



Dispute Resolution and Content Moderation:

Fair, Accountable, Independent,
Transparent, and Effective

Heidi Tworek

University of British Columbia

Ronan Ó Fathaigh

Institute for Information Law, University of Amsterdam

Lisanne Bruggeman

Institute for Information Law, University of Amsterdam

Chris Tenove

University of British Columbia

January 14, 2020

The Transatlantic Working Group Papers Series

Co-Chairs Reports

Co-Chairs Reports from TWG's Three Sessions:
Ditchley Park, Santa Monica, and Bellagio.

Freedom of Expression and Intermediary Liability

Freedom of Expression: A Comparative Summary of
United States and European Law
B. Heller & J. van Hoboken, May 3, 2019.

Design Principles for Intermediary Liability Laws
J. van Hoboken & D. Keller, October 8, 2019.

Existing Legislative Initiatives

An Analysis of Germany's NetzDG Law
H. Tworek & P. Leerssen, April 15, 2019.

The Proposed EU Terrorism Content Regulation:
Analysis and Recommendations with Respect to
Freedom of Expression Implications
J. van Hoboken, May 3, 2019.

Combating Terrorist-Related Content Through AI
and Information Sharing
B. Heller, April 26, 2019.

The European Commission's Code of Conduct for
Countering Illegal Hate Speech Online: An Analysis
of Freedom of Expression Implications
B. Bukovská, May 7, 2019.

The EU Code of Practice on Disinformation: The
Difficulty of Regulating a Nebulous Problem
P.H. Chase, August 29, 2019.

A Cycle of Censorship: The UK White Paper on
Online Harms and the Dangers of Regulating
Disinformation
P. Pomerantsev, October 1, 2019.

U.S. Initiatives to Counter Harmful Speech and
Disinformation on Social Media
A. Shahbaz, June 11, 2019.

ABC Framework to Address Disinformation

Actors, Behaviors, Content: A Disinformation ABC:
Highlighting Three Vectors of Viral Deception to
Guide Industry & Regulatory Responses
C. François, September 20, 2019.

Transparency and Accountability Solutions

Transparency Requirements for Digital Social Media
Platforms: Recommendations for Policy Makers
and Industry
M. MacCarthy, February 12, 2020.

Dispute Resolution and Content Moderation:
Fair, Accountable, Independent, Transparent,
and Effective
H. Tworek, R. Ó Fathaigh, L. Bruggeman & C. Tenove,
January 14, 2020.

Algorithms and Artificial Intelligence

An Examination of the Algorithmic
Accountability Act of 2019
M. MacCarthy, October 24, 2019.

Artificial Intelligence, Content Moderation, and
Freedom of Expression
E. Llansó, J. van Hoboken, P. Leerssen & J. Harambam,
February 26, 2020.

www.annenbergpublicpolicycenter.org/twg

**Dispute Resolution and Content Moderation:
Fair, Accountable, Independent, Transparent, and Effective[†]**

Heidi Tworek, University of British Columbia¹

Ronan Ó Fathaigh, Institute for Information Law, University of Amsterdam²

Lisanne Bruggeman, Institute for Information Law, University of Amsterdam³

Chris Tenove, University of British Columbia⁴

January 14, 2020

Contents

Introduction	2
1. Key questions	3
2. Content moderation and dispute resolution mechanisms	3
3. Essential standards	4
4. Range of regulatory measures	5
Social Media Councils	6
1. Current situation	6
2. Fundamental questions on social media councils	8
E-Courts	12
1. Fundamental questions on e-courts	12
Conclusions and recommendations	15
Appendix	17
Social Media Council Case Studies	17
E-Court Case Studies	21
Notes	28

[†] One in a series: A working paper of the Transatlantic Working Group on Content Moderation Online and Freedom of Expression. Read about the TWG: <https://www.ivir.nl/twg/>.

Introduction

In May 2019, Instagram announced that users could now appeal if they disagreed with the company's decision to remove a post.⁵ Let that sink in – until a few months ago, users could not appeal takedowns on one of the largest social media sites. Even now, those decisions are made internally with the appeal reviewed by a second content moderator.

Instagram's decision followed its parent company, Facebook, which had already allowed appeals. All these appeals rely upon trusting the legitimacy of private company processes. For any dispute resolution to function, everyone involved should feel that the dispute resolution mechanism holds legitimacy. That principle of legitimacy underpins our judicial systems and processes. It enables citizens to accept the results of court decisions and arbitrations. At present, social media companies do not command the same legitimacy. Many argue that they never can, nor should they as private entities.

Many users, scholars, civil society organizations, and policy makers have lost faith in social media companies' content moderation systems. The Electronic Frontier Foundation has described these systems as “fundamentally broken.”⁶ Problems abound: lack of respect for human rights standards, due process, and transparency; vague rules; inconsistent policy enforcement.

This paper offers suggestions for how to improve one specific aspect of content moderation: resolving disputes over takedowns. Dispute resolution may seem like a small area, but it actually encapsulates the legitimacy problems behind content moderation decisions. The mechanisms behind the decisions seem arbitrary to outside observers. Decisions are subject to no public scrutiny or accountability. Appeals can only be directed to the companies themselves, if appeals mechanisms even exist.

This paper considers the problem of legitimacy in content moderation decisions and suggests new institutions to rebalance the private and public interests in resolving disputes. David Kaye, the U.N. Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, has noted that appeal mechanisms and remedies are “limited or untimely to the point of nonexistence.”⁷ He has recommended that “[g]iven their impact on the public sphere,” social media companies “open themselves up to public accountability.”⁸ Our recommendations thus contribute to a broader conversation about rebalancing the relationship between companies, users, and governments. This is otherwise known as platform governance.⁹ Following the particular concerns of the Transatlantic Working Group, we seek solutions that safeguard freedom of expression.

Users seem to value companies' internal appeal mechanisms. In the first three months of 2019 alone, users challenged Facebook's removal of over 1.1 million pieces of content under its own rules on “hate speech.”¹⁰ Facebook restored just 10 percent of this content. There is no further public mechanism to dispute the decision. These numbers only cover users who took the time to complain, and only about hate speech, with millions more pieces of content removed under propaganda, violence, and harassment rules. The number of appeals raises the question of whether we need new institutions to adjudicate some of these decisions. This would help the companies, too, because currently they find themselves constantly under fire for their content moderation.

This paper makes two interventions. First, we propose five basic principles to bolster legitimacy in content moderation decisions: fairness, accountability, independence, transparency, and effectiveness (FAITE). Second, we suggest two possible institutions to achieve these principles in dispute resolution. First, social media councils could discuss terms of service, adjudication processes, and broader ethical questions. A second, complementary institution is e-courts.

The paper first introduces dispute resolution and content moderation. Section 2 analyses social media councils; Section 3 explores e-courts and online dispute resolution initiatives. Sections 2 and 3 draw on case studies from Europe and North America. The fourth section concludes with policy options and recommendations. The appendix provides further details on the case studies.

This paper is authored by Heidi Tworek, University of British Columbia (project lead); Ronan Ó Fathaigh, Institute for Information Law, University of Amsterdam; Lisanne Bruggeman, Information Law and Policy Lab (ILP Lab), Institute for Information Law, University of Amsterdam; and Chris Tenove, University of British Columbia. Nico van Eijk (Institute for Information Law, University of Amsterdam) was involved in the drafting phase until November 1, 2019. We are grateful to the members of the Transatlantic Working Group for their comments on various drafts of this paper before, at, and after our meeting in Bellagio in November 2019.

1. Key questions

- What principles should undergird dispute resolution mechanisms?
- How can we create legitimate dispute resolution mechanisms for content moderation?
- How do we strike a balance between privatized and public adjudication of online content?
- How do we better align content moderation with international human rights law and fundamental principles of freedom of expression?
- How do we bake freedom of expression into any institutions involved in social media?
- How do new institutions deal with the scale of online content?
- Who will finance new institutions?
- How do we ensure access to justice for content moderation questions?
- How do we create procedural safeguards to avoid any new systems being gamed?
- Who should be involved in discussing content moderation decisions?
- How should we deal with jurisdiction for content moderation?
- Who should be allowed to flag or complain about content moderation decisions?

2. Content moderation and dispute resolution mechanisms

Almost every day, new research documents the problems with social media companies' content moderation systems. These problems are acute for freedom of expression. One recent example was

that Google's AI tool to detect toxic comments often classified comments in African-American English as toxic.¹¹ Other minority and marginalised groups around the world fear suppression of their speech. Racist and misogynist content, often including death threats, may not be removed. At the same time, automated tools may amplify over-blocking and pre-publication censorship of political expression. And all of this is based on vague rules, which create chilling effects on discussing matters of public concern, like sexual health.¹² These are real-world harms, preventing individuals from exercising the most protected form of freedom of expression under international standards: political expression and expression on matters of public interest.

The time has arrived for democratic governments to play a more active role. The online environment has already become the main (and unprecedentedly open) space for exercising the right to freedom of expression. This means that the online world too is subject to international and regional human rights standards for freedom of expression. These standards stipulate that governments should not interfere with freedom of expression. Simultaneously, governments have a positive duty to enable a favourable environment for freedom of expression.¹³ This is particularly true for minority opinions and ideas, and those considered offensive and shocking by government authorities and the majority. Many democratic governments also bear responsibility to promote diversity and pluralism and to promote equitable access to the means of communication. This duty extends to protecting relations between individuals and private companies.

We argue that democratic governments' more active role extends to creating institutions specifically for dispute resolution. We see dispute resolution as fundamental for a legitimate regulatory framework for the online environment.¹⁴ We cannot assume that we can know what types of content are being deleted. Instead of focusing on particular types of content, we suggest institutions that can empower users. Social media were often initially praised as giving a voice to the voiceless. Yet AI tools used in content moderation now may disproportionately affect marginalised communities, including people of colour and women.¹⁵

Our paper suggests that e-courts could enable users to resolve disputes over content deletion through public scrutiny, when appropriate. Social media councils, in turn, could make content moderation more transparent, encourage information-sharing from companies, and create baseline standards for content moderation that safeguard freedom of expression and human rights.

Within dispute resolution, our paper mainly addresses the following scenarios:

1. A user's content has been removed, or a user's account has been suspended, based on a supposed violation of the social media company's terms of service, adopted community guidelines, or within the context of legally cognizable rights. The removal or suspension may have resulted from the company's own proactive review of content or a complaint (flagging) by another user or third party (e.g., civil society organisations).
2. A complaint that another user's content *should be* removed, but has not been removed. The complaint may have arisen from another user (e.g., a journalist) or a third party, such as a civil society organisation, and concerns content that is not removed despite being problematic.

