Primer: Respecting human rights in content regulation in the digital age

January 2018

*Prepared by Access Now for David Kaye, Special Rapporteur on the promotion and protection for the right to freedom of opinion and expression, responding to questions for the “*Study on Content Regulation in the Digital Age”

**Table of Contents**

Introduction 1

State content regulation laws and measures, including those concerning terrorism-related and extremist content 3

False news and disinformation 5

Right to be forgotten 6

Content-related requests from States other than those based on law or the company’s terms of service 7

Global removals 8

Individuals and communities at risk 10

“Doubleswitch” attacks on human rights activists 12

Appeals and remedies 13

Automation and content management 14

Notification, disclosure, and transparency 15

Recommendations 17

Conclusion 20

## Introduction

Access Now is an international organization dedicated to defending and extending the digital rights of users at risk.[[1]](#footnote-1) We provide thought leadership and policy recommendations for actors in the public and private sectors across the globe. Our work is aimed at ensuring that the internet remains an open platform for the exercise of human rights, and we lead inclusive advocacy efforts to engage all stakeholders in developing rights-respecting law and norms. Access Now also operates a 24/7 Digital Security Helpline that provides real-time direct technical assistance to vulnerable communities and individuals around the world.

We appreciate this opportunity to contribute to the Special Rapporteur’s report to the 38th session of the Human Rights Council. The internet is a powerful tool for promoting the right to freedom of expression -- an enabler of other rights -- and impacts our participation in civil and political life. Through open and secure access to the global internet, users confidently form and openly express their opinions and fully realize their right of access to information. Therefore, use of the internet requires the same protections as offline communications.

Existing legal and normative frameworks affirm and delineate the application of human rights online. Article 19 of the International Covenant on Civil and Political Rights (ICCPR) remains visionary for its projection of the rights to opinion and expression across borders and media. Freedom of expression restrictions must meet the “necessary and proportionate” test outlined in Art. 19(3) of the ICCPR. As the UN Human Rights Committee explains in General Comment 34, interpreting Article 19:

“Any restriction on the operation of websites, blogs, or any other internet-based electronic or other such information dissemination system, including systems to support such communication, such as internet service providers or search engines, are only permissible to the extent that they are compatible with paragraph 3.”

However, the foundational international human rights instruments did not factor in the rise of multinational corporations as powerful as the largest internet intermediaries and telecommunications companies. These instruments sufficiently and powerfully apply in this digital age, yet the role of internet intermediaries under them requires further exploration, as this project by the Special Rapporteur and other instruments have sought to provide through guidance and norms for the private sector. For example, the Organization of Economic Cooperation and Development Guidelines for Multinational Enterprises are voluntary, responsible business principles applicable to multinational enterprises. Pursuant to these principles, enterprises may be accountable for adverse human rights impacts, including the violation of privacy rights, linked to their operations and products. The 2011 Guidelines also recommend that enterprises carry out due diligence in order to protect against potential adverse impact of products and services, and in particular identify internet freedom for protection. The UN Guiding Principles on Business and Human Rights achieved consensus support at the Human Rights Council in 2011. Designed for broad applicability across business sectors, the Guiding Principles clarify the responsibility of businesses to respect human rights and participate in remedial action for harms connected to their operations.

Despite these clear tests and internationally applicable codes of conduct, global internet freedom is declining due to increasing government censorship and pressure on intermediaries and companies to take voluntary action for privatized enforcement of vaguely defined restrictions regarding types of information, identities, or modes of communication. In 2016, authorities in 38 countries arrested internet users based solely on social media content.[[2]](#footnote-2)

Government are now shifting their burdens to companies and pressing them to remove content — including deleting specific webpages, accounts, and social media posts. According to a Freedom House report, 65% of the countries surveyed required “companies, site administrators, and users to restrict online content of a political, social, or religious nature.”[[3]](#footnote-3)

Users need assistance in asserting their human rights online. In the past two years, between February 2016 and early January 2018, our Digital Security Helpline has handled approximately 204 cases, including 102 from Syria, relating to online content that was flagged, removed, or blocked by platforms. Our cases represent a small fraction of instances where users require help with expertise, contacts, channels, confidentiality, or simply trust to approach the companies that make decisions impacting their rights. That half of our cases arose in Syria shows the acute need in conflict zones for assistance mitigating content and account regulation, and the grave importance of decisions over content originating in dangerous places. This is particularly alarming because activists often rely on internet communication and platforms to document and expose human rights violations. When sites of atrocities face censorship, we should consider the impacts on access to information as well as peace, dignity, and justice.

