATION

In addition to developing a model of "rough justice", our analysis shows that with regards to access to culture and cultural diversity, decision *quality* should be emphasised as a separate factor from ex post mitigation mechanisms. Both the DSA and the CDSMD (including case law) provide starting points for this. The analysis also points to the fact that content moderation increasingly requires an understanding of contextual use but further work is needed on the potential risk of "bias carry-over" from datasets to content moderation. In this context, it is also worthwhile to point out that content moderation technology appears to be a blind spot in the AI Act proposal and legislative process.

Our departure point in this part of the analysis is that underlying most copyright content moderation scenarios there is a binary choice between whether the uploaded content is illegal (i.e. copyright infringement) or not. Whereas in some instances the decision is straightforward (e.g. for "manifestly illegal" content), on others it is not, as it might require detailed assessment by domain experts or even courts. In any case, we should be able to assess the "quality" of such content moderation decision. But what is the right benchmark for such assessment?

In our view, the "quality" of copyright content moderation is correlated to access to culture, because access to culture (as per our definition) is considered embedded in the existing copyright framework. Since the existing framework is assumed to strike the appropriate balance between exclusivity in copyright protection and access to culture, any variation in that balance – beyond the margin of interpretation allowed by law – will impact on access to

culture. Consequently, both excessive and insufficient content moderation will have a negative impact on access to culture. The consequence of this assumption is that the "quality" of content moderation can in simple terms be described in terms of correct and false results. The first set of outcomes that relates to correct result of content moderation (i.e., the absence of error). The second set of outcomes relates to false results of content moderation (i.e., the presence of error).

In this light, the principal question that arises is what error rate is acceptable under the legislative framework. After examining different explicit and references to content moderation error rates in the DSA and art. 17 CDSMD (including interpretations in the Commission's Guidance, the AG Opinion and CJEU judgment on Case C-401/19), we conclude that the issue of error rates in all these above examples can only consist of a contextual analysis. A first factor should relate to the volume of content moderation decisions taken. The goal cannot only be to have a low percentage of error (error rate) but rather a low number of actual "wrong" content moderation decisions. A second factor should relate to the "harm" caused by the wrong decision (and whether such harm can be mitigated ex-post).

3.1.3. LOOKING INTO THE FUTURE: CONTEXT AND BIAS IN CONTENT MODERATION[44]

In addition to our benchmarks for normative assessment, our project took the initial steps into the examination of the issue of bias in copyright content moderation. In simple terms, it can be stated that art. 17 CDSMD incentivizes OCSSPs to preventively filter content uploaded by users to comply with their "best efforts" obligations to deploy preventive measures against infringing content. Prior to the introduction of this legal regime, some platforms already "voluntarily" relied on similar automated content moderation (e.g., YouTube's ContentID and Facebook's Rights Manager). At the current state of technology, filtering appears to be done

[44] This section of the report is based on D.6.3. Final Evaluation and Measuring Report - impact of moderation practices and technologies on access and diversity" (2023), and in particular in the following working papers attached to that report: Thomas Margoni, João Pedro Quintais and Sebastian Felix Schwemer: "Algorithmic propagation: do property rights in data increase bias in content moderation?" [forthcoming 2023].

mainly through matching and fingerprinting. However, it is also now well-known that these tools are incapable of assessing contextual uses. Therefore, they are not suitable to ensure the required protection of freedom of expression-based exceptions like parody, criticism and review, as required by art. 17(7) CDSMD. Accordingly, more sophisticated tools seem necessary to enable preventive measures while respecting users' rights and freedoms, as recently confirmed by the CJEU in case C-401/19. This suggests that machine learning algorithms may increasingly be employed for copyright content moderation given their alleged superiority in identifying (understanding?) contextual uses.

Against this background, a crucial question emerges for the future of (copyright) online content moderation and fundamental rights in the EU: what happens when these tools are based on "biased" datasets? More specifically, if it is plausible that any bias, errors or inaccuracies present in the original datasets be carried over in some form onto the filtering tools developed on those data: (1) How do property rights in data influence this "bias carry-over effect"? and (2) what measure (transparency, verifiability, replicability, etc.) can and should be adopted to mitigate this undesirable effect in copyright content moderation in order to ensure an effective protection of fundamental rights?

