ARTICLE 19's Submission
to the **UN Special Rapporteur's consultation on online content regulation**

December 2017

I. Introduction

ARTICLE 19 welcomes the UN Special Rapporteur's consultation on online content regulation. We find that the forthcoming 2018 Special Rapporteur's report on this issue is particularly timely given recent policy developments in this area.[1] More than ever, the principles of immunity from liability for third-party content are under threat. Equally, many Internet companies appear to be removing content more than ever before, often under threat of regulation from governments.[2] Meanwhile, the Internet companies have been inconsistent in applying content regulation to majority and minority groups; their inadequate, at times clumsy response to the issue of online abuse, is of concern.

In this submission, ARTICLE 19 seeks to respond to the questions raised in the consultation.

II. Companies' compliance with state laws

Lack of clarity and transparency when dealing with state requests

ARTICLE 19 finds that the content regulation practices of Internet companies on their platforms lack transparency and are inconsistent.

We have recently conducted an analysis of the Community Guidelines of four major Internet companies, namely Facebook, Twitter, YouTube and Google, in the area of terrorism/extremism, 'fake news' and privacy violations.[3] In essence, three different types of scenarios can be distinguished when looking at companies' compliance with content regulation laws and measures imposed by governments:

- **Companies dealing with "legal requests", i.e. court orders or orders from government** agencies/public authorities: In this scenario, companies are generally *required* to take down the notified content when receiving such requests from the relevant authorities. Relevant sources of information about companies' processes in dealing with "legal requests" for content removal can usually be found in their Transparency Report[4] or pages dedicated to Law Enforcement.[5] Companies' policies do not usually explain whether "legal requests" are dealt with on a fast-track basis or not. In some cases, it appears that companies have established as 'single contact point' in order to deal with requests from certain governments.[6]

This scenario must be distinguished from content that is reported by users as *allegedly* unlawful. Companies are generally *not* required to take this content down but expose themselves to the risk of liability if the user in question later decides to take them to court if they decide not to remove the content at issue.[7] This corresponds to the 'conditional immunity from liability' regulatory model.

- Companies acting on the basis of their Terms of Service at the request of law **enforcement or other 'trusted flaggers'**: This is the main process used in relation to 'extremist' content, particularly in circumstances where law enforcement agencies do not have the power to order the removal of such content. In practice, this means that companies are *not* required to takedown content flagged as 'extremist': it is at their own discretion under their Terms of Service. Given the way in which 'extremist', 'violent extremist' and 'terrorist' content are increasingly used interchangeably, however, ARTICLE 19 finds it reasonable to assume that the same mechanism is used to request the removal of 'terrorist' content, despite the fact that law enforcement would generally have a specific power to order such a removal.[8] Finally, it is important to note that over time, social media companies – such as Facebook or Twitter – have amended their community guidelines seemingly to match more closely the legal definition of 'publicly condoning', 'promoting' or 'glorifying' terrorism in force in some countries (e.g. in France). It is therefore possible that, at least in some areas, the enforcement of the law and the enforcement of companies' Terms of Service are becoming indistinguishable as far as the definition of prohibited content is concerned. It is unclear, however, whether the distinction does lead to differences in terms of internal processing of removal requests.

- Search **engine operators dealing with "**right to be forgotten**"** requests: this scenario primarily concerns search engines under data protection legislation in the European Union and countries that have followed a similar interpretation to the Court of Justice of the European Union in the *Google Spain* case.[9] Under data protection law, search engines are *required* to deal with "right to be forgotten" requests and have developed forms to that effect.[10] If search engines refuse to de-list links upon request, the individual concerned can apply to his/her national data protection authority to re-examine the request. If the data protection authority finds in favour of the data subject, the search engine is required to comply. However, questions arise as to the scope of de-referencing orders (see below for more details).

  In general, ARTICLE 19 finds that the legal procedure developed so far to deal with "right to be forgotten" requests has been significantly lacking in due process safeguards for the protection of the right to freedom of expression. In particular, there is no obligation to notify companies or individuals that their content has been de-listed. Therefore, in the vast majority of cases, those individuals or companies have no opportunity to challenge a decision made by a search engine or a data protection authority to de-list their content. In other words, the protection of freedom of expression entirely relies on the likes of Google to do the right thing and strike an appropriate balance between freedom of expression and data protection.