3. Essential standards

Before exploring the possible models, we suggest essential criteria or standards for dispute resolution mechanisms. We derive these standards from the wealth of literature on independent self-regulatory bodies, particularly independent media regulatory authorities. We draw inspiration from legal instruments which set out criteria to guarantee the independence of national regulatory authorities from government and private interests.¹⁶ We see five fundamental principles (FAITE):

1. *Fairness*: Dispute resolution bodies must have fair procedures. All parties (users, social media companies, civil society organisations and other institutions/stakeholders) must be able to express their point of view, be provided with arguments and evidence put forward by other parties, and be able to comment on them. Dispute resolution bodies must have the necessary expertise, including members possessing the necessary knowledge and skills in the fields of freedom of expression and international human rights law.
2. *Accountability*: Dispute resolution bodies must be accountable to the public and civil society; all decisions taken and regulations adopted should be duly reasoned.
3. *Independence*: Dispute resolution bodies must be as independent as possible or their membership must be balanced to try to ensure independence. Members should hold terms of office long enough to ensure the independence of their actions. Members should disclose any circumstances that may, or may appear to, affect their independence or create a conflict of interest.¹⁷
4. *Transparency*: Any bodies should be established via a fully consultative and inclusive process. Their work should occur in an open, transparent, and participatory manner with transparent procedural and financial rules and decision-making procedures. Members must be appointed through a transparent procedure. Decisions must be published in easily accessible formats.
5. *Effectiveness*: Dispute resolution bodies must have adequate financial and human resources to carry out their function effectively; any complaint procedure should be free of charge and easily accessible. Dispute resolution bodies must have adequate powers to provide appropriate remedies.

4. Range of regulatory measures

Dispute resolution has long existed. Many industries have struggled with similar questions for decades, including the press, broadcasting, and advertising. We can thus draw on a wide range of self-regulatory, co-regulatory and statutory regulation models for dispute resolution.

First, there is **self-regulation**, which can take many forms.¹⁸ It can include self-regulation through *internal governance*, where companies themselves establish policies based on their own cultures, practices and control. Content moderation policies are usually developed by a social media company's legal officials, public policy and product managers, and senior executives.¹⁹ However, companies may also engage *external* parties to a certain degree, but still ultimately subject to the company's own governance structure. Examples include Twitter's Trust and Safety Council, which provides expert input on products, policies, and programmes;²⁰ and Google's short-lived Advanced Technology External Advisory Council, to monitor responsible development and use of AI.²¹ Facebook has also announced a planned Oversight Board.²²

This type of self-regulation can be complemented by self-regulation through *user regulation*, where users flag breaches of content rules, such as YouTube’s Trusted Flagger programme, assisting with enforcement of YouTube’s community guidelines.²³ Another example is Wikipedia’s Arbitration Committee, known as Wikipedia’s supreme court, which considers serious conduct disputes and appeals over blocked or banned users.²⁴

Further, there is self-regulation through *industry standards* and codes of conduct. The Global Network Initiative, for example, seeks to protect freedom of expression and privacy in the ICT industry by setting a global standard for responsible company decision-making.²⁵ The Internet Watch Foundation, established by the internet industry, reports potentially criminal child sexual abuse content online.²⁶ Crucially, this raises the question of the independence of regulatory bodies established by industry.²⁷

Second, there is **co-regulation**, which can also take many forms, and encompasses self-regulation plus an element of regulatory compulsion to enable effective enforcement. Government, legislation, or a statutory regulator can recognise self-regulatory bodies. Alternatively, government or regulators can approve codes of conduct, for example with the EU’s Business-to-Platform Regulation, which regulates disputes between certain online platforms and business users. While the legislation does not mandate that online platforms must resolve disputes through independent mediation, it sets out a list of statutory requirements (e.g., impartiality and independence) for any such alternative dispute resolution.

Finally, there is **statutory regulation**, where legislation sets the scope and coverage of regulation, which may be enforced by a statutory regulator. This can prove problematic and undermine freedom of expression. One example is Russia’s 2019 amendments to the Information, Information Technologies and Protection of Information Law, which provides that the communications regulator may order the removal of illegal online content without a court order, including “unreliable socially significant information.”²⁸

In the sections that follow, we discuss two potential and complementary options that adhere to international freedom of expression standards, and provide fair, independent, accountable, transparent, and effective dispute resolution mechanisms for content moderation.

Social Media Councils

In this section, we first discuss current proposals for social media councils, and some industry initiatives. We then delve into fundamental questions around social media councils.

1. Current situation

The idea of social media councils has gathered significant steam in the last two years. The U.N. Special Rapporteur described independent social media councils as “[a]mong the best ideas,” noting that “[e]ffective and rights-respecting press councils worldwide provide a model for imposing minimum levels of consistency, transparency and accountability to commercial content moderation.”²⁹ ARTICLE 19 put forward the idea of independent social media councils in early 2018.³⁰ Tenove, Tworek, and McKelvey proposed content moderation councils for Canada.³¹ In late 2018, Facebook announced the idea of its own oversight board.³² In early 2019, Stanford University’s Global Digital

Policy Incubator (GDPi), ARTICLE 19, and the U.N. Special Rapporteur held a multistakeholder conference on social media councils.³³

There are two main models for independent social media councils. One (the GDPi model) would begin from the global level, designed to provide guidelines and evaluate emblematic cases, not adjudicate individual complaints. The second (ARTICLE 19), a national or regional council, would allow users to bring a claim, but the council has discretion over its docket.

The GDPi model would be a global multistakeholder cross-platform social media council, designed to complement company-specific and national initiatives. The council would develop guidelines based on international human rights principles for how to approach content moderation online. It would *not* adjudicate individual cases or serve as an appeals body. Rather it would evaluate emblematic cases and create precedents for future cases. Further, it would advise platforms in developing and implementing their terms of service, mediate the interaction between governments and platforms, and provide advice, recommendations, and expertise to governments.

ARTICLE 19's model would be a network of national or regional social media councils, providing general guidance to social media platforms and, crucially, deciding individual users' complaints. National and regional councils would apply international standards on human rights, either directly or on the basis of a universal Code of Principles adopted globally. The social media councils would have an adjudicatory function. Individual users would have the right to bring a complaint, though only after adjudication at the platform level to avoid the risk of overload. A social media council would have full control of its docket, and could filter the cases it wants to review to "build a relationship of trust with the public."³⁴ In summer 2019, ARTICLE 19 published a consultation paper on social media councils as another phase in developing the idea.³⁵ These national councils could establish working relationships with any global body.

Social media companies, whether in cooperation or individually, are also establishing procedures for content moderation. An example of industry cooperation is the Global Internet Forum to Counter Terrorism (GIFCT), established in 2017 by Facebook, Twitter, YouTube and Microsoft. It enables coordination between social media companies seeking to remove "terrorist" content, and the GIFCT houses a hash database of "terrorist" images. However, as Tworek notes, there is no public oversight, and it remains a "mystery to those outside the companies or specific civil society organizations and governments who cooperate with the forum."³⁶ At a meeting between governments and technology companies at the U.N. General Assembly in September 2019, as part of the Christchurch Call to Action to eliminate terrorist content online, structural changes to GIFCT were proposed, including a government oversight board.³⁷ The GIFCT is now transforming into a nonprofit which may entail more transparency and ability for public oversight.

There are also company-specific plans. In September 2019, Facebook published further details of its planned oversight board.³⁸ Facebook proposes to establish an oversight board comprising 40 members who serve three-year terms. Facebook will select the co-chairs. Then, Facebook and the co-chairs select the rest of the members.³⁹ The board will review and decide on content according to Facebook's "content policies and values." If people disagree with the outcome of Facebook's decision and have exhausted their appeals, they can submit a request for review to the board.⁴⁰ The board chooses which requests to review and adjudicate. The board's resolution "will be binding and Facebook will

implement it promptly, unless implementation of a resolution could violate the law.”⁴¹ Facebook plans to start the system in 2020.⁴²

Facebook’s plans show that the company has recognised the problem of legitimacy in internal appeals. But questions remain over implementation and the value of a company-specific process. Social media councils offer a potential avenue to address these concerns. Below we consider some fundamental questions on how they might work.

2. Fundamental questions on social media councils

a. *Geography*: Should social media councils be national or international?

The geographical coverage of a social media council is central to all its potential functions, composition, and regulatory frameworks. The GDPi and ARTICLE 19 models offer two different approaches to this question, one international and one national. We suggest that a national social media council may be better placed to address three of the major concerns with the current system of content moderation and complaint-handling by social media companies: insufficient understanding of linguistic and cultural nuance; absence of rigorous human evaluation of context, including wholesale problems with addressing cultural context; and an inability to understand widespread variation of language cues, meaning, linguistic and cultural particularities.

This view is consistent with European press councils, which apply national codes of ethics and often deal with international media conglomerates. The Alliance of Independent Press Councils of Europe (a network of European press and broadcast councils) sees universal codes of ethics as impossible. Although press and media councils are national, almost all include international media firms. National councils and international companies could similarly cooperate on social media.

Further, national councils could complement each other in the transatlantic context, given the different freedom of expression standards. Countries could coordinate to create similar governance structures and composition. A national framework also recognises that different countries are at different stages of thinking on this issue. Some wish to move far faster than others, making international coordination between national councils more realistic at this point.

b. *Scope*: Should social media councils decide guidelines or examine individual cases?

Social media companies’ terms of service emerged on an ad hoc basis. Although these terms of service are now much more systematised, they did not place freedom of expression standards at their core. Ironically for companies generally built on the philosophy that more speech was better, there are now grave concerns that content moderation decisions actually harm freedom of expression. Because the basic problem emerged from arbitrary standards, a social media council could start by creating guidelines that embed freedom of expression standards in social media companies’ content rules.

But guideline-making would only go so far. The true test of guidelines is implementation and accountability. That is why the press industry established press councils *in addition* to codes of ethics. A social media council limited to guideline-making would leave implementation to social media companies themselves. It would also leave users without a crucial right to subject a social media company’s decision to independent review. Indeed, freedom of expression is so important that it

should include the procedural safeguard of review by an independent and impartial body. A social media council limited to guidelines-making would not offer such a procedural safeguard. One solution would be to send individual cases to e-courts.

Further, a social media council limited to guideline-making would not solve the crucial problems of (a) widespread inconsistency and unpredictability in content moderation decisions; (b) inability to challenge or follow-up on content-related complaints; (c) lack of transparency and due process; (d) widespread use of automated decision-making; and (e) insufficient remedies.

Alongside guidelines, social media councils could help to coordinate third-party research into companies, compile decisions from different companies, and compare their content moderation policies. They could create industry-wide boards of ethics for new questions like AI and develop best practices for crucial questions like labour standards for content moderators.⁴³

Finally, a decision-making mechanism could cure the major deficiencies in the remedies available. Current systems offer no publication of a reasoned decision, no right of reply, no publication of an apology or acknowledgment. Where content removal causes specific reputational, physical, moral and/or financial harm, there are few mechanisms for settlements. A national social media council with a decision-making function could provide some of these remedies. Social media councils could develop and apply national standards, akin to how press councils make decisions based on national press codes. Here, civil society engagement could ensure that social media companies' content moderation rules and systems reflect freedom of expression principles.

c. *Regulatory framework*: What type of body might a social media council resemble?

There are myriad regulatory frameworks spanning from self-regulation to statutory regulation. With social media councils, the major question is whether they should be voluntary. The voluntary nature of social media councils may help build trust and accountability between the public and social media companies. However, if social media companies are unwilling to send representatives, governments may start to contemplate other regulatory options. Canada, for example, offers three possible frameworks (see the appendix for more detail on each).

i) Self-regulation like the National NewsMedia Council or an Ads Standards Council.