Based on this, we explore the possible links between conditional data access regimes and content moderation performed through data-intensive technologies such as fingerprinting and, within the realm of AI in general, and machine learning algorithms in particular. More specifically, we look at whether current EU copyright rules may have the effect of favouring the propagation of bias present in input data to the algorithmic tools employed for content moderation and what kind of measures could be adopted to mitigate this effect. Algorithmic content moderation is a powerful tool that may contribute to a fairer use of copyright material online. However, it may also embed most of the bias, errors and inaccuracies that characterize the information it has been trained on. Therefore, if the user rights contained in art. 17(7) CDSMD are to be given an effective protection, simply indicating the expected results but omitting *how* to reach them may not be sufficient. The problem of over-blocking is not simply a technical or technological issue. It is a cultural, social and economic issue, as well and,

perhaps more than anything, it is a power dynamic issue. Recognizing parody, criticisms and review as "user rights", as the CJEU does in C-401/19, may be a first step towards the strengthening of users' prerogatives. But the road to reach a situation of power symmetry with platforms and right holders seems a long one. Ensuring that bias and errors concealed in technological opacity do not circumvent such recognition and render art. 17(7) ineffective in practice would be a logical second step.

Even though content recommendation is outside the core of our research, we note that these questions are also of high relevance there.

3.2. MEASURING THE IMPACT OF MODERATION PRACTICES AND TECHNOLOGIES ON ACCESS AND DIVERSITY

During the course of our project and indeed the implementation period of the CDSMD it has become clear that online platforms play a crucial role in contemporary societies, whilst AI technologies are increasingly presented as solutions to the major societal problems. Under increasing public and political pressure, social media platforms have expanded their efforts to moderate content they host. To do so, they have invested both in growing numbers of human moderators[45] and in algorithmic moderation.

The empirical component of our research attempts to gauge the impact of increasing content moderation practices, policies, and technologies, including for copyright, and of the CSDMD on access to culture and diversity. In this regard, both legal and social science research have identified such legislative and practical developments as relevant for the future role of platforms as intermediaries and their impact on cultural diversity and access to culture. From the legal perspective, as we explain above, art. 17 CDSMD poses serious concerns as regards the freedom of expression implications of preventive filtering and over-blocking. This concerned is amplified by the lack of transparency surrounding private platforms' algorithmic

[45] Note, however, the impact of the COVID-19 pandemic on human resources in content moderation, which led (temporarily) to increased reliance on purely algorithmic moderation decisions.

moderation systems. This raises the stakes for understanding better how platforms and copyright content moderation impact diversity and access to culture in the DSM.

This empirical part of the project tackles three dimensions of this problem. First, we have investigated all the aggregated data on copyright moderation provided by the platforms themselves (3.2.1). Second, we have analysed content level data of platforms with regard to changes and factors of cultural diversity on social media and streaming platforms, specifically YouTube (3.2.2). Third, we have explored creators' understanding and experiences of copyright moderation in relation to their creative work and the labour of media production on social media platforms (3.2.3).

Before highlighting the findings of our research along these three dimensions, it is important to briefly clarify our **empirical research design**. Building on existing research on diversity, content moderation and algorithms, we examined different options to assess the impact of copyright regulation and content moderation on diversity and access to culture. Unfortunately, the most the adequate option was no longer viable when starting the empirical work. Gray and Suzor (2020) had assessed the life-circle of content on YouTube by tapping into YouTube's API, collecting a random sample of content directly at the moment of upload. In defined periods later, they checked for the availability of those content items.[46] YouTube's APIv2 at that same provided information about reasons if content was no longer available. On these grounds, researchers could in fact evaluate the actual scope and effects of copyright content moderation. Unfortunately, due to restrictions of access to data in current YouTube's API v.3, this option no longer exists.

Against this background we have developed a research design that technically circles around the key question at hand, taking three different approaches investigating data on different levels:

[46] Joanne E Gray and Nicolas Suzor, 'Playing with Machines: Using Machine Learning to Understand Automated Copyright Enforcement at Scale' (2020) 7 Big Data & Society <https://journals.sagepub.com/doi/full/10.1177/2053951720919963> accessed 25 January 2023.

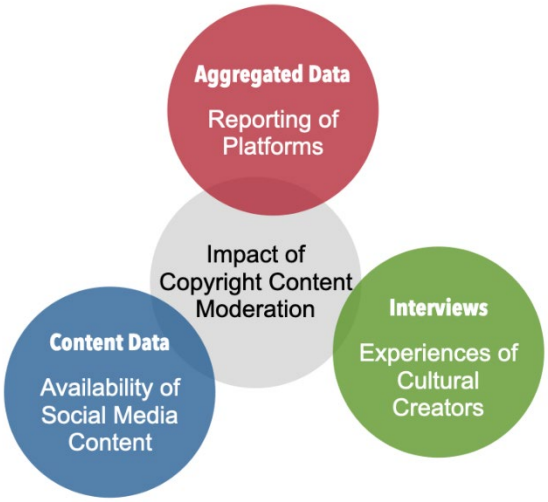

Figure 3. Empirical Research Design (Representation)

We have investigated the *aggregated data* on copyright and content moderation that platforms themselves publish; we have analysed *content level* data with regard to the sustaining availability and the diversity of content on social media platforms; and we have *interviewed cultural creators* with regard to their experiences with copyright content moderation.