Frustrated with intermediaries for their perceived reluctance or slow pace in combating extremism online, governments and intelligence agencies have created Internet Referral Units (IRUs). IRUs are designed to prevent and combat serious forms of crimes committed through the internet. They seek to do this by referring illegal internet content to online service providers for their voluntary consideration. This approach is outside the rule of law and can violate international human rights standards, an issue that we will discuss in this report.

This submission contains information Access Now and our partner organizations have compiled regarding the human rights implications of content management practices. We exhort the Special Rapporteur and other stakeholders to consider this information as they continue to investigate and advocate for the application of the right to freedom of expression online.

## State content regulation laws and measures, including those concerning terrorism-related and extremist content

Pressure is mounting on the internet intermediaries to “do something” about terrorism and extremism. Unfortunately, both public and private actors are proceeding without clear definitions of the problem, without objective and open evidence on what may work and any unintended consequences, and without transparency or accountability for their actions.

Governments coerce companies to assume the role of police, judge, and jury of online content. They compel companies to implement content management programs either by advancing new legislation or by threatening to censor speech outside of the legal process. For instance, Germany recently enacted the “Enforcement on Social Networks” law, commonly known as “NetzDG.” This law requires social media platforms to remove hate speech within 20 hours to seven days, depending on the difficulty of the content evaluation. A company’s failure to remove hate speech may result in a 50 million euro fine.[[4]](#footnote-4) Such extreme penalties weigh the scales toward blocking. This dynamic results in the removal of lawful material and the discriminatory profiling of account holders, especially those marginalized and vulnerable users that do not have the means to appeal corporate decisions or social capital and security to raise their voices openly.

The risks of private enforcement measures for freedom of expression are exacerbated when they take place within the opaque, closed, top-down structure typical of many technology firms. One illustration of the potential for harm is the unilateral decision by Cloudflare CEO Matthew Prince to terminate services for the Daily Stormer, a neo-Nazi and white supremacist website. In a company email, Prince admitted that the decision was arbitrary and he made it “because they were assholes.”[[5]](#footnote-5) We agree with the concerns voiced by Special Rapporteur Kaye on the decision; as he stated, character traits are a poor reason for restricting freedom of expression.[[6]](#footnote-6) This incident is particularly notable given that, traditionally, web hosting and internet infrastructure companies are not expected to make decisions about which content to host or protect. The arbitrary nature of the decision shows that no community standards had been developed.

Given these risks, companies should implement procedural safeguards to ensure their content management programs do not impose disproportionate and unnecessary restrictions to freedom of expression. They must also incorporate human rights law and norms as a core consideration when they assess whether to restrict online content or drop hate groups from services. Moreover, if companies do undertake private enforcement, the companies and those who review the content cannot be primarily responsible for evaluating whether it is legal in the absence of rule-of-law mechanisms. The staff should be well-trained in evaluating context and other factors, and the process should be safeguarded by ensuring that firms have brought in outside experts, and formed advisory councils, to aid staff in content review, in addition to other potential measures. Staff that review content see a barrage of difficult and offensive material, and they require sufficient psycho-social, financial, and other support. Furthermore, companies should ensure that corporate structures are in place to adequately oversee enforcement mechanisms, with Board-level expertise and robust review and audit procedures. In addition to implementing rights-respecting regulations, it is essential to employ expert third-party assessment and regular civil society consultation, of the kind that the Global Network Initiative enables, to ensure that corporations respect human rights in the digital age. Companies should report any government flagging of the content on their platforms through corporate transparency reports, per best practices and indicators like those identified by the Ranking Digital Rights Corporate Accountability Index. At a broader level, platforms must ensure that governments do not use flagging to circumvent legal protections for the freedom of expression.

### False news and disinformation

Among the most discussed global trends in 2017 was the use and effectiveness of disinformation campaigns via websites and social media.