It is tempting to frame social media councils as updated media councils. We cannot forget, however, the fundamental differences between user-generated social media content and journalism produced by a comparatively small number of newspaper and broadcast professionals. In both cases, relatively few companies are involved. But the content producers are very different – journalists versus almost every member of the public. The scale differs vastly too. Any social media council would need to account for these fundamental differences.

ii) Co-regulation like a Broadcasting Standards Council.

Private radio and television broadcasters created the Canadian Broadcasting Standards Council (CBSC) to devise industry codes and to address audience complaints about their programming, including when it is streamed online. Membership in the CBSC is voluntary, but non-members have complaints addressed directly by the government regulator of the entire broadcasting sector (the

Canadian Radio-television and Telecommunications Commission, or CRTC). The CRTC functions as an appellate body for those unsatisfied by CBSC judgments.

A social media council based on this framework would help develop and implement standards in a largely self-regulatory manner, while involving stakeholder input and government oversight. Consultations to create the council must include major national and international internet companies, and government must offer appropriate incentives – or threats of more intrusive regulation – to secure participation. To prevent the council’s composition becoming lopsided, the participation of other key stakeholders would be encouraged and supported, ranging from Indigenous communities to human rights organizations to political parties.

A social media council constituted under the auspices of the CRTC would apply a new group-based approach, pursuing “broadly based agreements tailored to and established with a few dozen specific companies or affiliated groups of companies, individually or collectively offering a variety of services (service groups)” in order to provide “public scrutiny and should set out specific binding commitments applicable to the service group.”⁴⁴ While we caution against using this experimental framework for online broadcasting or digital common carriers, the content moderation industry could fit as a group in this new “nimble regulatory” approach. The social media council would then be responsible for enforcing the group’s service agreements.

iii) A human rights framework and human rights bodies.

At the federal and provincial levels, human rights legislation sets out means to address rights. Human rights bodies could provide guidance and act as a complaints body to address violations of certain rights through the use of social media platforms. Human rights bodies have issued decisions regarding hate speech and discrimination on websites and message boards. More recently, the Canadian and Ontario human rights commissions jointly demanded that Facebook introduce safeguards to protect against discrimination in targeted employment advertising.⁴⁵ The Canadian Human Rights Commission (CHRC) is most likely to have jurisdiction over issues related to social media content moderation.

Nevertheless, the CHRC faces several challenges in addressing content moderation issues. First, current human rights legislation applies to a very narrow set of issues, particularly since the federal Conservative government removed the provision addressing hate speech in the Canadian Human Rights Act in 2013. Second, there are complex jurisdictional questions regarding the application of federal or provincial human rights frameworks. Different bodies have jurisdiction over different issue areas (e.g., employment, hate speech, etc.), and subject entities (e.g., broadcasters, web publishers, etc.). The CHRC’s decisions are enforceable at the national level, but the CHRC’s framework may be coordinated globally via the Global Alliance for National Human Rights Institutions and through reference to international laws and guidelines. Greater clarity would be necessary, either through new legislation or precedent-setting decisions.

d. *Composition*: Who should sit on a social media council?

Composition is a complicated and crucial question. As with European press councils, each country may take a different approach. The Netherlands Press Council, for example, has a chairperson, four

vice-chairs, 10 member-journalists, and 10 non-journalist members. In Germany, however, only publishers and journalists comprise the Press Council board with no independent representatives.

i) How should civil society be involved?

Obviously, not every social media user can have the right to sit on a social media council. But different groups of users need representation for a council to gain any public legitimacy. Civil society involvement is crucial, unlike current press or media councils. Civil society involvement in creating social media councils, and broad representation, is vital for public trust. Civil society could possibly lodge complaints with councils, support complaints, and lodge third-party submissions. Finally, civil society would operate as an important check on the operation of social media councils, and assess whether they are contributing to an enabling environment for freedom of expression online. Each country would have to consider carefully and consult about which civil society organisations or representatives might contribute to the council. The council might particularly try to include those who have suffered most from hate speech or abusive speech.

ii) What size of social media company should participate?

In policy circles, “social media companies” is often used as a synonym for Facebook and Twitter. But this overlooks influential alternative networks such as Reddit, or Mumsnet in the UK. Given that participating in councils can involve significant company time, we should be attentive to the potential burdens of membership. One approach is only to include companies over a certain size. Another is to apply a sliding scale of participation based on size, market capitalisation, or another measure.

Another approach is to think about the council starting from the perspective of smaller companies. Any initiatives in this space should be wary of moves that unintentionally “lock in” the big players and stifle smaller companies. European press councils, for example, often involve not only dominant and large pan-European media companies, such as Axel Springer or Sanoma, but also small online-only news publications, with limited readership and staff.

e. *Diversity of approaches*: Is it valuable to maintain multiple approaches to content moderation?

Smaller and medium-size companies often bring very different approaches to content moderation. Even the major companies differ significantly, as the recent policy decisions on political advertising show. This raises a broader concern about whether social media councils should support a diversity of approaches to content moderation or create more uniform standards.

At base, this is a question about competition: do we want social media councils to encourage uniformity or do we want an ecosystem where different social media companies make different decisions? There are advantages both to uniformity and multiplicity (especially in a truly competitive environment with adequate substitutes). One way to thread the needle is for social media councils to create baseline standards for freedom of expression. Above that baseline, companies could implement their own content moderation policies. This could address the widespread inconsistency and unpredictability in content moderation decisions.

Rather than imposing uniform content moderation, a social media council can create uniform consistency and predictability of content moderation decisions.

f. *Jurisdiction.*

As with any online dispute, jurisdiction is a major consideration. Here, we suggest it could depend upon the type of dispute. We can look at this through the two different types of disputes we sought to address in this paper.

- i) The first category is content removed or an account suspended based on a supposed violation of the social media company's own content rules or terms of service. The removal or suspension may have resulted from the company's own proactive review of content or having been flagged by another user (e.g., journalist) or third party (e.g., civil society organisations). Here, jurisdiction could be based upon the location of the user whose content or account was flagged.
- ii) The second category is a complaint that another user's content should be removed, based on the company's own content rules or terms of service, but it is not. Here, jurisdiction is trickier. The complaint may have arisen from another user, or third party, such as a civil society organisation or government agency, based in another country from the content's author. If social media councils allow these types of public-interest complaints where complainants are not directly affected, then jurisdiction could depend upon the location of the content's author.

Most realistically, social media councils could facilitate discussion with civil society, create baseline standards and ethics, and improve company transparency. We suggest that individual appeals about content removal might best be directed to a second new institution: e-courts.

E-Courts

A second possible dispute resolution model is e-courts or internet courts. E-courts exist throughout the world to resolve disputes between governments and individuals, companies and individuals, disputes between individuals, and disputes between companies. In the Netherlands, for example, various types of e-courts started in 2009. However, there is no agreed definition of e-courts, and they take many forms. As one legal scholar notes, online dispute resolution bodies "generally facilitate settlement or substantive determination on the merits," while e-courts are "more limited to ending the dispute or providing a remedy or result based on limited parameters."⁴⁶ For content moderation, e-courts could be crucial to upholding freedom of expression standards.

Dispute resolution is already fundamental for many online companies. One of the best-known examples is eBay's online dispute resolution mechanisms, which resolve over 60 million disputes between eBay traders and users annually.⁴⁷ One programme is for feedback disputes, where an eBay seller may challenge a review posting about a seller's businesses. As part of eBay's online dispute resolution procedure, an impartial third-party reviewer from a professional dispute resolution service can examine the challenged posting and determine whether to affirm, withdraw or take no action on the review.⁴⁸ Here, we consider how to create public e-courts.

1. Fundamental questions on e-courts

- a. *Regulatory model:* What is a feasible regulatory model for dispute resolution of content?

There are three main appropriate models for e-courts. Unsurprisingly, the feasibility of e-courts fundamentally depends upon the type of e-court. Both Europe and North America have already created myriad e-courts which can help us to understand what has worked and what has not.

i) Online judicial adjudication.

This model uses technology to bring speed, efficiency, and lower costs. These are court-operated and judges make decisions. The model includes public or private online platforms for electronic case management, filings and communications. Examples include the Michigan Online Case Review in the United States and the European Small Claims Procedure, both discussed in further detail in the appendix. These e-courts digitise court processes such as small civil claims or minor offences, using technology to bring speed, efficiency, and lower costs.

Online judicial adjudication would provide a fast, simple, and cheap mechanism for users and others to challenge content moderation decisions made by social media companies. It would be fully online with no physical presence of the parties. Specially trained judges would make decisions. The simplified procedure would resemble small claim procedures; there would be no right of appeal to the general courts. Social media companies would assign specialised lawyers to process claims before the e-courts. The e-courts would regularly publish case-law compilations. Finally, the social media company would have to acquiesce to e-court jurisdiction to operate in a state.

Notably, this model would probably require legislation. The European Small Claims Procedure required legislation to be enacted. An e-court modelled on such a procedure at the European or national level would seem to require legislation – particularly if there were no right of appeal to general courts, and social media companies were subject to mandatory jurisdiction and the rules in order to operate.

ii) Online dispute resolution with built-in independent adjudication.

This model encourages early resolution through online dispute resolution, while preserving independent adjudication as a last resort. The model goes beyond digitising court systems, and involves online negotiation, independent mediation, arbitration, and adjudication. One of the best examples is the British Columbia Civil Resolution Tribunal, with its tiered approach: (a) online solution platform; (b) online facilitated negotiation; (c) online facilitated settlement; and (d) online hearing and adjudication.

Similar to the first model, legislation would be required to establish such an e-court, as it would include rules on access to courts, and judicial review of decisions. This model is also specifically designed to encourage early dispute resolution, and adjudication only as a matter of last resort, and therefore may save resources. Further, the independent adjudication would be carried out by officials established by legislation, rather than judges. This raises the question of whether this model is akin to statutory regulation, as the Civil Resolution Tribunal acts similarly to a statutory regulator, like a broadcasting regulator in Europe.

iii) Online independent dispute resolution, which may be operated by both private and public bodies.

This has a similar procedure to the previous model, but without independent adjudication. The best example operated by a public body is the European Online Dispute Resolution platform, and would involve: (a) online solution platform; (b) online facilitated negotiation; and (c) online facilitated settlement. This model can be put on a legislative footing, but it is not necessary.

Unlike the other models, this process may be established by private companies themselves. This would be the most feasible model not requiring legislation, and could also be built to encourage early dispute resolution, rather than proceedings through courts. It would also offer users an independent out-of-court dispute resolution mechanism. Of course, such mechanisms would need to be consistent with the EU's Alternative Dispute Resolution Directive, and some social media companies do already offer similar procedures for business users.

b. *Scalability*: How can e-courts cope with the potential scale of cases?

The most obvious question about e-courts is whether the low barrier to entry and admirable principle of access could create an unmanageable caseload. One method to cope within the EU is to scale national e-courts up to a regional e-court, and take the European Small Claims Procedure and European Online Dispute Resolution as possible models for a network of national e-courts.

Further, extant e-courts suggest a manageable scale. In 2017, the European Online Dispute Resolution platform handled over 24,000 complaints, the Dutch Stichting e-Court handled more than 20,000 cases, and in its latest figures, the Civil Resolution Tribunal in Canada handled nearly 12,000 disputes. The Civil Resolution Tribunal has been gradually extending its competence beyond the initial small claims it handled, and could provide a model for how to scale an e-court with more competences. Finally, for a benchmark from a statutory regulator, the UK broadcasting regulator (Ofcom) dealt with over 55,000 complaints last year.⁴⁹

c. *Accessibility*: How can e-courts ensure accessibility while preventing gaming of the system?