In the first step, we have compared *aggregated data from transparency reports* published by major platforms present in the EU. In this sub-study we have analysed both the kinds of data platforms have started to disclose in the recent years, as well as the substantial numbers on copyright content moderation.

In the second step, we have analysed on the *content level* availability and diversity of content on a selected platform, YouTube. For a timeframe from 2019 (before CDSMD) and 2022, we have analysed both the scale of copyright-based content deletion and blocking, as well as measured differences in the diversity of content available on the platform across time and selected countries.

Further on we collected samples of channels from all the four countries, and their descriptions, in order to compare the changes in diversity supply that happened from 2019 till 2022.

Finally, we conducted semi-structured interviews with creators on various platforms: the sample was derived from those taking part in the survey on digitalization of creative work from the same project ReCreating Europe.[47]

3.2.1. ASSESSING TRANSPARENCY REPORTS[48]

In this first study we investigate the historical evolution and current situation of transparency reporting with a focus on copyright-based content moderation. We further examine the convergence and divergence in social media platforms' content moderation practices along with the transparency habits in a broader sense also by elaborating on substantial numbers of content moderation data.

Our analysis highlights that transparency reporting has a number of important limitations that potentially jeopardize platforms' perceived accountability and positive effects of the reporting on their legitimacy in the eyes of external stakeholders. First, as noted by other scholars "aggregated data in transparency reports only shows the platforms' own assessments, and not the merits of the underlying cases [and] researchers cannot evaluate the accuracy of takedown decisions or spot any trends of inconsistent enforcement".[49] Additional limitations of transparency reports in their current form are that they largely focus on the removal of content (and accounts) rather than other (often called "softer") forms of moderation. More recently, practices described as "shadow banning" have taken hold on

[47] Joost Poort and Abeer Pervaiz, 'D3.2/3.3 Report(s) on the Perspectives of Authors and Performers' (Institute for Information Law (IViR) 2022) reCreating Europe Reports <https://zenodo.org/record/6779373> accessed 25 January 2023.

[48] This section of the report is based on D.6.3. Final Evaluation and Measuring Report - impact of moderation practices and technologies on access and diversity" (2023), and in particular in the following working paper attached to that report: Christian Katzenbach, Selim Basoglu and Dennis Redeker: "Finally Opening up? The evolution of transparency reporting practices of social media platforms" (forthcoming 2023).

[49] Daphne Keller and Paddy Leerssen, 'Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation', *Social Media and Democracy: The State of the Field and Prospects for Reform* (Cambridge University Press 2019) 228 <https://papers.ssrn.com/abstract=3504930> accessed 4 May 2021.

platforms.[50] Users' content is not outright deleted but instead merely not shown to wider audiences, effectively stymying free expression.[51] Due to the lack of notice of users and their resulting inability to dispute such a moderation measure, shadow banning or the related downranking of content are controversial issues. Even the extent of such "softer" practices is still relatively opaque as "platforms like Instagram, Twitter and TikTok vehemently deny the existence of the practice".[52] Shadow banning is likely less relevant for copyright-based moderation, since there are more categorical issues when intellectual property is being reproduced without permission. In general, the lack of information on how moderation algorithms work is a shortcoming for platform transparency, with platforms often engaging in "black box gaslighting" to deflect critique.[53] All in all, our research in this topic shows that there is still significant room for improvement of platform transparency practices, as there is for their moderation practices. Better quality and potentially a standardisation of transparency practices by platforms would be crucial for a better understanding and assessment of their copyright content moderation and, as a result, for evidenced-based policy making in this area.