  This lack of procedural safeguards may partly be due to the guidelines initially developed by the Article 29 Data Protection Working Party on the implementation of the *Google Spain* judgment.[11] ARTICLE 19 finds that the Article 29 Data Protection Working Party has taken a particularly extreme approach to the protection of personal data, dismissing freedom expression concerns as overblown. The EU General Data Protection Regulation only partially remedies these shortcomings by providing that data controllers should take 'reasonable steps' to notify content providers that their content

has been de-listed under the "right to be forgotten".[12] However, as noted above, this falls short of an obligation to notify in circumstances where a de-listing decision may infringe the right to freedom of expression. ARTICLE 19 has detailed our recommendations for an appropriate balance between freedom of expression and data protection in our policy *The Right to be Forgotten: Remembering Freedom of Expression*[13] and *The Global Principles on Freedom of Expression and Privacy*.[14]

ARTICLE 19 believes that the Special Rapporteur should reiterate in his report that companies should demonstrate their commitment to respect human rights. As such, they should challenge government and court orders that they consider to be in breach of international standards on freedom of expression and privacy. In practice, this means that as a matter of principle, companies should:

- Resist government legal requests to restrict content in circumstances where they believe that the request lacks a legal basis or is disproportionate; this includes challenging such orders before the courts;

- Resist individuals' legal requests to remove content in circumstances where they believe that the request lacks a legal basis or is disproportionate;

- Appeal court orders demanding the restriction of access to content that is legitimate under international human rights law;

- Moreover, as a matter of principle, companies should resist informal government requests to restrict content on the basis of their Terms of Service. In this regard, companies should make clear to both governments and private parties that they are under no obligation to remove content when a takedown request is made on the basis of their Terms of Service.

Lack of clarity and transparency under self-regulatory initiatives

In addition to the above, companies have developed a range of self-regulatory initiatives to deal with "terrorist" or "extremist" content as well as "fake news":

- "Terrorist/Extremist" content: since 2016, Facebook, Microsoft, Twitter and YouTube have formalised a partnership to combat 'terrorism' online through the Global Internet Forum to Counter Terrorism (GIFCT).[15] Under the partnership, companies agree to share a hash database, i.e. a database containing images/videos deemed terrorist under their Terms of Service. The images/videos have a digital fingerprint allowing participating companies to identify and remove matching content quickly from their own networks under their Terms of Service. Another partnership with the UN Security Council Counter-Terrorism Executive Directorate (UN CTED) and the ICT4Peace Initiative has also led to a broad knowledge-sharing network to:
  - engage with small companies;
  - develop best practices, including on 'online hate'; and
  - develop counter-speech initiatives.[16]

- "Fake news": Spreading "fake news" or "false information" is prohibited in some countries. We find that most major internet companies do not ban "fake news" *per se* in their Terms of Service but have a number of provisions banning spam, impersonation or bots (e.g. Facebook or Twitter). On YouTube, misleading metadata or tags can lead to content removal, whilst Google has recently strengthened its ad policies to prohibit "ads or destinations that intend to deceive users by excluding relevant information or giving misleading information about products, services, or businesses."[17]

In addition, in 2016 Facebook started working with fact-checking organisations in order to put in place a "fake news" labelling system.[18] Under this new initiative, readers are able to alert Facebook that a story might be false, if enough people report that story as fake, it is then sent to trusted third-party fact-checkers. If the story is deemed unreliable, it becomes publicly flagged as "disputed by third-party fact checkers" and a warning appears when users decide to share it. However, the initiative has been criticised for its lack of effectiveness.[19] Meanwhile, Facebook has explained that it has been hunting down "fake accounts" and worked with government and civil society partners to defend its platform from "malicious interference", particularly during elections.[20] It is also strengthening its enforcement of its ad policies[21] and continues to create new tools to help its users better understand the context of the articles in its news feed.[22]

Twitter has recently stepped up its crackdown on Russian "fake" accounts that allegedly meddled in the US election.[23] These efforts rely on the company's internal "systems" to detect "suspicious" activity on the platform, including suspicious accounts, tweets, logins and engagement. The company does not disclose how these "systems" are used in order to prevent "bad actors" from gaming the system.