One major advantage of e-courts is that they can enable greater access to justice. With reduced costs and the ability to file from home, an e-court could lower the barrier to entry. It could also help to prevent some current speech-law cases that create a Streisand effect, meaning that a case unintentionally draws even more attention to the information or speech that the case tried to prevent.

A separate concern in the online environment is that malevolent actors may try to game the system and launch multiple complaints against a user's content to try to get that person removed or their content blocked. The e-court system could perhaps use automated detection of repeated submissions to prevent anyone coordinating to flood the system. More broadly, any institution would need to have procedural safeguards to prevent the court system being used as a tool of trolling or harassment, the very things it was designed to prevent.

d. *Compatibility*: How are e-courts compatible with European, American, or Canadian law?

Both North American and European legal systems have already given legislative footing to e-court models. At the European level, the European Online Dispute Resolution and European Small Claims Procedure provide a framework. The examples from the Netherlands in the appendix demonstrate the operation of national e-courts.

One crucial point about e-courts is that government legislation may be needed to create the institution. But judges would decide individual cases, not government agencies. This is critical to uphold international freedom of expression standards.

e. *Finances*: Who will pay for e-courts?

Like any judicial institution, public authorities would seem the obvious choice to cover the cost of e-courts. Their online nature should make them comparatively inexpensive. Another option is that platforms bear the cost of the service. There are currently discussions over tax policy related to online services. Any new or additional tax revenue from platforms could be used to pay for e-courts.

f. *Jurisdiction*.

This raises similar issues discussed in the previous section on social media councils, including whether individuals may make complaints where they are not directly affected. The e-court model would have to adopt rules on whether only individuals directly affected by content may initiate a claim, or whether civil society organisations, for example, may initiate group-based or public-interest claims. Although complex, these jurisdictional questions apply to similar issues like the right to be forgotten in the EU.

Conclusions and recommendations

This paper seeks to provide an appropriate framework for building dispute resolution mechanisms which are fair, accountable, independent, transparent, and effective. Having considered the range of social media council models in Section 2, and the range of e-court models in Section 3, we offer the following conclusions and recommendations:

1. *Fairness, accountability, independence, transparency and effective standards (FAITE) are essential*: Dispute resolution mechanisms for online content moderation must be consistent with the essential standards of fairness, accountability, independence, transparency and effectiveness. Both social media councils and e-courts can satisfy these standards, provided the standards are applied during the creation and operation of these bodies.
2. *The model adopted must connect to identified problems*: It is crucial to adopt a model that best remedies the problems associated with current content moderation, particularly legitimacy, and best provides an enabling environment for freedom of expression online. This paper addresses a subset of the most relevant problems in content moderation, such as when companies' terms of service/community guidelines or behaviour based on or following from other (legal) instruments may violate a person's fundamental rights.
3. *Dispute resolution mechanisms are a key part of establishing legitimacy and creating public accountability*: Independent complaint/dispute mechanisms can remedy myriad other problems of the current system operated by social media companies: widespread inconsistency and unpredictability in content moderation decisions; inability to challenge content actions or follow up on content-related complaints; lack of transparency, due process and notice; widespread problematic use of automated decision-making; and wholly insufficient remedies. Complaint/dispute mechanisms would provide the necessary public accountability of the freedom of expression standards implemented into social media

companies' content rules, and provide a crucial procedural safeguard of review by an independent and impartial body for users and the public.

4. *Social media councils and e-courts are two institutional options:* We have outlined two detailed models for institutions that could create greater public accountability and transparency. In an ideal world, these institutions can benefit everyone involved: users, civil society groups, companies, and governments. These two institutions will not solve all the problems of platform governance. But they will start toward finding a legitimate balance between the public and private interests in promoting freedom of expression.
5. *Social media councils and e-courts are not mutually exclusive:* Their compatibility depends upon the models for social media councils and e-courts. If a social media council is limited to guideline-making, an e-court would provide decisions in actual cases. If social media councils have a decision-making function, there could be coordination issues. A social media council could be combined with an e-court or social media councils could be involved in decisions preceding e-court decision-making.
6. *Freedom of expression should be a core value for any new institutions:* The current content moderation system has many potential detrimental effects on freedom of expression, which flow from the lack of freedom of expression standards built into social media companies' content rules. Freedom of expression principles do not form the basis of content moderation decisions. The opposite should hold for new democratic institutions like social media councils or e-courts. Freedom of expression concerns should be baked into the design, composition, and creation of any new institutions.
7. *Social media companies can help themselves by acting now:* Even without creating a social media council or e-court, social media companies themselves can already begin to alleviate these problems by incorporating freedom of expression standards into their content rules, and freedom of expression standards into their content moderation decisions and complaint-handling systems. Social media companies should invite relevant stakeholders to assist in this process, and formulate the appropriate freedom of expression standards upholding the FAITE standards of freedom of expression online.

Dispute resolution is complicated in every industry. But fair procedures are crucial to generate trust and legitimacy. This is even more true for such a sensitive area as speech. These two institutions of social media councils and e-courts are not panaceas for all the problems of platform governance. But they could go a long way in addressing some of the most pressing concerns.

Appendix

Here we provide deep dives on our particular case studies of councils and e-court systems. This material provides further details for those interested in the mechanics of particular systems.

Social Media Council Case Studies

1. Case study on social media councils in North America

For a North American case study, we focus on Canada. Canada does not currently regulate social media as a sector, though a range of existing regulatory frameworks apply. These include federal and provincial privacy regulation and regulation on advertising. They also include criminal, civil, administrative and human rights laws that address certain forms of prohibited speech, including defamation, hate propaganda, counselling terrorism and the non-consensual distribution of intimate images.

Multiple parliamentary committees and the Canadian government have recommended action to impose duties on social media companies “to remove manifestly illegal content in a timely fashion, including hate speech, harassment and disinformation, or risk monetary sanctions commensurate with the dominance and significance of the social platform, and allowing for judicial oversight of takedown decisions and a right of appeal.”⁵⁰ The Liberal Party has formed a minority government following the federal election in October 2019. The party’s platform claims that it will “target online hate speech, exploitation and harassment,” and “when social media platforms are used to spread these harmful views, the platforms themselves must also be held accountable.”⁵¹

Canada, as in many areas, pursues a regulatory framework that is partway between American and more interventionist European approaches. On the one hand, Canada has strong constitutional protections of freedom of expression, relatively minimal regulation of publishers and light regulation of broadcasters with respect to content. On the other hand, Canada has criminal laws against hate propaganda, and its human rights bodies have taken action on discriminatory or hateful communication disseminated by websites, publishers, broadcasters, and – in a few cases – social media companies.⁵²

One report proposed a social media council as a policy measure.⁵³ Within Canada, there are three potential routes for a social media council: a self-regulatory body (similar to Canada’s National NewsMedia Council or Ad Standards), a co-regulatory body facilitated by the Canadian Radio-television and Telecommunications Commission (similar to the Canadian Broadcast Standards Council), and a human rights approach (primarily pursued through the Canadian Human Rights Commission).

Self-regulatory social media council: Canada already has self-regulatory bodies for two cognate sectors to social media: the National NewsMedia Council (for news media other than broadcasters) and Ad Standards (the advertising sector). The National NewsMedia Council (NNC) is Canada’s main press council. It is a voluntary, self-regulatory body for the news media industry, with over 500 member organizations. The council promotes media ethics and operates a process to hear complaints.

The Council does not have its own standards, but relies on “the news outlet’s code of practice or generally accepted journalistic standards.”⁵⁴ If a complaint process finds that the news organization failed to uphold these standards, and that adequate corrective measures were not taken, the organization can “publish a fair report including, at minimum, a summary of the decision, and a link to the decision on the NNC website.” The NNC’s enforcement capacities are very weak, leading to questions about its relevance.

Ad Standards is a self-regulatory body for Canada’s advertising sector. It administers the *Code of Advertising Standards*, and operates a consumer complaints process about advertisements that run in Canadian media. The Code was created in 1963 after broad consultation with industry and stakeholders, and is regularly updated. Among its provisions, it prohibits advertisements that condone discrimination based upon race, national or ethnic origin, religion, gender identity, sex or sexual orientation, age or disability; that condone violence; that demean individuals or groups; or that “undermine human dignity.” The Consumer Complaints Process encourages advertisers to take voluntary actions and address the complaints, and may forward the complaints to a Standards Council to make a decision. Ad Standards enforces a council’s rulings by calling on media organizations to no longer host the advertising, and informing the Competition Bureau or other regulatory bodies that the advertiser is not compliant with the Code.

A co-regulatory social media council, under the auspices of the CRTC: A social media council could also follow the approach taken by private broadcasters in Canada. Private radio and television broadcasters created the Canadian Broadcasting Standards Council (CBSC) to address audience complaints about their programming, including when streamed online. Describing its mandate, the CBSC states:

Broadcasters have the ability to influence opinion, modify attitudes and shape minds. That’s why the industry created a voluntary system of codes that set high standards for all of their programming. Through these codes, private broadcasters promise to respect the interests and sensitivities of the people they serve, while meeting their responsibility to preserve the industry’s creative, editorial and journalistic freedom.⁵⁵

Membership in the CBSC is voluntary, but non-members will have complaints addressed directly by the government regulator of the entire broadcasting sector (the Canadian Radio-television and Telecommunications Commission, or CRTC), and the CRTC functions as an appellate body for those unsatisfied by CBSC judgements.

The CBSC applies industry codes on issues relating to ethics, violence on television, equitable portrayal, journalistic ethics, and cross-media ownership. When it receives complaints, the CBSC may constitute an Adjudicating Panel to investigate, arrive at, and publish a decision. If the broadcaster is found to have violated a code, it must announce that result several times on air. The CBSC cannot levy fines or revoke broadcast licenses.

A social media council based on this framework would help develop and implement standards in a largely self-regulatory manner, while involving stakeholder input and government oversight. Once established, the council could help social media platforms coordinate their actions to mitigate harmful speech, disinformation and other threats to democratic discourse online. Consultations to create the

council must include major national and international internet companies, and government must offer appropriate incentives – and threats of more intrusive regulation – to secure their participation. To ensure the composition of the council is not lopsided, the participation of other key stakeholders would be encouraged and supported, ranging from Indigenous communities to human rights organizations to political parties.

A social media council constituted under the auspices of the CRTC would apply a new group-based approach, pursuing “broadly based agreements tailored to and established with a few dozen specific companies or affiliated groups of companies, individually or collectively offering a variety of services (service groups)” in order to provide “public scrutiny and should set out specific binding commitments applicable to the service group.”⁵⁶ While we caution against using this experimental framework for online broadcasting or digital common carriers, the content moderation industry could fit as a group in this new “nimble regulatory” approach. The social media council would then be responsible for enforcing the group’s service agreements.

A human rights framework: Canada also addresses disputes over mediated communication through a human rights framework and human rights bodies. This offers an alternative or a productive complement to a social media council. At the federal and provincial levels, human rights legislation sets out means to address rights violations. Human rights bodies have issued decisions regarding hate speech and discrimination on websites and message boards. More recently, the Canadian and Ontario human rights commissions jointly demanded that Facebook introduce safeguards to protect against discrimination in targeted employment advertising.⁵⁷ The Canadian Human Rights Commission (CHRC) is most likely to have jurisdiction over issues related to social media content moderation.