3.2.2. MEASURING CONTENT BLOCKING AND DELETION ON PLATFORMS, AND ITS IMPACT ON DIVERSITY[54]

In addition to the screening of aggregated data in transparency reports, this second part of the empirical assessment has sought to find evidence about the impact of copyright content moderation on the *content level of social media platforms*. How does copyright content

[50] On the concept of shadow banning, see e.g. Paddy Leerssen, 'An End to Shadow Banning? Transparency Rights in the Digital Services Act between Content Moderation and Curation' <https://osf.io/7jg45/> accessed 23 November 2022.
[51] Laura Savolainen, 'The Shadow Banning Controversy: Perceived Governance and Algorithmic Folklore' (2022) 44 Media, Culture & Society 1091.
[52] Savolainen (n 51) 1092.
[53] Kelley Cotter, '"Shadowbanning Is Not a Thing": Black Box Gaslighting and the Power to Independently Know and Credibly Critique Algorithms' (2021) 0 Information, Communication & Society 1.
[54] This section of the report is based on D.6.3. Final Evaluation and Measuring Report - impact of moderation practices and technologies on access and diversity" (2023), and in particular in the following working paper attached to that report: Daria Dergacheva, Christian Katzenbach: "Mandate to Overblock? Understanding the impact of EU's Art. 17 on automated content moderation on YouTube" [forthcoming 2023].

moderation impact on the availability of content and its diversity? While a systematic study on the diversity of content circulating of social media platforms is already challenging, pinning down and isolating the impact of copyright regulation and content moderation is ambitious.

Against this background, this empirical study investigates the changes and influences in access and cultural diversity on social media and streaming platforms, specifically YouTube, in the timeframe 2019 to 2022, focusing on the period between the approval of the CDSMD and the end of 2022, where most national implementation laws in Member States have just been passed or are still in the final stages of discussion.

Our results consist of two parts. The first part presents general findings on the copyright takedowns on YouTube in the EU after 2019. The second part measures the diversity of content available on the platform in selected four countries of the EU in 2019 vis-à-vis 2022. For measuring diversity we use the diversity index developed by Stirling and adapted by the UN. Countries were selected depending on specifics of their national copyright regime and the CDSMD. As such, they function as proxies for the impact of copyright regulation in this area. In this data, we investigate if there were any changes in content supply diversity during that time and whether it varies by the countries in the sample.

Summing up this data-driven investigation of content blocking and content availability on YouTube with a focus on content diversity, it is possible to offer three main conclusions.

First, we found a *high share of blocked and deleted content* in our sample. While previous research has identified a share of roughly 1%, our sample identified a share of 3.8%. Due to restricted access to data, though, it is difficult to really pin down and isolate the exact reasons for content deletion and take-down. These 3.8% might include other types of content deletion and blocking, although we have applied the measures available to clean the data.

Second, we have found a general *decrease of diversity* with regard to available content. Within the four countries under study (Estonia, France, Germany, and Ireland), three countries display a noticeable decrease in the diversity index, with Ireland representing a contrary development with a light increase. The country differences do not correlate, though,

with national differences in copyright regulation and specifically with the variation in substance and timing of the national implementation of the CDSMD. This makes it hard to assess and isolate the actual impact of copyright content moderation and the implementation of the CDSMD on content diversity. Is the general decrease of diversity a result of the (then forthcoming) national implementation of the CDSMD? Or rather the product of changing monetization strategies of media companies, shifting media usage routines, or YouTube's algorithmic systems? Some of these research limitations concern the timeline of the study: actual national implementation of art. 17 CDSMD is not yet fully in place in the countries under study, and it is possible that we could not yet see its full-scale influence. At the very least, it will take some time post-implementation to assess its effects, namely as regards judicial practice and behaviour of private parties (e.g. platforms and users). Future and continuing research is needed to assess these questions, when the policy implementations become effective and visible at full scale.

Third, and most important from our perspective, we have been confronted with the limitations of research in this space due to lack of data access. In the current landscape, it results close to impossible to systematically study the questions posed in this project. What is the impact of copyright regulation and content moderation on content diversity? In fact, this research is not only highly limited, but also dependent on internal decisions of platforms on giving access to (different types of) data. This is a common refrain also for our legal research. Hence, there is urgent need for more robust rules on data access for researchers. Mandatory data access clauses such as those included in the German NetzDG, the German CDSMD implementation as well as in the DSA pave an important avenue in this regard. Yet, it remains to be seen how robust and effective these clauses are, since they demand highest levels of data security and infrastructure facilities on the side of researchers and their institutions. Finding practical and fair solutions as well as best practices for data access that are not only accessible to researchers at elite and perfectly-equipped institutions is a key challenge for policy and research in the next decade.

3.2.3. SOCIAL MEDIA CREATORS' PERSPECTIVE ON COPYRIGHT CONTENT MODERATION IN THE EU[55]

In the third sub-study, we have taken another angle at understanding copyright content moderation – understanding the experiences of cultural creators who share their work primarily on social media platforms. As social media creators and users in the EU may see a rise in algorithmic copyright moderation after implementation of art. 17 CDSMD, we focus this sub-study on creators' understanding and experiences of copyright moderation in relation to their creative work and the labour of media production on social media platforms. To what extent does copyright moderation on the former influence the creations that are posted there? What about the changes to one's creative process? In order to answer these questions, we have interviewed creators with regard to their experiences and descriptions of their interaction with copyright moderation and algorithms. This allows us to better understand the changes and influences that automated copyright moderation brings to creative work.