So-called troll armies are being used to execute concerted disinformation campaigns. “Troops” are comprised of groups of individuals as large as 300,000 to two million people. In some cases, these troops are comprised of government employees. These individuals assume a false identity to participate in internet platforms, such as Twitter or Facebook. A member of a troll army might create several online identities in one forum and fake an argument between two, and then post false information in order to smear the reputation of a target opposition leader.[[7]](#footnote-7)

The use of “troll armies” to shape online discourse has grave implications for the freedom of expression because they can be used to discredit or silence dissenting voices, demonize opposition, and disseminate propaganda.[[8]](#footnote-8) Troll armies are particularly disruptive to access to information, since many internet users rely on the information they obtain through social media platforms. According to a study by Stanford Journal of Economic Perspectives, during the 2016 election, 62% of adults in the U.S. relied on and believed information from social media news sources.[[9]](#footnote-9)

However, censorship is not the solution for troll armies, given that this tactic could easily be abused to silence activists and damage the internet’s capacity for disseminating information at an unprecedented speed and scale. Governments should commit not to use troll armies, either domestically or abroad. Platforms should take measures to address this matter, including transparent enforcement of terms of service and rights-respecting policies for regulation of content and accounts.

### Right to be forgotten

European data protection laws, such as the General Data Protection Regulation (GDPR), empower individuals to control their personal data, including the retention and listing of certain data. The “right to be forgotten” actually comprises two different aspects: first, what is commonly referred to “as the right to erasure,” and second, the “right to de-list.” The right to erasure gives individuals the right to request the deletion of all personal data pertaining to them when they leave a service or application. The right to de-list, elaborated in the *Google Spain* opinion of the Court of Justice of the European Union, provides individuals with the right to request that search engines de-list a search result associated with their names that “appears to be inadequate, irrelevant or no longer relevant or excessive… in light of the time that had elapsed.”[[10]](#footnote-10) The court’s broad interpretation of this right raises the concern that the right to de-list will be used to curtail freedom of expression and access information. As countries look to the European Union’s model of data protection, and expand the right to de-list, we have already seen damage to the free flow of information online.[[11]](#footnote-11)

South Korea was among the first countries to debate the “right to be forgotten” after the *Google Spain* ruling. The right to de-list first rose to public consciousness there when MC Mong — a South Korean rapper — sought exemption from mandatory military service. The government requires that all South Korean men serve in the military for two years, while exempting from service those with poor health. After finding remarks that Mong made in an online Q&A platform in 2005 questioning whether he would be exempt from service if he lost a number of teeth, some people began to suspect he had purposely extracted his teeth to dodge the draft. The Q&A platform does not allow people to remove questions they ask once they are answered.

Following intense public debate and consultation regarding the right to de-list this kind of information, the Korea Communications Commission (KCC), a government agency, developed guidelines aimed at protecting that right. These guidelines diverge significantly from the EU right to de-list. Firstly, the guidelines are principles not established by law. Secondly, they are primarily focused on users’ online posts rather than articles posted by a third party, since Korean law already grants people the right to request the deletion of information by a third party if it is deemed damaging to one’s reputation. The guidelines also expand the right to de-list beyond search engines, ordering internet companies to accept removal requests in some “exceptional cases” where a user’s control of content was formerly limited or blocked. Last but not least, reports on the guidelines suggest that if a user requests de-listing, the content in question will be deleted. This approach is at odds with the right to de-list developed in the EU, raising serious concerns for free expression and creating risks for censorship.

While Access Now supports the right to erasure, we cannot support development of the right to de-list or a right to obscurity by governments around the world. We recommend that the right to de-list be exercised within the data protection context and limited to circumstances where the sole objective is protection of the personal data of non-public figures. The implementation of the right to de-list should fall solely on the data controller, the entity responsible for determining when, how, and for what purpose personal data should be processed. Lastly, the right to de-list should not lead to the deletion of online content; although the web address may be delisted, the content must remain online.[[12]](#footnote-12)

## Content-related requests from States other than those based on law or the company’s terms of service

In addition to enacting legislation to compel companies to implement content management, several governments are pursuing mass take-down of content through the aforementioned Internet Referral Units (IRUs). This practice has a widespread chilling effect on freedom of expression. As noted by Special Rapporteur Kaye, “states should not require or otherwise pressure the private sector to take steps that unnecessarily or disproportionately interfere with freedom of expression, whether through laws, policies, or extralegal means.”[[13]](#footnote-13)