YouTube has pledge to offer trainings to teenagers to help them identify "fake" videos.[24] More recently, it was reported that in the wake of the Las Vegas shooting, it was looking to change its search algorithm to produce more authoritative sources in response to searches. It remains unclear how YouTube determines which sources are more "authoritative."[25]

Other state requests

Companies' Terms of Service and policies do not make it clear whether special priority status is assigned to requests received from either State actors or so-called 'trusted flaggers'. In practice, it is known that requests made by 'trusted flaggers', whether governments or non-State actors such as intellectual property associations or anti-discrimination organisations, are fast-tracked. It is unclear whether companies treat different 'trusted flaggers' differently depending on their status as State or non-State actors or on the basis of the particular content at issue (e.g. priority being given to terrorist content over hate speech or intellectual property).

In general, it appears that States have encouraged companies to collaborate in counter-speech efforts, in the same way that they have sought to encourage free speech and other groups to engage in these efforts.

ARTICLE 19 suggests that the Special Rapporteur in his report highlights that companies decisions to block or remove content must follow the minimum standards set out in ARTICLE 19's policy on blocking, filtering, and free speech[26] and the Manila Principles on Intermediary Liability. In particular, companies' Terms of Service should comply with international standards on freedom of expression.

Global removals

ARTICLE 19 has identified several cases of "global takedown orders" where the search engine has been obligated to de-list the specific respondent URL addresses that contained

the objectionable material from its search results globally, beyond the territory in which the order has been issued. For example, such cases have been recently decided by courts in France[27] and in Canada.[28] The French case has been recently referred to the Court of Justice of the European Union.[29]

ARTICLE 19 believes that global orders in "right to be forgotten" cases are an inherently disproportionate interference with the right to freedom of expression in almost all but the most exceptional cases. The precedents set by the recent cases in Canada and France pose a significant threat to international human rights, including the international right to freedom of expression.[30]

For global removals, we recommend that the Special Rapporteur highlights that States should not oblige a search engine operator to remove the results displayed on all of the domain names used by its search engine worldwide. Search engine operators should only be required to de-reference results for searches made from within the State where a national court or data protection authority is satisfied that such a step is necessary and proportionate in all the circumstances.

III. Companies and individuals at risk

ARTICLE 19 finds that companies' Terms of Service do not adequately reflect the interests of users who face particular risks on certain grounds.

Company standards often include "hate speech" provisions that make reference to users who face particular risks. For instance, Facebook community guidelines provide that "Facebook removes hate speech, which includes content that directly attacks people based on their race, ethnicity, national origin, religious affiliation, sexual orientation, sex, gender or gender identity, or serious disabilities or diseases." However, it is unclear how the interests of these groups are taken into account in practice or if companies apply different considerations in relation to different groups (including taking into account context, language, religion, culture, politics etc.).[31] For instance, while Facebook goes on to specify that "People can use Facebook to challenge ideas, institutions and practices" and that humour, satire or social commentary related to hate speech is allowed, it does not give specific examples of the way in which these standards are applied in practice.

Moreover, overbroad policies in certain areas, combined with algorithmic bias,[32] have backfired, resulting in the censoring of groups at risks.

- **Terrorist or 'extremist' content**: most social media platforms have overbroad content policies on terrorist content. Under increasing pressure from governments, they have also increasingly deployed algorithms to flag and delete 'extremist' content. However, this can result in the removal of legitimate content by Muslim or other groups. For instance, Facebook recently closed the accounts of some Rohingya activists in Myanmar.[33] While it is reasonable to assume that this was a mistake, it seemingly reveals bias in the use of algorithms to target primarily Muslim 'extremism'. Similar concerns apply in relation to company policies that use official lists of proscribed organisations as a basis for the removal of content or closure of accounts. It is well-known that decisions to include particular groups on these lists can be intensely political. More generally, groups regarded as 'terrorist' in some countries are considered legitimate or regarded as 'freedom fighters' elsewhere. For instance, the US list of proscribed organisations includes the Kurdish Democratic Party (PKK) but this group is not included on the UN list.