Cultural creators mainly seeking audiences online are strongly dependent on social media platforms. They have to constantly be involved in pursue of algorithmic visibility as measured by quantified metrics such as likes, views, and shares.[56] At the same time, the way platforms curate and govern content and interactions on their sites and its dynamic and opaque character evokes the threat of "invisibility" to creators, a development that has been described as "dangerous" for creators.[57]

[55] This section of the report is based on D.6.3. Final Evaluation and Measuring Report - impact of moderation practices and technologies on access and diversity" (2023), and in particular in the following working paper attached to that report: Daria Dergacheva, Christian Katzenbach and Paloma Viejo Otero: "Losing authenticity: social media creators' perspective on copyright restrictions in the EU" [forthcoming 2023].

[56] Brooke Erin Duffy and Colten Meisner, 'Platform Governance at the Margins: Social Media Creators' Experiences with Algorithmic (in)Visibility' [2022] Media, Culture & Society 01634437221111923; Taina Bucher, 'The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms' (2017) 20 Information, Communication & Society 30.

[57] Stuart Cunningham and David Craig, *Social Media Entertainment: The New Intersection of Hollywood and Silicon Valley*, vol 7 (NYU Press 2019) <https://www.jstor.org/stable/j.ctv12fw938> accessed 25 January 2023.

The main takeaway from our study is that users of social media platforms that do creative work are influenced by algorithmic content moderation. Perhaps our most important finding, which extends understanding on how algorithmic content moderation influences creative work on platforms, is that creators engage in self-censorship. That is to say, creators do avoid posting certain content or adjust it in advance in order to cater to the perceived functioning of platforms algorithmic content moderation. For many artists, anticipation of platform "punishments" (i.e. restrictive moderation actions) directly influenced the cultural products that they produced. In addition, because the regulative dimension of algorithmic copyright moderation is opaque for creators, they engage in "algorithmic gossip"[58] and use user folk theories[59] to try and guess which practices are accepted and which are not. These are important policy implications from this research, such as that more transparency in platform governance is needed, both from policy makers and platforms, so that the automated content moderation does not add to the uncertainty and insecurity of the creators' media production work on social media platforms.

[58] Sophie Bishop, 'Managing Visibility on YouTube through Algorithmic Gossip' (2019) 21 New Media & Society 2589.

[59] Michael A DeVito, Darren Gergle and Jeremy Birnholtz, '"Algorithms Ruin Everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media', *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery 2017) <https://doi.org/10.1145/3025453.3025659> accessed 25 January 2023.

4. RECOMMENDATIONS FOR FUTURE POLICY ACTIONS

In the following, we summarise the key recommendations for future policy actions based on our research.

Definition of OCSSPs

- Considering the potential for legal uncertainty and fragmentation of the digital single market as regards copyright content moderation, we recommend that the Commission reviews its Guidance on art. 17 CDSMD (COM/2021/288 final) in order to provide clearer guidelines on the definition of OCSSPs, especially for small and medium-sized online platforms and coordinates its application across Member States.

User Rights - recognition

- National legislators should review their national transpositions of art. 17 CDSMD to fully recognize the nature of the exceptions and limitations in paragraph (7) as "user rights" in accordance with CJEU jurisprudence, rather than mere defences.

User Rights – operationalisation

- We further recommend that the Commission reviews its Guidance in order to provide guidelines from the perspective of EU law as to the concrete implications of a "user rights" implementation of paragraph (7) in national laws. This should include, to the extent possible, concrete guidance on what type of actions users and their representatives (e.g., consumer organisations) may take against OCSSPs to protect their rights.

Complaint and redress safeguards – complementary nature

- National legislators should review their national transpositions of art. 17 CDSMD to ensure that *ex post* complaint and redress mechanisms under paragraph (9) are not the only means to ensure the application of user rights, but rather a complementary means, in line with the Court's judgment in case C-401/19.
- We further recommend that the Commission's Guidance is updated to fully reflect the Court's approach in case C-401/19, as regards the complementary role of complaint and redress mechanisms under paragraph (9).