The CHRC and some provincial human rights commissions bodies can constitute tribunals to hear and investigate specific complaints, compel evidence and enforce remedies. This human rights approach offers several advantages:

- The CHRC and other human rights bodies can constitute tribunals to hear and investigate specific complaints (from individuals or groups), compel evidence, and enforce remedies. Complaints may address those responsible for creating or disseminating problematic communications. These investigative processes follow stronger rule of law provisions than self-regulatory bodies, while generally being more accessible to complainants than civil or criminal processes.
- The CHRC and other human rights bodies can issue remedies that are legally enforceable. These can include retractions or removals of content and fines. Failure to adhere to decisions can lead to criminal prosecution.
- In addition to addressing individual complaints, the CHRC can study human rights issues, including systemic risks to rights. These studies enable the CHRC to comment on legislative proposals and play a public education role.
- Human rights bodies are bound by a commitment to freedom of expression as a constitutionally protected right of Canadians, and restrictions on speech through the rulings of Canadian human rights bodies can be reviewed by court systems.

- The Canadian Human Rights Commission applies a Canadian legal framework. However, it is a member of the Global Alliance for National Human Rights Institutions, which helps state-based human rights institutions to achieve shared goals and engage with the United Nations' Human Rights Council and Treaty Bodies. National human rights bodies may therefore be able to take a shared approach on issues related to social media content moderation, drawing on international human rights law and bodies.

Nevertheless, the CHRC faces several challenges in addressing content moderation issues that may affect the enjoyment of human rights. There are complex jurisdictional questions regarding the application of federal or provincial human rights frameworks. Different bodies have jurisdiction over different issue areas (e.g., employment, hate speech, etc.), and subject entities (e.g., broadcasters, web publishers, etc.). Greater clarity is needed, either through new legislation or precedent-setting decisions. Another challenge is that the provision addressing hate speech in the Canadian Human Rights Act was removed in 2013 by the federal Conservative government.⁵⁸ Policy makers and experts in Canada have argued for a new provision to address hate speech, but no concrete provision has been put forward to Parliament.⁵⁹

Human rights bodies could provide guidance and act as a complaints body to address violations of certain rights through the use of social media platforms. Decisions by the CHRC are enforceable at the national level, but the framework applied by the CHRC may be coordinated globally via the Global Alliance for National Human Rights Institutions and through reference to international laws and guidelines. However, there remain significant jurisdictional challenges, and current human rights legislation applies to a very narrow set of issue areas.

2. *Case study on media councils in Europe*

Social media councils could be modelled on press councils, as they enable industry-wide complaint mechanisms and appropriate remedies. Press councils – although mainly active in a professional environment – may provide a helpful starting point for social media councils as they have a long history of dealing with the various issues at stake, such as independence, impartiality, fairness, and effectiveness (including remedies).

First, press and media councils in Europe are all national and apply national codes of ethics, with no pan-European press council. The Alliance of Independent Press Councils of Europe (a network of press and broadcast councils from across Europe) has two “core beliefs” on jurisdiction. First, regulation should be based on nations’ differing cultures. Second, a universal code of ethics is impossible, and there should be no supranational codes and regulatory organisations, whether European or global.⁶⁰ While these principles concern media ethics codes, such views may also apply online. One criticism of the current content moderation system is that social media companies have “insufficient understanding of linguistic and cultural nuance,” an “absence of rigorous human evaluation of context,” and problems with “addressing context, widespread variation of language cues and meaning and linguistic and cultural particularities.”⁶¹ National content rules and national social media councils may be more appropriate. The “national” scope of the press and media councils doesn’t exclude the fact that many of the involved companies are international.

Second, independence is a key principle for press and media councils. For example, the Netherlands Press Council is composed of a chairperson, four (vice) chairs, 10 member-journalists and 10 non-journalist members. The chair is a high-profile journalist, and the four vice-chairs are (former) members of the judiciary. The Netherlands Press Council is established by a foundation called the Stichting Raad voor de Journalistiek, which is composed of the Netherlands Union of Journalists, the Society of Chief-Editors, printed press organisations, and public and commercial broadcasting organisations. This foundation appoints the members of the press council. Another example is Germany, where only publishers and journalists sit on the Press Council board, with no independent representatives.⁶² If these models were applied to social media councils, the council would include social media company officials, and be established by a foundation comprising social media interest groups. It is difficult to see how such a social media council would satisfy the principle of independence.

Third, there is not a uniform view on who can complain to press councils. In countries such as Sweden, Denmark, and Ireland, only a person affected by the material can complain. In contrast, in Finland and Germany, the Councils will accept a complaint from any complainant, which could include complaints about general issues of misleading reporting or the failure to separate fact from opinion. There are similar issues for online content moderation decisions. Should any user (or civil society organisation) be able to complain that another user's content should be removed, based on the company's own content rules or terms of service? Of course, press and media councils mainly address issues on content which falls under direct editorial responsibility, and the essential difference with social media companies is that content is generated by users.

Finally, European press councils offer ideas about potential remedies. Most press councils only provide remedies in the form of publication of a decision. The Netherlands Press Council has no power to impose sanctions or to order a correction, rectification, or reply. The vast majority of press councils do not have the power to issue fines, or order corrections, apologies or removal of content. Only one press council, the Swedish Press Council, can impose an administrative fee that is imposed on an upheld complaint, and is tiered depending on the circulation of the publication.⁶³ In addition, press councils can also operate arbitration procedures.⁶⁴

E-Court Case Studies

1. European dispute resolution mechanisms

The EU has adopted specific legislation in relation to small claim court procedures, out-of-court dispute resolution, and online dispute resolution, which can inform our discussion. Some examples concern EU cross-border initiatives and build upon national systems.

a. Alternative Dispute Resolution Directive

First, there is the Alternative Dispute Resolution Directive of 2013.⁶⁵ This directive aims to harmonise rules across the EU to ensure that consumers can voluntarily submit complaints against companies to ADR bodies, which offer independent, impartial, transparent, effective, fast and fair alternative dispute resolution procedures. The ADR Directive applies to procedures for the out-of-court resolution of domestic and cross-border disputes concerning contractual obligations stemming from sales or service

contracts between a trader and a consumer. Crucially, the directive includes “quality requirements” for ADR bodies and procedures, including detailed rules on (a) expertise, independence and impartiality; (b) transparency; (c) effectiveness; and (d) fairness.

First, the rules on *expertise, independence and impartiality* include that ADR officials must possess the necessary knowledge and skills in the field of alternative or judicial resolution of consumer disputes, as well as a general understanding of law; are appointed for a term of office of sufficient duration to ensure the independence of their actions, and are not liable to be relieved from their duties without just cause; are not subject to any instructions from either party or their representatives; are remunerated in a way that is not linked to the outcome of the procedure; and must disclose conflict of interests.

Second, the rules on *transparency* include that ADR bodies must publish details of their expertise, impartiality and independence; the types of disputes handled; procedural rules governing the resolution of a dispute; the costs, if any, to be borne by the parties; the average length of the ADR procedure; and the legal effect of the outcome of the ADR procedure. Further, ADR bodies must also publish annual reports, including the number of disputes received, average time taken to resolve disputes, and rate of compliance, if known.

Third, the rules on *effectiveness* include that the ADR procedures must be available and easily accessible online and offline to both parties; parties must have access to the procedure without being obliged to retain a lawyer or a legal advisor; ADR procedures must be free of charge or available at a nominal fee for consumers; and the outcome of the ADR procedure must be made available within a period of 90 days.

Finally, the rules on *fairness* include that parties have the possibility of expressing their point of view; may seek independent advice or be represented or assisted by a third party at any stage of the procedure; parties are notified of the outcome of the ADR procedure in writing, and are given a statement of the grounds on which the outcome is based. Importantly, parties must, before agreeing to or following a proposed solution, be informed of the legal effect of such a proposed solution, and before expressing their consent to a proposed solution or amicable agreement, be allowed a reasonable period of time to reflect.

Importantly, the directive does not apply to dispute resolution mechanisms where the ADR officials are “employed or remunerated exclusively by an individual trader,” but the directive does include a clause that member states can bring such procedures under the directive, where certain requirements are satisfied, including the “specific requirements of independence and transparency.” Further, the directive does not apply to consumer complaint-handling systems operated by traders, and attempts made by a judge to settle a dispute in the course of a judicial proceeding concerning that dispute.

Of particular note, the directive contains rules on the binding and non-binding natures of ADR agreements. An agreement between a consumer and a trader to submit complaints to an ADR entity is not binding on the consumer if it was concluded before the dispute has materialised, and if it deprives the consumer of their right to bring an action before the courts to settle the dispute. Further, the solution imposed may be binding on the parties only if they were informed of its binding nature

in advance and specifically accepted this. As such, any proposed e-court model that included an ADR mechanism would need to be consistent with the ADR Directive's rules.

b. European Online Dispute Resolution Regulation

The EU has also enacted the Online Dispute Resolution Regulation.⁶⁶ This Regulation established a European Online Dispute Resolution platform, which would facilitate independent, impartial, transparent, effective, fast and fair out-of-court resolution of disputes between consumers and traders online. The Regulation required the European Commission to establish such a platform, designed to be a “single point of entry for consumers and traders seeking the out-of-court resolution” of certain disputes, and free of charge. It applies to out-of-court resolution of disputes concerning contractual obligations stemming from online sales or service contracts. In 2016, the Commission established the European Online Dispute Resolution platform. In 2017, over 24,000 consumer complaints were lodged with the platform.⁶⁷ Over a third of the complaints concerned cross-border purchases within the EU; and most complaints concerned clothing and footwear, airline tickets and information and communication technology goods.

The ODR Regulation sets down the rules and procedure for the operation of the platform. First, a consumer can submit an electronic complaint form; and the ODR platform automatically notifies the company, which has a period of 10 days to reply, of the request. Second, the ODR platform facilitates the exchange of messages directly through a dashboard, and permits sending attachments such as product photos, and scheduling an online meeting. Third, parties have 30 days from the submission of the complaint to agree on a dispute resolution body; and the ODR platform will provide recommended dispute resolution bodies. However, the parties may choose outside of this recommended list. Once the parties have agreed on a dispute resolution body, the ODR platform will provide the ODR body with the consumer's complaint. The body then has three weeks to inform the parties whether it accepts jurisdiction to handle the complaint. Once the dispute resolution body has accepted jurisdiction, it will handle the complaint in line with its procedures, and must offer a suggested solution with 90 days. The final decision of the body will be available on the ODR platform. Whether or not the suggested solution is legally binding on the parties depends upon the rules of the particular dispute resolution body handling the complaint.