- Nudity and gender: Similar issues have arisen in the context of policies on nudity. For instance, Twitter recently failed to return results when entering the hashtag 'gay', 'lesbian' or 'bisexual', among other LGBTI terms. Twitter later fixed the glitch but it appears that this was due to the application of filters to weed out pornography.[34] Equally, social media platforms are regularly criticised for the apparent bias with which they apply their policies, easily removing legitimate images containing nudity, e.g. mammograms or gay images,[35] but failing to tackle xenophobic content online.[36]

In addition, the willingness of Internet companies to strike deals with certain governments, for instance the creation of single points of contact in countries such as VietNam,[37] may well have significant ramifications for human rights defenders, journalists and others, including minority groups, depending on the policies of the government of the day.[38]

Finally, it is important to highlight that the availability of tools and policies enabling anonymity and/or pseudonymity online is particularly important as a form of protection for vulnerable groups. For instance, anonymity and pseudonymity might be the only way for some trans people to interact with others and access information relating to sexual orientation/gender identity. In this context, real name registration (or even requirements to provide identifying information privately) may chill expression and deter them from joining platforms. For this reason, the evolution of Facebook's real-name policy[39] is to be welcomed though it remains too limited.[40] Equally, the various user-controls developed by companies to deal with harassment, and a series of options including alternatives to requesting removal (e.g. to message the person explaining why the content is upsetting etc.) are positive steps that can help address the needs of particular communities.

IV. Content regulation processes

As ARTICLE 19 highlighted in the previous section, Internet companies rely on a number of different processes and tools to implement content restrictions:[41]

- Court or government orders: they require companies to remove content deemed illegal under national law;

- **'Trusted flagger' system**: this system has been developed by large internet companies such as Facebook or YouTube. 'Trusted flaggers' report potentially unlawful content or content deemed unacceptable under companies' Terms of Service. Trusted flaggers' can either be government agencies (e.g. law enforcement flagging 'extremist' content, that may not necessarily be unlawful content under domestic law) or associations (e.g. representing the interests of intellectual property rights holders or charities fighting against "hate speech" or other forms of discrimination, organisations protecting the rights of children etc.). It is unclear how companies award 'trusted flagger' status to various organisations. We note that the European Commission is looking at strengthening this system by strongly encouraging Internet platforms to adopt certain criteria or revoke the 'trusted flagger' status of organisations who fail to provide detailed takedown notices;[42]

- Individual reporting: individuals can also report potentially unlawful content or content deemed unacceptable under companies' Terms of Service. However, processing of content reported on this basis is likely to be slower;

- **Algorithms and 'hash' tags**: over the years, companies have become more upfront about their use of algorithms to identify undesirable content. In particular, they are

under intense pressure to prevent certain types of content from being even uploaded online in the first place. This primarily concerns 'extremist' or terrorist content and, following the *Delfi* judgement by the European Court of Human Rights and various initiatives from the European Commission on "hate speech".[43] In particular, some Internet companies have recently developed a database of "hashes" (i.e. images) deemed terrorist or 'extremist'. This database is shared with participating companies who can then make their own decision as to whether the particular image should be taken down.[44] Equally, Google proactively de-list 'revenge pornography' as a matter of company policy;[45] Facebook is using hashes for "revenge pornography."[46]

Beyond their Terms of Service and community guidelines, however, there is very little information about the way in which companies assess content for takedown. For instance, it is only through leaks to the Guardian that users have recently been given a window in Facebook's internal policies for the removal of content.[47] More generally, companies use a graduated approach to sanctions for unlawful removal.[48]

V. Appeals and remedies

ARTICLE 19 finds that Internet companies do not generally provide a clear complaints mechanisms for wrongful removal. Complaints seem to be handled on an entirely discretionary basis. To begin with, it appears that users are not notified systematically that their content has been flagged or removed as a matter of policy. When users are notified that their content has been removed, it is not clear whether they are provided with information about ways to challenge companies' decisions to remove that content. Equally, it does not appear that companies ever give reasons for the decision to remove, presumably because that would not be efficient or allow them to deal with the millions of requests they deal with on a regular basis. More generally, it appears that the most effective way to challenge companies' removal of content is by going public either on social media like Twitter or the press.

ARTICLE 19 believes that the Special Rapporteur in his forthcoming report should reiterate that Internet companies should develop clear redress mechanisms for individuals whose content has been taken down in circumstances where that content is legitimate under international human rights law. Such mechanisms must meet a due process threshold as defined by international human rights law. At the very least, companies should notify users that their content has been removed and give them the basic reasons for their decision. They should also provide users with an opportunity to challenge those decisions, particularly when the content at issue is lawful under national or international human rights law. These basic due process safeguards should be put in place in practice and made clear in companies' policies. At a bare minimum, companies should provide an email address to enable individuals to complain about wrongful removals of their content.