Permissible preventive filtering

- The Commission should review its Guidance to clearly align it with the Court's judgment in C-401/19, namely by clarifying that: (1) OCSSPs can only deploy *ex ante* filtering/blocking measures if their content moderation systems can distinguish lawful from unlawful content without the need for its "independent assessment" by the providers; (2) such measures can only be deployed

for a clearly defined category of "manifestly infringing" and strictly defined category of "equivalent" content; and (3) such measures cannot be deployed for other categories of content, such as (non-manifestly infringing) "earmarked content". Member States should further adjust their national implementations of art. 17 CDSMD to reflect these principles.

- In implementing these principles, the Commission and Member States could take into consideration the approach proposed by the AG Opinion on how to limit the application of filters to manifestly infringing or "equivalent" content, including the consequence that all other uploads should benefit from a "presumption of lawfulness" and be subject to the *ex ante* and *ex post* safeguards embedded in art. 17, notably judicial review. In particular, the AG emphasized the main aim of the legislature to avoid over-blocking by securing a low rate of "false positives". Considering the requirements of the judgment, in order to determine acceptable error rates for content filtering tools, this approach implies that the concept of "manifestly infringing" content should only be applied to uploaded content that is identical or nearly identical to the information provided by the rightsholder that meets the requirements of art. 17(4) (b) and (c) CDSMD.

Relationship art. 17 CDSMD and DSA - clarification

- The Commission should review its Guidance to clarify which provisions in the DSA's liability framework and due diligence obligations Chapters apply to OCSSPs despite the *lex specialis* of art. 17 CDSMD, within the limits of the Commission's competence as outlined in art. 17(10) CDSMD.

Relationship art. 17 CDSMD and DSA – Terms and Conditions and Fundamental Rights

- The Commission should clarify in its Guidance that the obligations of Article 14 DSA apply to OCSSPs, in particular the obligation in paragraph (4) to apply and enforce content moderation restrictions with due regard to the fundamental rights of the recipients of the service, such as freedom of expression. The authorities and courts of the Member States should equally interpret their national law in a manner consistent with the application of art. 14 DSA to OCSSPs.

Monetisation and restrictive content moderation actions

- At EU level, EU institutions and in particular the Commission should explore to what extent the copyright *acquis* already contains rules addressing content moderation actions relating to monetization and related restrictive content moderation actions (e.g. shadow banning and downranking) of copyright-protected content on online platforms (e.g., in arts. 18 to 23 CDSMD), and to what extent policy action is needed in this area. Further research is needed specifically on the imbalanced nature of the contractual relationship of online platforms and uploading users, as well as in the transparency and fairness of their remuneration.

Recommender systems and copyright content moderation

- Although our research has focussed on issues of content *moderation*, we note the related but separate issue of content *recommendation*.[60] Whereas the actual phenomena are somewhat related, however, they relate to a different set of issues and perspectives. We note that more research is needed in the field of copyright content recommendation as well as copyright's role in content recommendation with a view to access and diversity. We therefore recommend that the EU institutions (e.g. the Commission through its Joint Research Centre) takes steps to carry out such research.

Transparency and robust data access for researchers

- At EU level, EU institutions and in particular the Commission should explore the application of the DSA's provisions on transparency and access to date to OCSSPs and non-OCSSPs hosting copyright protected content (see art. 40 DSA on data access and scrutiny[61]), as well as study and, if adequate, propose EU level action that imposes transparency and access to data obligations on online platforms regarding their copyright content moderation activities. Inspiration could be drawn by the design and implementation of the German national transposition law under Section 19(3) UrhDaG as regards rights to information. In that context, special care should be taken to: (1) ensure mandatory rules for data access for researchers; (2) carefully define the scope of beneficiary researchers, research institutions and research activities so not to be overly restrictive; (3) design a regime that avoids the potential negative effects of requiring researchers to reimburse the platforms' costs related to complying with such requests; (4) fund and support academic initiatives to build up collaborations and institutional capacity to develop and coordinate the necessary expertise and infrastructure to process this data, including database creation and secure processes for data access. To the extent possible, the Commission should advance recommendations in this direction in its revised version of the Guidance on art. 17 CDSMD.

Trade secret protection and transparency of content moderation systems

- In order to make transparency meaningful, in our view, proper account must be take on trade secrets protection, which likely extends to different aspects of human and algorithmic copyright

[60] NB that in the DSA the act of recommending content is conceptually included in the definition of "content moderation" (art. 3(t)), the obligations imposed on service providers for "recommender systems" (defined in art. 3(s)) are separate in the DSA from those on content moderation. The copyright acquis, and in particular the CDSMD, do not regulate this topic. On the relation between DSA and AIA on this matter, see also Sebastian Felix Schwemer, 'Recommender Systems in the EU: From Responsibility to Regulation' (2022) 1 Morals & Machines 60.