The European Online Dispute Resolution platform is an example of an online dispute resolution mechanism established by legislation, and facilitating online negotiation, and independent dispute resolutions. There is a network of national online dispute resolution bodies as part of the platform, and EU member states are required to designate at least one such body. A European social media dispute resolution platform could be created along these lines to facilitate dispute resolution between users and social media companies.

c. European Small Claims Procedure

In addition to online alternative dispute resolution mechanisms, the EU has enacted a European Small Claims Procedure Regulation,⁶⁸ which was designed to simplify and speed up litigation concerning small claims in cross-border cases, and to reduce costs. It was designed with a number of features: (a) an online procedure that does not require a physical presence in court; (b) eliminates the intermediate proceedings necessary to enable recognition and enforcement of judgments from another member

state; (c) is an alternative to national procedures; and (d) a decision given in the European Small Claims Procedure is recognised and enforceable in all member states. The regulation applies to cross-border cases involving civil and commercial matters, where the value of a claim does not exceed 5,000 euro. Notably, the regulation does not apply to “violations of privacy and of rights relating to personality, including defamation.”

The European Small Claims Procedure is now available on the EU’s European e-Justice online platform,⁶⁹ and operates as follows. To start the procedure, a claim form is sent, with relevant supporting documents, such as receipts, invoices, etc., to the national court with jurisdiction. Member states must ensure the procedure is accessible through relevant national websites. Once the court receives the application, it completes a certain form, and within 14 days the court should serve a copy of it, along with the form, on the defendant. The defendant has 30 days to reply, by filing another form. The court must then send a copy of any reply to the plaintiff within 14 days.

Within 30 days of receiving the defendant’s answer, the court must either give a judgment on the small claim, or summon the parties to an oral hearing. The Regulation provides that this oral hearing should use any appropriate distance communication technology, such as videoconference or teleconference, unless it would be unfair. Further, a party summoned to be physically present at an oral hearing may request the use of distance communication technology, and it is not necessary to be represented by a lawyer. The court is required to use the “simplest and least burdensome method of taking evidence” and the court “shall not require the parties to make any legal assessment of the claim.” Whenever appropriate, the court must seek to reach a settlement between the parties. Finally, once a decision is reached, a certificate is issued by the court, and the judgment is enforceable in all member states, without any further formalities.

The European Small Claims Procedure is an e-court model based on online judicial adjudication, and established by legislation. Social media disputes could follow a similar online procedure, with online judicial decision-making, and the imposition of time limits.

d. Business-to-Platform Regulation

It may also be helpful to refer to EU legislation applicable to disputes between certain online platforms and their *business* users, called the Business-to-Platform Regulation 2019.⁷⁰ While the regulation is only applicable to platforms offering online marketplaces (e.g., Amazon, eBay, and mobile app stores) it is still relevant to some questions on content moderation decisions.

First, it has rules on platform *terms and conditions*, including that the rules must clearly set out the grounds for decisions to suspend, terminate, or restrict a business user’s account. Platforms must give business users notice of any planned changes to the terms and conditions. Second, there are detailed rules on *account suspension and terminations*, including the requirement that when a platform decides to suspend or terminate a business user account, it must provide a statement of reasons for that decision. Third, platforms must establish *internal complaint-handling systems*, including the obligation to provide individualised decisions drafted in plain and intelligible language. Fourth, and importantly, the regulation requires platforms to identify *impartial and independent mediators* in their terms and conditions which they are willing to engage for out-of-court dispute resolution. Article 12 sets down rules for these mediators, including that they must be impartial and independent, affordable for business users,

easily accessible, and have sufficient understanding of general business-to-business commercial relations, enabling them to contribute effectively to dispute settlement.

The Business-to-Platform Regulation is the first piece of sector-specific regulation for online platforms. While only applicable to business users, it attempts to bring fairness and transparency for these business users. The regulation's rules on clear terms and conditions, internal complaint-handling, and account suspension could serve as a helpful model for best practices for social media companies. The regulation's rules on independent dispute resolution could also be used for best practices. For example, social media companies could recognise third-party online dispute resolution bodies they would be willing to engage to resolve content moderation disputes.

2. Case study on e-courts in the Netherlands

The term e-court covers a broad spectrum of online dispute resolution in the Netherlands, including private and public initiatives. As regards public initiatives, e-courts are often a way of digital litigation in national courts, for example the former "eKantonrechter."⁷¹ Private initiatives fall under alternative dispute resolution and are usually a form of arbitration.⁷² The best-known private Dutch e-courts are Stichting e-Court,⁷³ Stichting DigiTrage⁷⁴ and Stichting Arbitrage Rechtspraak Nederland.⁷⁵

E-courts began in the Netherlands in 2009 with the Stichting e-Court.⁷⁶ It was followed by Stichting Arbitrage Rechtspraak Nederland in 2012,⁷⁷ the eKantonrechter in 2013⁷⁸ and Stichting DigiTrage in 2014.⁷⁹ All these initiatives had the goal to lower the threshold for dispute resolution by introducing a simplified, faster and (as for private initiatives) less costly procedure. However, only the e-courts of Stichting DigiTrage and Stichting Arbitrage Rechtspraak Nederland are still in use. The eKantonrechter was terminated in 2018 by the Dutch Council for the Judiciary because it "would no longer fit well within the ambition to simplify."⁸⁰ Also, Stichting e-Court has been stalled since the beginning of 2018 due to negative reporting⁸¹ and related lawsuits.⁸²

The Dutch Code of Civil Procedure provides a national legal basis for e-courts' provisions on arbitration.⁸³ In these cases, the arbitration institutions will be often included in the general terms and conditions as agreed between parties.⁸⁴ An opponent can opt out within (at least) one month.⁸⁵ To enforce the decision, an instrument permitting the enforcement issued by a national court (an *exequatur*) is still required.⁸⁶ The national court examines the arbitration decision summarily.⁸⁷ No regulatory body is involved. The Netherlands Arbitration Institute⁸⁸ promotes alternative dispute resolution by, *inter alia*, publishing draft texts for arbitration clauses.⁸⁹ The eKantonrechter programme operated by courts also found its legal basis in the Dutch Code of Civil Procedure.⁹⁰

Online dispute resolution initiatives in the Netherlands differ in the types of disputes that they address. Stichting e-Court handles mainly debt collection cases of the largest Dutch insurance companies (more than 20,000 cases in 2017).⁹¹ Five out of six decisions published by the court concern unpaid contributions for health insurance.⁹² Stichting DigiTrage focusses on debt collection cases of small and medium-size enterprises.⁹³ The eKantonrechter was allowed to handle essentially (uncomplicated) civil matters up to 25,000 euro (approximately 27,800 USD) and labour and rental matters,⁹⁴ but has only handled fourteen cases in a four-year period.⁹⁵ One case involved claiming damages for "non-conformity" in the sale of a house.⁹⁶ It is unclear which specific type of cases Stichting Arbitrage Rechtspraak Nederland deals with.

The general procedure at e-courts is as follows. First, the party seeking redress submits an application on an online platform. Then, the defendant lodges a statement of defence on the same platform. A round of online adversarial hearings occurs. If necessary, an oral hearing can be held (online or offline), and witnesses summoned. Finally, judgment will be rendered. Each body can deviate from this procedure.⁹⁷ Occasionally, appeal against the arbitration decision is possible.⁹⁸

Compared with regular courts, e-courts generally offer shorter deadlines and faster procedures, lower and clearer costs, more convenience (litigation from home and the possibility to access your dossier anytime and anywhere) and the possibility to litigate in comprehensible language. The main difficulty of (private) e-courts is that they lack transparency. Decisions are rarely published, and a list of arbitrators is not always available.⁹⁹ Another major concern is that e-courts financially depend on large parties which often use this type of dispute resolution, such as large insurance companies. This can influence negatively the equality of arms, which requires a fair balance between parties.¹⁰⁰ Another difficulty especially relating to Stichting e-Court is that arbitrators rule in absentia in favour of the applicant.¹⁰¹ Besides, use of the term “e-court” for arbitration purposes can be misleading.

Multiple questions arise about Dutch e-court initiatives, particularly related to the essential criteria mentioned above, and the requirements of the ADR Directive. First, all the e-courts claim to be independent. However, it is not clear whether there is an arbitrators’ term of office and whether this term is sufficiently long to guarantee their independence, which raises questions under the ADR Directive’s rules on expertise, independence, impartiality, and transparency. Second, there can be an appearance of partiality because the arbitration institution depends financially, albeit indirectly, on professional parties that include the institution in their general terms and conditions.¹⁰² Besides, a lot of arbitrators are attorneys. It can be questioned whether they have (without specific training) the appropriate skills to judge impartially. This again raises issues under the rules on expertise, independence and impartiality in the ADR Directive. Third, many e-court initiatives lack transparency as decisions are never or rarely published, and a list of arbitrators is not always available. None of the websites contain annual reports, which is required under transparency rules.

Fourth, e-courts seem effective, as they are simple, fast and relatively cheap to use, and assistance by a lawyer is not mandatory. However, not all procedures are also available offline, which is required by the rules on effectiveness. Finally, professional parties use an arbitration institution more often than a consumer. This can have a negative effect on the equality of arms. Apart from that, there are concerns about whether consumers are properly informed that the arbitration institution is not a regular court. This could constitute an infringement of the rules on fairness.

3. Case study on e-courts in North America

There are some excellent examples in North America of e-court initiatives that involve courts adopting online case resolution systems, where individuals with minor disputes engage through an online platform with police, prosecutors, and judges. First is Michigan’s Online Case Review programmes for resolving traffic disputes, warrant disputes and misdemeanours.¹⁰³ It comprises online platforms for defendants to submit their cases, including all arguments or explanations, such as why they cannot pay their fines. It allows police and prosecutors to review cases before a judge makes a decision. The online format provides for the resolution of traffic disputes without in-person court appearances.

There are further incentives for individuals, under which resolutions may be reached with prosecutors, resulting in, for example, no driver licence points.

Second, some e-court initiatives go beyond the digitisation of court systems (electronic filings, online hearings, etc.), and involve online negotiation, mediation, arbitration, and adjudication. One leading example is the Civil Resolution Tribunal in British Columbia, which is “Canada’s first online tribunal,” and was established under British Columbia’s Civil Resolution Tribunal Act 2012.¹⁰⁴ The purpose of the Civil Resolution Tribunal is to encourage early resolution of disputes through online dispute resolution, while preserving adjudication as a last resort. Under the act, the tribunal has authority to resolve certain small claims, up to a value of CAD\$5,000 (it has no jurisdiction over claims involving libel); and also provides that certain small claims *must* go through the tribunal before going to provincial court.

The tribunal process follows a stepped online-dispute resolution process. First, a free, anonymous, and confidential online platform helps complainants assess their problem and decide the best option for how to proceed. Second, if the user cannot resolve the issue through the Solution Explorer platform, the process moves to an online dispute resolution portal, which begins with party-to-party negotiation. It is an online guided and structured negotiation. While it is an independent process, tribunal staff monitor the negotiations and are available to provide case-specific suggestions and support. Third, if there no is agreement, the process moves to the facilitated settlement phase. This involves neutral tribunal case managers who are dispute resolution experts and discuss the issues and settlement options. Under the act, if the parties reach a facilitated agreement, they can apply to the tribunal for a consent resolution order. If parties still cannot reach a solution, the claim proceeds to resolution by tribunal hearing. An independent tribunal member will make a decision about the dispute. If hearings are not needed, the tribunal member may render a decision based solely on digital evidence and submissions. The Act has detailed rules on enforcement of tribunal orders, and judicial review of these decisions.