These private enforcement programs are at odds with freedom of expression, transparency, and the rule of law. They pressure companies to police content, processes, and users’ activity outside of legal process. One illustration is the Europol IRU program, which supports EU states in preventing and countering serious crimes committed through the internet. This unit flags what may be illegal content to online service providers. The providers may then voluntarily examine the flagged content and evaluate whether it meets their terms and conditions, and then if they see fit, subsequently remove it.[[14]](#footnote-14) According to a July 2016 Europol press release, to date, there have been more than 11,000 referrals to approximately 31 online platforms via this program, and 91.4% of the flagged content has been removed.

The Europol IRU, similar to other government flagging programs, circumvents the rule of law and raises serious human rights issues because it lacks procedural safeguards. If law enforcement agencies are to challenge the legality of an activity, whether it occurs online or off, they should be required to pursue the matter using legal channels, processes, and procedures. These units instead delegate to intermediaries the responsibility for dealing with alleged illegal content without providing individuals with the due process protections of a judicial or quasi-judicial proceeding.

As such, the Europol IRU represents a clear circumvention of the rule of law, and like other programs of its nature, stands in violation of international human rights standards. Illegal content is a serious matter, but removal of such content should take place only when the content has been specifically adjudicated as being illegal and a court order has been issued.

## Global removals

Governments and private parties are demanding that internet intermediaries take down or restrict access to content not only in their own jurisdictions, but in others as well.

In December of 2013, the U.S. Department of Homeland Security (DHS) asked GoDaddy take down the 1dmx.org website. This site hosted numerous posts that were critical of the Mexican government. On December 2, GoDaddy took the site down, days after it had been used to memorialize the one-year anniversary of the incidents of police brutality that took place after Enrique Pena Nieto’s inauguration. GoDaddy, located in the U.S., said it took down 1dmx.org at the request of a DHS official stationed in Mexico City.[[15]](#footnote-15) Although the GoDaddy restored the website four months later, neither the company nor the governments involved have since provided an explanation for the takedown.

Unfortunately, there are other easy routes for forcing intermediaries to take down content. The U.S. Digital Millennium Copyright Act (“DMCA”) is also applied internationally to chill freedom of expression rights. This law was originally intended to prevent online copyright infringement. Under the DMCA, internet intermediaries are shielded from liability for hosting materials that infringe copyright laws so long as they remove the content after they receive a complaint. Many intermediaries, including Facebook, YouTube, and Twitter, are incorporated in the U.S., a party to various international intellectual property treaties, and that means that they are required to recognize copyright laws in other jurisdictions. These platforms typically remove content promptly following receipt of a DMCA complaint, and the scope of such removal is vast. For instance, Google received 75 million requests for removal of content in only the month of February 2016.[[16]](#footnote-16) Users can appeal content takedown and make their case under “fair use” exceptions to copyright, but it can take weeks before the content it restored. In the U.S., material under copyright may be shared if it is used for purposes such as criticism, commentary, or parody, but there are no bright lines, and context is critical for a fair use determination.

These factors make the DMCA ripe for abuse to censor free expression. For instance, Ares Rights, a Spanish law firm, has relied on the DMCA as a tool to silence political opposition. The firm has filed copyright complaints alleging that content on social media platforms violate intellectual property rights, including “photos and videos of Correa and his government.” As a result of these claims, Facebook, YouTube, and Twitter have removed content published by Ecuadorian activists and organizations. The most controversial removal case involved the takedown of an image that included an official government logo, used to mock the Ecuadorian president’s weekly Saturday address.