[61] The DSA enables data access to very large online platforms (VLOPs) and very large online search engines (VLOSEs) for "vetted researchers" under certain conditions. Under art. 40(4) and (8) DSA on data access and scrutiny, researchers can be granted the status of "vetted researchers" for the "sole purpose of conducting research that contributes to the detection, identification and understanding of systemic risks in the Union (…) and to the assessment of the adequacy, efficiency and impacts of the risk mitigation measures (…)" put in place for VLOPs and VLOSEs.

content moderation by platforms.[62] Consequently, achieving meaningful transparency in this area will likely require legislative intervention that exempts platforms algorithmic moderation systems from trade secrets protection, at least for purposes of data access and scrutiny by researchers and policy makers. EU institutions and in particular the Commission should explore the limitations of the current legal framework in this respect and propose the required legislative intervention to ensure this access.

Relationship art. 17 CDSMD and AI Act Proposal

- We recommend that the Commission studies the legal interplay between legislation on AI and platform regulation, in particular the issue of whether and to what extent algorithmic content moderation systems might be covered by the AIA proposal. Any such study should consider the future scenario and potential impact of algorithmic content moderation systems that rely on machine learning which will be deployed to assess contextual uses covered by user rights under art. 17(7) CDSMD, and how this might affect the permissibility of preventive filtering measures.

Human competences in copyright content moderation

- Our research indicates that competences of human moderators directly impact the quality of the content moderation system. This much is recognized in the DSA, CDSMD and expert recommendations the codes we reviewed, which require human review at minimum in the appeal process, partly as a means to mitigate the risks of automated content moderation. From our viewpoint, a certain level of human involvement should also be required to reduce biases and errors and ensure accuracy in the first stage of automated moderation. One way to achieve this would be to mandate or incentivize random accuracy tests by human intervention at this stage. We therefore recommend that the Commission explore the best practices and mechanisms to mandate or incentivize such random accuracy test for OCSSPs.

[62] In the EU, see Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure.

REFERENCES

[*Additional extensive reference lists, legislation and case tables available in Reports D.6.2 and D.6.3.*]

Angelopoulos C, 'Articles 15 & 17 of the Directive on Copyright in the Digital Single Market Comparative National Implementation Report' (Coalition for Creativity (C4C); CIPIL 2022) <https://informationlabs.org/copyright/> accessed 15 December 2022

Bishop S, 'Managing Visibility on YouTube through Algorithmic Gossip' (2019) 21 New Media & Society 2589

Bucher T, 'The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms' (2017) 20 Information, Communication & Society 30

Cotter K, '"Shadowbanning Is Not a Thing": Black Box Gaslighting and the Power to Independently Know and Credibly Critique Algorithms' (2021) 0 Information, Communication & Society 1

Cunningham S and Craig D, *Social Media Entertainment: The New Intersection of Hollywood and Silicon Valley*, vol 7 (NYU Press 2019) <https://www.jstor.org/stable/j.ctv12fw938> accessed 25 January 2023

DeVito MA, Gergle D and Birnholtz J, '"Algorithms Ruin Everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media', *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery 2017) <https://doi.org/10.1145/3025453.3025659> accessed 25 January 2023

Duffy BE and Meisner C, 'Platform Governance at the Margins: Social Media Creators' Experiences with Algorithmic (in)Visibility' [2022] Media, Culture & Society 01634437221111923

Gray JE and Suzor N, 'Playing with Machines: Using Machine Learning to Understand Automated Copyright Enforcement at Scale' (2020) 7 Big Data & Society

<https://journals.sagepub.com/doi/full/10.1177/2053951720919963> accessed 25 January 2023

Hacker P, Engel A and List T, 'Understanding and Regulating ChatGPT, and Other Large Generative AI Models: With input from ChatGPT' [2023] Verfassungsblog <https://verfassungsblog.de/chatgpt/> accessed 24 January 2023

Keller D and Leerssen P, 'Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation', *Social Media and Democracy: The State of the Field and Prospects for Reform* (Cambridge University Press 2019) <https://papers.ssrn.com/abstract=3504930> accessed 4 May 2021

Leerssen P, 'An End to Shadow Banning? Transparency Rights in the Digital Services Act between Content Moderation and Curation' <https://osf.io/7jg45/> accessed 23 November 2022

Margoni T, Quintais JP and Schwemer SF, 'Algorithmic Propagation: Do Property Rights in Data Increase Bias in Content Moderation? – Part II' (*Kluwer Copyright Blog*, 9 June 2022) <https://copyrightblog.kluweriplaw.com/2022/06/09/algorithmic-propagation-do-property-rights-in-data-increase-bias-in-content-moderation-part-ii/> accessed 24 January 2023