VI. Automation and content moderation

Algorithmic and automated decision-making is increasingly the method of choice for regulating online content. Algorithms can refer to any computer code that carries out some set of instructions, and is essential to the way in which computers process data.[49] Automated decision-making involves the collection of large data sets, which are processed by algorithms, and the automatic execution of decisions based on the application of the algorithm to the data set.[50] While these new mathematical models hold great promise, they also come with risks, especially for the right to freedom of expression.[51]

Often, online content regulation is automated or happens through algorithmic filters with limited human oversight. One of the most widely accepted uses of algorithmic decision-making is in the removal and filtering of child sex abuse images/videos by online platforms and Internet Service Providers (ISP). This filtering system is now automated in both the US and the UK, based on the use of "hash" technology to identify content for removal and filtering; it is also increasingly used for removals of other type of content.[52] This practice has been criticised due to the lack of judicial oversight mechanisms, lack of transparency and the risks of over-blocking.[53]

Algorithmic decision-making is also widely used in the context of copyright removals. In these cases, there is usually human input in the process, as copyright owners are asked to upload their material within a specific program and to decide what consequence a breach of copyright should have. These programs have also been widely criticised; in particular in the context of YouTube, there is a strong perception that its appeal system places copyright owners in the position of "playing judge, jury and executioner".[54]

Algorithmic decision-making has also been implemented in the context of abusive messages on social media. On Twitter, for example, "abusive messages" are filtered out of the recipient's notifications, while still remaining visible on the platform. Staff members then make a decision regarding possible bans and suspensions for the abusive user. It is unclear however the extent to which 'algorithms' are able to take into account context, cultural or linguistic differences in determining what amounts to 'abusive' language.

More generally, algorithms play a part in the distribution of online content by prioritising news from particular sources. Algorithms have also been deployed to flag "fake news". In short, algorithms play a significant role in determining the type of information that users consume online.[55] Yet there is almost no transparency concerning the way in which algorithmic decision-making works in this area, raising concerns for media pluralism.

Overall, over-blocking, vague community guidelines, bias and lack of transparency in relation to the use of algorithms, are recurrent problems in the context of algorithmic or automated decision-making. Of particular concern is companies' reliance on automated machine learning on the basis of undefined terms such as 'extremism' or "fake news". This is compounded by the lack of clear complaints mechanisms to deal with wrongful removals. In practice, online intermediaries unilaterally decide what redress mechanisms, if any, are made available under their Terms of Service, and the level of human oversight. Users' ability to challenge these decisions before the domestic courts is also extremely limited.[56]

However, the impact of algorithms and automation goes beyond the top layers of the Internet and extends into its architecture. Often overlooked, but of equal importance to enabling freedom of expression online, is the increased use of algorithms for network management of critical infrastructure, from the electrical grid[57] to Internet routing.[58] There is limited knowledge about the current topology of the network, how data is routed across it, and who is connected to whom, and how. Adding further automation to the process increases the risk of making the Internet's operation into a black(er) box. This can have serious consequences for activist and academics interested in understanding the Internet's technical workings to ensure its resilient, distributed nature can continue to be leveraged towards enabling human rights.

All these issues taken together have led to a renewed focus on, and increasing concern about, the impact of algorithms and automated decision-making on the right to freedom of expression.

ARTICLE 19 finds that further research is needed about the use, effects, and impact of algorithms on the architecture of the Internet. We need to understand the broader implications of the various forms of algorithmic decision-making for freedom of expression and access to information beyond their use in particular instances. As applications of automated decision making become more sophisticated and widespread, understanding the potential, limitations and dangers of the technology *per se* becomes important.

ARTICLE 19 suggests that the Special Rapporteur addresses these issues in this report. At minimum, he should recommend that:

- Automated decision-making shall include a sufficient level of human oversight. The level of human intervention and nature of human input must be made more easily understandable, and accountable;

- Individuals should have a right not to be subjected to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. In practice, this means that they should have a right to challenge such decisions

VII. Transparency

As ARTICLE 19 highlighted above, users are not notified about content restrictions as a matter of principle but on a discretionary basis. Equally, it is unclear whether individuals are notified of the reasons for content removal in specific instances. It appears that in the vast majority of cases, this is not the case. Similarly, the major Internet companies do not provide clear complaints mechanisms to challenge wrongful removals. Whereas Internet companies generally provide detailed procedures or mechanisms to report violations of their Terms of Service, no such equivalent exists for content removals.