Lastly, the unrestricted application of the right to de-list in several jurisdictions is undermining the right to access to information. One case that illustrates the problem is **Google v. Equustek**, which involves a dispute between Equustek Solutions Inc. (“ESI”) and Datalink Technologies.[[17]](#footnote-17) ESI, a Canadian company, manufactures and sells network interface software. ESI alleged that Datalink engaged in trademark infringement. A Canadian court ruled the Datalink breached trademark law, but the company continued to sell the infringing products online. Subsequently, the Canadian court granted an injunction requiring that Google remove from its index all links to Datalink sites. Google removed 345 Datalink websites from its Canadian domain (Google.ca), but was not willing to de-index the site globally. It is important to highlight that this application of the right to de-list goes far beyond what was introduced under EU law, since it implicates de-indexing (not just de-listing what is in the index), which is not foreseen under the GDPR for data protection. This decision could have significant ramifications for human rights, given that authoritarian regimes could demand that courts apply their broad censorship laws to what search engines can index and make discoverable globally.[[18]](#footnote-18)

To avoid takedowns or other content restriction practices that undermine the freedom of expression globally, companies should ensure that their takedown and appeal policies are clear and public. They should interpret governments’ jurisdictional authority appropriately narrowly to minimize the adverse impact of takedown orders on the right to freedom of expression. Lastly, companies that get takedown requests should avoid defaulting to a process whereby they remove the disputed content first and then notify the user of the redress procedure.

## Individuals and communities at risk

Company content standards often fail to protect vulnerable communities from online harassment. Since 2013, Access Now’s Digital Security Helpline has handled 166 cases for groups or individuals that defend women’s rights. Around half of these clients requested security consultations and sought advice on assessing their practices and protecting the privacy of their communications, while the other half asked for rapid response assistance. For the past two years, the most common requests to our Helpline from these types of groups have been for assistance with compromised or potentially compromised accounts; harassment incidents; and secure communications and protection of websites.

Our organization has developed resources to help promote an inclusive and digitally secure community, including *A First Look a Digital Security*,[[19]](#footnote-19) a booklet that guides people in taking the first steps to improving their security and safety online. Access Now also advocates against corporate practices that put women and other vulnerable and marginalized communities at risk of human rights violations.

An Access Now grantee, Japleen Pasricha, the founder and director of Feminism in India.com, wrote that cyber violence against women is “not restricted to: cyber stalking, cyber bullying, cyber harassment, identity theft, breach and violation of privacy/confidentiality, voyeurism, and image-based sexual violence, popularly known as ‘revenge porn.’”[[20]](#footnote-20) In a post, Pasricha shared data from a Feminism in India report, “*Violence” Online in India: Cybercrime Against Women & Minorities on Social Media,* showing “that thirty-six percent of the respondents who had experienced harassment online took no action at all. Twenty-eight percent reported that they had intentionally reduced their online presence after suffering online abuse.” Pasricha believes that in order to address cyber harassment, women, social actors, and law enforcement agencies must be educated about the importance of prosecuting individuals that use social media as a forum to perpetrate violence against women and minorities.

Company policies like Facebook’s “authentic name” requirement pose just this sort of risk. Facebook requires that users provide their real name and information to use the service, and when people use pseudonyms and are flagged, the company has unilaterally altered the name on accounts to the legal name. A Honduran blogger named “La Gringa,” who writes about Honduran politics, had her Facebook account suspended after she used a pseudonym.[[21]](#footnote-21) She stated “ [t]his week I started writing a series of blog articles about crime and narco trafficking in Honduras — and it is likely what prompted the complaint about my account, just as posting of my political articles were blocked by Facebook for a time last year and the year before because of a false complaint […] By asking for a copy of my ID, Facebook is asking me to put my life in danger. By disabling my account, Facebook is silencing one of the few internet voices in English in Honduras.”

Vietnamese writers and activists who used a pen name were also flagged and barred from using the service. When Facebook asked for documents to verify their identities, and the activists supplied them, the company then changed the name on their accounts to their legal names. One of those forced to use her real name was a mother using the platform to fight for the release of her two detained sons.[[22]](#footnote-22)

To show the links between identity policies and human rights, Access Now joined the “Nameless Coalition,” which includes groups defending privacy and other human rights, women’s rights, and the interests of members of marginalized and oppressed groups such as LGBTQI people, indigenous people, and people in oppressed religious, ethnic, and other communities. Together, the coalition denounced Facebook’s policy, asserting that it is culturally biased and technically flawed. Access Now recommends that online platforms commit to allowing the use of pseudonyms in appropriate circumstances. If they persist with real name requirements, they should require that complaints against people based on a real name policy are backed by evidence so these complaints are not abused to victimize people.[[23]](#footnote-23) Additionally, to protect digital security, platforms should let people confirm their identities without having to submit government-issued identification, which could be leaked, stolen, and used for identity theft. Intermediaries should also notify users about what happens when they submit their identity information, including providing information about where and how their data will be stored. Users should have the capacity to protect their identification by using PGP or encrypted communication. Lastly, these companies should institute a robust appeals mechanism to ensure that when people are locked out of their accounts, they have the ability to request a second review, to submit evidence, and to discuss their investigation with staff, particularly in cases where personal safety is a concern.