Peukert A and others, 'European Copyright Society – Comment on Copyright and the Digital Services Act Proposal' (2022) 53 IIC - International Review of Intellectual Property and Competition Law 358

Poell T, Nieborg D and Dijck J van, 'Platformisation' (2019) 8 Internet Policy Review <https://policyreview.info/concepts/platformisation> accessed 18 February 2022

Poort J and Pervaiz A, 'D3.2/3.3 Report(s) on the Perspectives of Authors and Performers' (Institute for Information Law (IViR) 2022) reCreating Europe Reports <https://zenodo.org/record/6779373> accessed 25 January 2023

Quintais J and Angelopoulos C, 'YouTube and Cyando, Joined Cases C-682/18 and C-683/18 (22 June 2021): Case Comment' [2022] Auteursrecht 46

Quintais JP and others, 'Copyright Content Moderation in the EU: An Interdisciplinary Mapping Analysis' (2022) reCreating Europe Report <https://papers.ssrn.com/abstract=4210278> accessed 7 September 2022

Quintais JP, Appelman N and Fahy R, 'Using Terms and Conditions to Apply Fundamental Rights to Content Moderation' [2023] German Law Journal

Quintais JP, Gregorio GD and Magalhães JC, 'How Platforms Govern Users' Copyright-Protected Content: Exploring the Power of Private Ordering and Its Implications [Forthcoming]' [2023] Computer Law & Security Review

Quintais JP and Schwemer SF, 'The Interplay between the Digital Services Act and Sector Regulation: How Special Is Copyright?' (2022) 13 European Journal of Risk Regulation 191

Savolainen L, 'The Shadow Banning Controversy: Perceived Governance and Algorithmic Folklore' (2022) 44 Media, Culture & Society 1091

Schwemer SF, 'Digital Services Act: A Reform of the E-Commerce Directive and Much More' in A Savin (ed), *Research Handbook of EU Internet Law [Forthcoming]* (Edward Elgar 2022)

——, 'Recommender Systems in the EU: From Responsibility to Regulation' (2022) 1 Morals & Machines 60

Schwemer SF and Schovsbo J, 'What Is Left of User Rights? – Algorithmic Copyright Enforcement and Free Speech in the Light of the Article 17 Regime', *Paul Torremans (ed), Intellectual Property Law and Human Rights* (4th edition, Wolters Kluwer 2020) <https://ssrn.com/abstract=3507542>

van Dijck J, Poell T and de Waal M, *The Platform Society* (Oxford University Press 2018) <https://oxford.universitypressscholarship.com/10.1093/oso/9780190889760.001.0001/oso-9780190889760> accessed 20 February 2022

YouTube, 'YouTube Copyright Transparency Report H1 2022' (YouTube 2022) Copyright Transparency Report <https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-22_2022-1-1_2022-6-30_en_v1.pdf>

References - forthcoming publications in project

D.6.3. Final Evaluation and Measuring Report - impact of moderation practices and technologies on access and diversity" (2023), and attached draft articles to be submitted to peer-reviewed journals

- Thomas Riis: "A theory of rough justice for internet intermediaries from the perspective of EU copyright law"
- Sebastian Felix Schwemer: "Quality of Automated Content Moderation: Regulatory Routes for Mitigating Errors"
- Thomas Margoni, João Pedro Quintais and Sebastian Felix Schwemer: "Algorithmic propagation: do property rights in data increase bias in content moderation?"
- Christian Katzenbach, Selim Basoglu and Dennis Redeker: "Finally Opening up? The evolution of transparency reporting practices of social media platforms", submitted to ICA 2023.
- Daria Dergacheva, Christian Katzenbach: "Mandate to Overblock? Understanding the impact of EU's Art. 17 on automated content moderation on YouTube", submitted to ICA 2023.
- Daria Dergacheva, Christian Katzenbach and Paloma Viejo Otero: "Losing authenticity: social media creators' perspective on copyright restrictions in the EU" submitted to ICA 2023.

The ReCreating Europe project aims at bringing a ground-breaking contribution to the understanding and management of copyright in the DSM, and at advancing the discussion on how IPRs can be best regulated to facilitate access to, consumption of and generation of cultural and creative products. The focus of such an exercise is on, inter alia, users' access to culture, barriers to accessibility, lending practices, content filtering performed by intermediaries, old and new business models in creative industries of different sizes, sectors and locations, experiences, perceptions and income developments of creators and performers, who are the beating heart of the EU cultural and copyright industries, and the emerging role of artificial intelligence (AI) in the creative process.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870626