ARTICLE 19 notes that the main Internet companies that publish transparency reports do not generally provide information about content removed on the basis of their Terms of Service. However, some progress has been made with Twitter reporting content which has been removed on the basis of their Terms of Service, but at the request of governments rather than users or other 'trusted flaggers'. However, no such information is provided in relation to content proactively removed by companies or to content removed on the basis of companies' Terms of Service at the request of private parties, including 'trusted flaggers'.

In short, ARTICLE 19 considers that current transparency reporting is insufficient. Companies should provide information about content removed on the basis of their Terms of Service, whether at the request of governments, private parties, including 'trusted flaggers', or on a proactive basis by the company itself, in disaggregated format. Equally, they should publish their internal guidelines for the removal of content.

VIII. Examples

ARTICLE 19 submits that it is currently extremely difficult to find out about wrongful removal of content on the basis of Terms of Service. Companies do not publish data about content removal on that basis in their Transparency Report. One has to rely on media reporting or initiatives such as Lumen[59] or Onlinecensorship.org.[60] One recent example of wrongful removal of legitimate content includes the takedown of content posted by Rohingya activists on Facebook.[61]

[1] See ARTICLE19's analysis of the German NetzDG law on the regulation of social networks, September 2017.
[2] See e.g. Samuel Gibbs, EU warns tech firms: remove extremist content faster or be regulated, The Guardian, 7 December 2011.
[3] Forthcoming publication, available upon request from ARTICLE 19 in 2018.
[4] See e.g. Google Transparency report, Governments' requests to remove content.
[5] See e.g Twitter: https://help.twitter.com/en/rules-and-policies/twitter-law-enforcement-support
[6] E.g. this seems to the case under the newly adopted Germany law on social networks Germany. See, e.g. David Meyer, Germany's new hate speech law goes live: So who's in its sights?, The German View, 2 October 2017.
[7] Usually because they believe that a court would decide that the content is lawful.
[8] See e.g. UK Terrorism Act 2006.
[9] However, the EU's expansive interpretation of personal data suggests that social media platforms such as Facebook may well be required to comply with similar requests in future. See EUGDPR, Questions about the Incoming GDPR.
[10] See e.g. Google, Request removal of content indexed on Google Search based on data protection law in Europe.
[11] Guidelines on the implementation of the Court of Justice of the European Union judgment on Google Spain and Inc v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, C-131/12.
[12] See Article 17 GPDR, Position of the Council at first reading with a view to the adoption of a Regulation of the European Parliament and the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 6 April 2016.
[13] ARTICLE 19, Policy Brief: The Right to be Forgotten, 29 March 2016.
[14] ARTICLE 19, The Global Principles on Freedom of Expression and Privacy, 2017.
[15] See e.g. Twitter Public Policy, Global Internet Forum to Counter Terrorism, 26 June 2017; or Google, Update on the Global Internet Forum to Counter Terrorism, 4 December 2017.
[16] See Twitter Public Policy, *op.cit*.
[17] Google, How we fought bad ads, sites and scammers in 2016, 25 January 2017.
[18] Amber Jamienson & Olivia Solon, Facebook to begin flagging fake news in response to mounting criticism, 15 December 2016.
[19] Sam Levin, Facebook promised to tackle fake news. But the evidence shows it's not working, 16 May 2017.
[20] Facebook Newsroom, Update on German Elections, 27 September 2017.
[21] Facebook Newsroom, Improving Enforcement and Transparency of Ads on Facebook, 2 October 2017.
[22] Facebook Newsroom, News Feed FYI: New Test to Provide Context About Articles, 5 October 2017.
[23] Twitter Public Policy, Update: Russian Interference in 2016 US Election, Bots, & Misinformation, 28 September 2017.
[24] BBC, YouTube to offer fake news workshops to teenagers, 21 April 2017.
[25] Market Watch, YouTube cracks down on conspiracies, fake news, 5 October 2017.
[26] See ARTICLE 19, Freedom of Expression Unfiltered: How blocking and filtering affect free speech, 8 December 2016.
[27] See ARTICLE 19, France: ARTICLE 19 supports claim challenging lawfulness of administrative website blocking, 30 July 2015.
[28] See, ARTICLE 19, Canada: Canadian Supreme Court must uphold freedom of expression standards in injunction case, 7 October 2016.
[29] See, ARTICLE 19, Civil society tells EU Court: Right to be forgotten should not be global, 30 November 2017.
[30] The submissions are available from here: https://www.article19.org/resources/canada-canadian-supreme-court-must-uphold-freedom-of-expression-standards-in-injunction-case/
[31] There can be a perception that internet companies default to censorship when they don't fully grasp the context in which they operate, see e.g. Meeran Karim, As American Tech Firms Move to India, Many Choose to Self-Censor, The Slate, 18 July 2017; or Kristen V. Brown, What's Behind Facebook's Censoring Of Atheists In India, The Huffington Post, 16 November 2015.
[32] See e.g. Clair Cain Miller, When Algorithms Discriminate, The New York Times, 9 July 2015.
[33] See, e.g. Albert Fox Cahn, Why is Facebook Censoring Rohingya Accounts of the Genocide, 2 October 2017.
[34] Tom McKay, Search for 'Bisexual' on Twitter Right Now, and No News, Photos, or Videos Show Up Gizmodo, 5 November 2017.
[35] See e.g. Christopher Harity, Why Does Facebook Censor Gay Images?, Advocate, 30 January 2015.