Although not currently operational, it is worth mentioning a similar initiative, but involving members of the judiciary, rather than tribunal officials. The UK Civil Justice Council’s advisory group has recommended the creation of an e-court system (HM Online Courts), designed to reduce the need for judges in many cases.¹⁰⁵ There would be three tiers: the first tier would be online evaluation, to help users with grievances evaluate their problems, and understand both their entitlements and the options available to them. This would be a form of information and diagnostic service and would be available at no cost to court users. The second tier would be online facilitation, which would involve trained and experienced facilitators, working online, who can review papers and statements from parties, and then help by mediating, advising, or encouraging them to negotiate. This stage would be designed to bring many if not most disputes to speedy, fair conclusions without the involvement of judges; and users would incur a court fee. Then tier three would be online judges, where cases would be decided on an online basis, largely on the basis of papers submitted to them electronically, as part of a structured but still adversarial system of online pleading and argument. The decisions of online judges would be binding and enforceable, enjoying the same status as decisions made by judges in traditional courtrooms. A court fee would be payable, but much lower than normal court fee.

Notes

¹ Dr. Heidi Tworek is assistant professor at the University of British Columbia as well as a non-resident fellow at the German Marshall Fund of the United States and the Canadian Global Affairs Institute. Her book *News from German: The Competition to Control World Communications, 1900-1945*, was published in 2019 by Harvard University Press.

² Dr. Ronan Ó Fathaigh is a researcher at the Institute for Information Law (IViR), University of Amsterdam, whose doctoral dissertation at Ghent University was on freedom of expression and the chilling effect.

³ Lisanne Bruggeman participated in the Glushko & Samuelson Information Law and Policy Lab at the Institute for Information Law (IViR), University of Amsterdam, and is completing the master's programme in information law at the University of Amsterdam.

⁴ Dr. Chris Tenove is a postdoctoral research fellow at the University of British Columbia who studies political theory and international relations, with a focus on issues of democracy, public policy, global governance, and digital politics.

⁵ Josh Constine, "Instagram will let you appeal post takedowns," *TechCrunch*, 9 May 2019, <https://techcrunch.com/2019/05/09/instagram-vaccine-hashtags/>

⁶ Jillian C. York and Corynne McSherry, "Content Moderation is Broken. Let Us Count the Ways," Electronic Frontier Foundation, 29 April 2019, <https://www.eff.org/deeplinks/2019/04/content-moderation-broken-let-us-count-ways>.

⁷ Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Human Rights Council, A/HRC/38/35, 6 April 2018, para. 38, https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/38/35.

⁸ Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Human Rights Council, A/HRC/38/35, 6 April 2018, para. 72.

⁹ On platform governance, see the series of essays, "Models for Platform Governance," Centre for International Governance Innovation, 29 October 2019, <https://www.cigionline.org/publications/models-platform-governance>.

¹⁰ Facebook, Community Standards Enforcement Report, May 2019 - Hate Speech - How much content did we restore after removing it?, <https://transparency.facebook.com/community-standards-enforcement#hate-speech>.

¹¹ Jonathan Vanian, "Google's Hate Speech Detection A.I. Has a Racial Bias Problem," *Fortune*, 16 August 2019, <https://fortune.com/2019/08/16/google-jigsaw-perspective-racial-bias/>.

¹² For empirical analysis of the chilling effect, see Jonathon W. Penney, "Internet surveillance, regulation, and chilling effects online: a comparative case study" (2017) 6(2) *Internet Policy Review*, DOI:10.14763/2017.2.692. For a European discussion, see Ronan Ó Fathaigh, *Article 10 and the Chilling Effect: A critical examination of how the European Court of Human Rights seeks to protect freedom of expression from the chilling effect* (PhD Thesis, Ghent University, 2019), <https://biblio.ugent.be/publication/8620369>.

¹³ See *Dink v. Turkey* (App. nos. 2668/07, 6102/08, 30079/08, 7072/09 and 7124/09) 14 September 2010, para. 137; Human Rights Committee, General comment No. 34, CCPR/C/GC/34, 12 September 2011, para. 7; and Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Human Rights Council, A/HRC/38/35, 6 April 2018, para. 6. See Brittan Heller and Joris van Hoboken, "Freedom of Expression: A Comparative Summary of United States and European Law," working paper of the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression (2019), https://www.ivir.nl/publicaties/download/TWG_Freedom_of_Expression.pdf.

¹⁴ As stated in the memo presented by European Court of Human Rights Judge Róbert Spanó during the Bellagio meeting in November 2019. Other parts of the memo are reflected in the paragraph and elsewhere in the paper.

¹⁵ Heidi Tworek, "How Transparency Reporting Could Incentivize Irresponsible Content Moderation," Centre for International Governance Innovation, 10 December 2019, <https://www.cigionline.org/articles/how-transparency-reporting-could-incentivize-irresponsible-content-moderation>.

¹⁶ Amy J. Schmitz, "Expanding Access to Remedies through E-Court Initiatives" (2019) 67 *Buffalo Law Review* 89.

¹⁷ More input on the criteria of independence can be derived from: K. Irion, W. Schulz, W., P. Valcke, *The Independence of the Media and Its Regulatory Agencies: Shedding New Light on Formal and Actual Independence Against the National Context*, Intellect, Bristol UK / Chicago USA, 2014.

¹⁸ We're not discussing the definition of self-regulation (or co-regulation) as such. There is rich literature on this: e.g., Michael Latzer, Natascha Just, & Florian Saurwein, "Self- and co-regulation: Evidence, legitimacy and governance choice," in Monroe E. Price and Stefaan Verhulst, eds., *Routledge Handbook of Media Law*, New York, 2013, 373-398.

¹⁹ Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Human Rights Council, A/HRC/38/35, 6 April 2018, para. 25.

²⁰ The Twitter Trust and Safety Council, https://about.twitter.com/en_us/safety/safety-partners.html#Trust-council.

²¹ Jane Wakefield, "Google's ethics board shut down," *BBC News*, 5 April 2019, <https://www.bbc.com/news/technology-47825833>.

²² Facebook, "Establishing Structure and Governance for an Independent Oversight Board," 17 September 2019, <https://newsroom.fb.com/news/2019/09/oversight-board-structure>.

²³ YouTube Trusted Flagger program, <https://support.google.com/youtube/answer/7554338?hl=en>.

²⁴ See Wikipedia: Arbitration Committee, https://en.wikipedia.org/wiki/Wikipedia:Arbitration_Committee. See also Corinne Ramey, "The 15 People Who Keep Wikipedia's Editors From Killing Each Other," *The Wall Street Journal*, 7 May 2018, <https://www.wsj.com/articles/when-wikipedias-bickering-editors-go-to-war-its-supreme-court-steps-in-1525708429>; and Piotr Konieczny, "Decision making in the self-evolved collegiate court: Wikipedia's Arbitration Committee and its implications for self-governance and judiciary in cyberspace," (2017) 32(6) *International Sociology* 755.

²⁵ Global Network Initiative, Our Mission, <https://globalnetworkinitiative.org/team/our-mission>.

²⁶ Internet Watched Foundation, <https://www.iwf.org.uk/become-a-member/join-us/our-members>.

²⁷ See, for example, Kristina Irion et al., *The independence of media regulatory authorities in Europe* (European Audiovisual Observatory, 2019), <https://rm.coe.int/the-independence-of-media-regulatory-authorities-in-europe/168097e504>.

²⁸ See Andrei Richter, *Disinformation in the media under Russian law* (European Audiovisual Observatory, 2019), p. 12, <https://rm.coe.int/disinformation-in-the-media-under-russian-law/1680967369>.

²⁹ Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Human Rights Council, A/HRC/38/35, 6 April 2018, para. 72.

³⁰ ARTICLE 19, *Self-regulation and 'hate speech' on social media platforms* (2018), <https://www.article19.org/resources/self-regulation-hate-speech-social-media-platforms/>.

³¹ Chris Tenove, Heidi Tworek, and Fenwick McKelvey, *Poisoning Democracy: How Canada Can Address Harmful Speech Online* (2018), <https://ppforum.ca/publications/poisoning-democracy-what-can-be-done-about-harmful-speech-online/>.

³² Mark Zuckerberg, "A Blueprint for Content Governance and Enforcement," Facebook, 15 November 2018, https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/?hc_location=ufi.

³³ Stanford Global Digital Policy Incubator, Article 19, and UN Special Rapporteur on Freedom of Opinion and Expression, *Social Media Councils: From Concept to Reality - Conference Report* (2019), <https://fsi.stanford.edu/content/social-media-councils-concept-reality-conference-report>.

³⁴ ARTICLE 19, *The Social Media Councils: Consultation Paper* (Article 19, 2019), <https://www.article19.org/resources/social-media-councils-consultation/>.

³⁵ ARTICLE 19, *The Social Media Councils: Consultation Paper* (2019), <https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf>.

³⁶ Heidi Tworek, *Social Media Councils*, Centre for International Governance Innovation, October 2019, <https://www.cigionline.org/articles/social-media-councils>.

³⁷ Jacinda Ardern, Prime Minister of New Zealand, "Significant progress made on eliminating terrorist content online," 24 September 2019, <https://www.beehive.govt.nz/release/significant-progress-made-eliminating-terrorist-content-online>.