[36] See, e.g. Amar Toor, Facebook still has a nipple problem, The Verge, 12 October 2016.
[37] See, e.g. Mark Scott& Mark Isaac, Facebook Faces a New World as Officials Rein In a Wild Web, The New York Time, 17 September 2017.
[38] See for instance, the ambiguity over what agreement, if any, facebook reached with the Pakistani government over blasphemy- removal that affects religious minorities, including freethinkers and dissenters; see e.g. Asif Shahzad & Saad Sayeed, Facebook meets Pakistan government after blasphemy death sentence, Reuters, 7 July 2017.
[39] See, e.g. Dawn Enis, Facebook Responds to Criticism of 'Real Name' Policy, Advocate, 2 November 2015,
[40] See, e. EFF, Changes to Facebook's "Real Names" Policy Still Don't Fix the Problem, 18 December 2015.
[41] For a summary, see e.g. the European Commission's paper on Tackling Illegal Content Online, October 2017.
[42] See, ARTICLE 19, EU fails to protect free speech online, again; 5 October 2017.
[43] *Op.cit.* See also FN 23 above and ARTICLE 19, EU: European Commission's Code of Conduct for Countering Illegal Hate Speech Online and the Framework Decision, 30 August 2016. This presumably also applies to child sex abuse images and pornography.
[44] See Twitter Public Policy, Update on the Global Internet Forum to Counter Terrorism, 4 December 2017.
[45] See e.g. Woodrow Hartzog, Google's action on revenge porn opens the door on right to be forgotten in US, 25 June 2015.
[46] See e.g. Did Facebook finally figure out that consent is more important than nipples?, Take Back the Tech.
[47] See The Guardian, Facebook Files.
[48] See e.g. Twitter Rules.
[49] Centre for Internet and Human Rights, The Ethics of Algorithms: from radical content to self-driving cars – final draft background paper, GCCS 2015.
[50] M. Perel & N. Elkin-Koren, Accountability in algorithmic copyright enforcement, Stanford Technology Review, forthcoming.
[51] See e.g. CDT, Mixed Messages: the Limited of Automated Social Media Content Analysis, November 2017; or Daphne Keller, The European Commission, for one, welcomes our new robot overlords, Stanford Law Review, 9 October 2017.
[52] EDRi, Algorithms - censorship a la carte, 12 July 2016.
[53] ARTICLE 19, Algorithms and automated decision-making in the context of crime prevention, 2 December 2016.
[54] C. Hassan, What about all that copyright takedown abuse, YouTube?, Digital Music News, 29th February 2016.
[55]See ARTICLE 19, Submission of Evidence to the House of Lords Select Committee on Artificial Intelligence, 6 September 2017.
[56] In particular, the legal basis for any court challenge is contract law, where the standard is generally the lack of fairness of contractual terms, i.e. a very high threshold for consumers. Moreover, social media platforms Terms of Use usually contain jurisdiction clauses forcing users to use the courts in California rather than the courts of their place of residence.
[57] See, e.g. How Artificial Intelligence is shaping the future of energy, Open Energi, 9 February 2017.
[58] For example, see Hao Bai, A Survey of AI for Network Routing Problems.
[59] https://lumendatabase.org/
[60] See Online Censorship.
[61] See, Betsy Woodruff, Facebook Silences Rohingya Reports of Ethnic Cleansing, Daily Beast, 18 September 2017. See also, EFF, Adult Content Policies: A Textbook Case of Private Censorship, 7 December 2017.