-
- ³⁸ Brent Harris, “Establishing Structure and Governance for an Independent Oversight Board,” Facebook, 17 September 2019, <https://newsroom.fb.com/news/2019/09/oversight-board-structure>.
- ³⁹ Facebook, Oversight Board Charter, September 2019, p. 4, https://fbnewsroomus.files.wordpress.com/2019/09/oversight_board_charter.pdf.
- ⁴⁰ Ibid.
- ⁴¹ Ibid., p. 7.
- ⁴² Facebook, “Establishing Structure and Governance for an Independent Oversight Board,” 17 September 2019, <https://newsroom.fb.com/news/2019/09/oversight-board-structure/>.
- ⁴³ For further suggestions, see Heidi Tworek, “Social Media Councils,” Centre for International Governance Innovation, 2019, <https://www.cigionline.org/articles/social-media-councils>.
- ⁴⁴ Canadian Radio-television and Telecommunications Commission, *Harnessing Change: The Future of Programming Distribution in Canada*, 2018, <https://crtc.gc.ca/eng/publications/s15/>.
- ⁴⁵ Mandhane, Renu, and Marie-Claude Landry, “Addressing Discriminatory Advertising on Facebook in Canada,” Ontario Human Rights Commission, 7 June 2019, http://www.ohrc.on.ca/en/news_centre/addressing-discriminatory-advertising-facebook-canada.
- ⁴⁶ Amy J. Schmitz, “Expanding Access to Remedies through E-Court Initiatives” (2019) 67 *Buffalo Law Review* 89, 92.
- ⁴⁷ Civil Justice Council’s Online Dispute Resolution Advisory Group, *Online Dispute Resolution for Low Value Civil Claims* (2015), <https://www.judiciary.uk/wp-content/uploads/2015/02/Online-Dispute-Resolution-Final-Web-Version.pdf>.
- ⁴⁸ Schmitz, 100.
- ⁴⁹ The Office of Communications Annual Report & Accounts For the period 1 April 2018 to 31 March 2019, p. 149, https://www.ofcom.org.uk/_data/assets/pdf_file/0024/156156/annual-report-18-19.pdf.
- ⁵⁰ Standing Committee on Access to Information, Privacy and Ethics, *Democracy Under Threat: Risks and Solutions in the Era of Disinformation and Data Monopoly*, (Committee Report 17. Ottawa, ON: House of Commons, Canada, 2018) <http://www.ourcommons.ca/DocumentViewer/en/42-1/ETHI/report-17>. See also (Government of Canada 2019). While some groups welcomed this, others have expressed concern about this vague aspiration (see, for instance, Zwibel 2019).
- ⁵¹ Liberal Party of Canada, “Forward: A Real Plan for the Middle Class,” Federal Liberal Agency of Canada, 2019, p. 47, <https://2019.liberal.ca/wp-content/uploads/sites/292/2019/09/Forward-A-real-plan-for-the-middle-class.pdf>.
- ⁵² E.g. Mack Lamoureux, “YouTube Pulls Canadian Anti-Islam Vlogger Following Huge Defamation Loss,” *Vice*, 14 May 2019, https://www.vice.com/en_ca/article/597ddk/youtube-pulls-canadian-anti-islam-vlogger-kevin-j-johnston-following-record-defamation-lawsuit-loss
- ⁵³ Chris Tenove, Chris, Heidi J. S. Tworek, and Fenwick McKelvey, “Poisoning Democracy: How Canada Can Address Harmful Speech Online,” 2018, Ottawa, ON: Public Policy Forum, <https://www.ppforum.ca/wp-content/uploads/2018/11/PoisoningDemocracy-PPF-1.pdf>.
- ⁵⁴ National News Media Council, “Frequently Asked Questions,” <https://mediacouncil.ca/frequently-asked-questions/>.
- ⁵⁵ Canadian Broadcast Standards Council, <https://www.cbcs.ca/about-us/>.
- ⁵⁶ Canadian Radio-television and Telecommunications Commission, *Harnessing Change: The Future of Programming Distribution in Canada*, 2018, <https://crtc.gc.ca/eng/publications/s15/>.
- ⁵⁷ Mandhane, Renu, and Marie-Claude Landry, “Addressing Discriminatory Advertising on Facebook in Canada,” Ontario Human Rights Commission, 7 June 2019, http://www.ohrc.on.ca/en/news_centre/addressing-discriminatory-advertising-facebook-canada.
- ⁵⁸ Section 13 of the Act barred individuals or groups from using telecommunications – later expanded to include the internet – to distribute messages likely to expose a person or group to hatred or contempt based on their race, ancestry, religion, and other characteristics.
- ⁵⁹ Report of the Standing Committee on Justice and Human Rights, “Taking Action to End Online Hate,” June 2019, <https://www.ourcommons.ca/Content/Committee/421/JUST/Reports/RP10581008/justrp29/justrp29-e.pdf>

-
- ⁶⁰ Alliance of Independent Press Councils in Europe, <http://www.aipce.net/aboutAipce.html>.
- ⁶¹ See Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Human Rights Council, A/HRC/38/35, 6 April 2018, para. 54.
- ⁶² See, for example, Lara Fielden, *Regulating the Press: A Comparative Study of International Press Councils* (Reuters Institute for the Study of Journalism, 2012), <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-11/Regulating%20the%20Press.pdf>.
- ⁶³ *Ibid.*, p. 15.
- ⁶⁴ See, eg., IMPRESS. Arbitration, <https://www.impress.press/regulation/arbitration.html>.
- ⁶⁵ Directive 2013/11/EU on alternative dispute resolution for consumer disputes (ADR Directive).
- ⁶⁶ Regulation (EU) No 524/2013 on online dispute resolution (ODR Regulation).
- ⁶⁷ See the European Online Dispute Procedure, <https://ec.europa.eu/consumers/odr/main/?event=main.home.howitworks>.
- ⁶⁸ Regulation (EC) No 861/2007 of the European Parliament and of the Council of 11 July 2007 establishing a European Small Claims Procedure – consolidated text of 14 June 2017, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02007R0861-20170714>.
- ⁶⁹ https://e-justice.europa.eu/content_small_claims-42-en.do.
- ⁷⁰ Regulation (EU) 2019/1150 on promoting fairness and transparency for business users of online intermediation services (Business-to-Platform Regulation).
- ⁷¹ See, Nu ook online rechtspraak voor burgers bij de eKantonrechter, 3 June 2014, <https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Nieuws/Paginas/Nu-ook-online-rechtspraak-voor-burgers-bij-de-eKantonrechter.aspx>; and see also: J.W. Langelaar, ‘De eKantonrechter’, *Tijdschrift voor de Procespraktijk* 2015-3.
- ⁷² Another type of alternative dispute resolution e-courts in certain cases deal(t) with is “binding advice.”
- ⁷³ <http://www.e-court.nl/>; see also: D.E. Thiescheffer, *E-court naast overheidsrechtspraak: Online rechtspraak als alternatieve geschillenbeslechting en de garanties voor consumenten* (Celsus juridische uitgeverij, 2018).
- ⁷⁴ Digitage, “Digitale arbitragevoor incassozaken,” <https://www.digitrage.nl/>.
- ⁷⁵ Stichting Arbitrage Rechtspraak Nederland, <https://www.arbitragerechtspraak.nl/>.
- ⁷⁶ “Nieuwe rechtbank opent haar deuren,” 12 January 2010, <http://www.e-court.nl/persberichten/>.
- ⁷⁷ Trade register of the Dutch Chamber of Commerce (consulted on 26 September 2019).
- ⁷⁸ “Nu ook online rechtspraak voor burgers bij de eKantonrechter,” 3 June 2014, <https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Nieuws/Paginas/Nu-ook-online-rechtspraak-voor-burgers-bij-de-eKantonrechter.aspx>
- ⁷⁹ Emerce, “Digitrage gestart met online geschillenbeslechting,” 2 February 2014, <https://www.emerce.nl/wire/digitrage-gestart-online-geschillenbeslechting>
- ⁸⁰ “Reset digitalisering van de Rechtspraak,” 10 April 2018, <https://www.rechtspraak.nl/SiteCollectionDocuments/2018-brief-reset-digitalisering.pdf>, p. 5.
- ⁸¹ See, Karlijn Kuijpers, Thomas Muntz and Tim Staal, “Vonnis te koop,” *De Groene Amsterdammer*, 17 January 2018, <https://www.groene.nl/artikel/vonnis-te-koop>; Sociaal Werk Nederland, “Rechtspraak op bestelling?! Stop commerciële rechtspraak,” January 2018, https://schuldingo.nl/fileadmin/Publicaties/Rechtspraak_op_bestelling_Stop_commerciële_rechtspraak.pdf; and the responses of Stichting e-Court: <http://www.e-court.nl/wp-content/uploads/2018/01/Persbericht-2018-01-19-reactie-publiciteit.pdf>; <http://www.e-court.nl/wp-content/uploads/2018/01/Reactie-LOSR-2018-01-22.pdf>.
- ⁸² See: e-Court, e-Court naar het Europese Hof van Justitie, 16 February 2018, <http://www.e-court.nl/wp-content/uploads/2018/02/Persbericht-2018-02-16-e-Court-naar-EHvJ.pdf>; and e-Court, “Dutch ‘e-Court’ takes

Supreme Court justices and other magistrates to Court,” 16 November 2018, http://www.e-court.nl/wp-content/uploads/2018/11/Persbericht-2018-11-16-RvdR_Eng.pdf.

⁸³ Articles 1020-1076 Wetboek van Burgerlijke Rechtsvordering.

⁸⁴ See also: Article 1020 Wetboek van Burgerlijke Rechtsvordering.

⁸⁵ Article 6:236, paragraph n, Burgerlijk Wetboek.

⁸⁶ Article 1062, paragraph 1, Wetboek van Burgerlijke Rechtsvordering.

⁸⁷ Article 1063, paragraph 1, Wetboek van Burgerlijke Rechtsvordering.

⁸⁸ Netherlands Arbitration Institute, <https://www.nai-nl.org/en/>.

⁸⁹ Netherlands Arbitration Institute, “Clauses,” <https://www.nai-nl.org/en/documents/clauses/>.

⁹⁰ Article 96 Wetboek van Burgerlijke Rechtsvordering.

⁹¹ “Robotrechter e-Court reageert: ‘De rechterlijke macht werkt ons tegen,’” *Het Financieele Dagblad*, 28 January 2018, <https://fd.nl/economie-politiek/1239049/robotrechter-e-court-reageert-de-rechterlijke-macht-werkt-ons-tegen>.

⁹² See decision numbers 21-04-2017 IPVRW 24009030000, 30-06-2017 IPVRW 24363052000, 17-11-2017 IPVRW 24562467000, 17-11-2017 IPVRW 24590256000 and 29-09-2017 IPVRW 24127585000; available here: <http://www.e-court.nl/uitspraken>.

⁹³ DigiTrage, “Geschil Aanmelden,” <https://www.digitrage.nl/geschil-aanmelden.html>; and DigiTrage, “Bedrijven met DigiTrage-beding,” <https://www.digitrage.nl/bedrijven-met-digitrage-beding.html>.

⁹⁴ Article 96, paragraph 1, and article 93 Wetboek van Burgerlijke Rechtsvordering.

⁹⁵ “Samen naar de eKantonrechter kan niet meer,” *Mr.*, 29 May 2018, <https://www.mr-online.nl/samen-naar-de-ekantonrechter-kan-niet-meer/>.

⁹⁶ ECLI:NL:RBOBR:2014:1145, 13 March 2014, <https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBOBR:2014:1145>.

⁹⁷ For more information, see the procedural regulations on the institutions’ websites.

⁹⁸ Article 30 Procesreglement DigiTrage 01-01-2015; Article 16 Procesreglement e-Court 2017; Article 30 Arbitragereglement Stichting Arbitrage Rechtspraak Nederland 15-08-2018.

⁹⁹ Only six decisions have been published by Stichting e-Court: <http://www.e-court.nl/uitspraken/>. Only Stichting DigiTrage (<https://www.digitrage.nl/over-digitrage/de-organisatie.html>) and Stichting e-Court (<http://www.e-court.nl/persberichten/>) have published lists of arbitrators on their websites.

¹⁰⁰ See also: D.E. Thiescheffer, *E-court naast overheidsrechtspraak: Online rechtspraak als alternatieve geschilbeslechting en de garanties voor consumenten* (Celsus juridische uitgeverij, 2018), p. 81.

¹⁰¹ Article 12, paragraph 2, Procesreglement e-Court 2017, unless the arbitrator judges the claim to be unlawful or unfounded.

¹⁰² See also: C.N.J. de Vey Mestdagh and A. Kamphorst, “e-Court of de moderne Prometheus vs. het recht, een achterhoedegevecht?,” *Tijdschrift voor Internetrecht* (2018), p. 14.

¹⁰³ See, for example, 74th District Court Online Case Review, <https://www.baycounty-mi.gov/News/74th-District-Court-Online-Case-Review.aspx>.

¹⁰⁴ See Civil Resolution Tribunal, <https://civilresolutionbc.ca>; and the Civil Resolution Tribunal Act 2012, http://www.bclaws.ca/civix/document/id/complete/statreg/12025_01.

¹⁰⁵ Civil Justice Council’s Online Dispute Resolution Advisory Group, *Online Dispute Resolution for Low Value Civil Claims* (2015), p. 3.