In June 2017, after numerous complaints, Facebook revised the implementation of its identity policy.[[24]](#footnote-24) The company now allows individuals who do not want to use their real names to explain their circumstances. However, the Nameless Coalition has determined that the amendments to the policy are not sufficient to protect vulnerable communities, who are still at risk of arbitrary enforcement. The coalition finds that the revised policy fails to provide adequate personal data safeguards, and should allow for exceptions that keep people out of danger.

Other online platforms and applications are making progress in protecting vulnerable communities. Grindr, Hornet, and Scuff, developed for LGBTQI communities, have successfully implemented strategies to help keep at-risk users safe from online harassment, targeting, and human rights violations. Grindr allows people in the Middle East, the Gulf, and North Africa to prevent harassment and protect the app’s content by changing the Grindr thumbnail icon on their device and setting a special passcode. Grindr and Hornet have also added additional safety tips in Arabic for people in Egypt.

### “Doubleswitch” attacks on human rights activists

Activists around the world rely on social media as a tool to communicate and advocate for human rights. These activists have been targeted with an attack called the “Doubleswitch.” The perpetrators of this type of attack gain access to a victim’s account; change the victim’s handle as well as the email associated with the account; and finally, create a new account with the victim’s original handle, but linked to the hacker’s email address. When the second “switch” occurs, the hacker achieves full control, and in any attempt by the original account holder to regain control of their account and identity, recovery messages will be emailed to the hacker.

Access Now’s Digital Security Helpline helped to identify the attack and reported an increase in its use against activists, journalists, and politicians across the globe. These attacks cripple the work of advocates because hackers may utilize the very reach and legitimacy of the target to undermine their campaigns, damage their reputations, and misinform their followers.

Internet platforms should protect their users from “Doubleswitch” attacks by enabling multi-factor authentication and encouraging users to activate it, thereby preventing hackers from gaining control of accounts in the first instance. The method should use app-based authenticators, given that text-based phone authentication exposes the identity of the users and places them at a greater risk.[[25]](#footnote-25) Social media platforms should also consult with “Doubleswitch” victims and update their rules in accordance with victims’ input.

## Appeals and remedies

Companies have largely failed to meet their responsibility to provide an effective remedy for violations of free expression they cause or contribute to. The United Nations “Ruggie” Framework for Business and Human Rights affirms that businesses have a responsibility to avoid infringing human rights and should take appropriate steps to address human rights impacts by preventing, mitigating, and remediating.[[26]](#footnote-26) Pursuant to the UN Guiding Principles on Business and Human Rights (Guiding Principles), when companies identify that they have caused or contributed to an adverse impact, they should afford the affected parties remediation.[[27]](#footnote-27) States also play a vital and primary role in ensuring adequate judicial and non-judicial processes are in place for adversely impacted parties to seek legal redress.

Remedy takes both procedural and substantive forms. Access Now has identified a host of measures and policies, from staff training to preservation of evidence and appeals mechanisms, that internet and telecom companies should employ to better provide access to remedy for affected users.[[28]](#footnote-28) Companies should implement grievance and remedy mechanisms to address users’ freedom of expression and privacy concerns. They should also build multi-stakeholder bodies that could monitor the development and deployment of content management tools under strict adherence to best practices in grievance mechanisms, as outlined in Principle 31 of the Guiding Principles.

## Automation and content management

In 2017 we noted a marked increase in public discussion about how technology companies and governments are using algorithms to deal with hate speech, violent extremism, false news, child pornography, and more. To develop rights-respecting policy and ensure that artificial intelligence (AI) and machine-assisted decisions do not harm human rights, it is imperative to put in place measures for transparency and accountability. Otherwise, neither the general public nor decision-makers will have the necessary information to prevent harm. Access Now is especially concerned about the haphazard, uncoordinated development of regulatory proposals for using automated technology to flag, filter, or otherwise manage content online, without a clear pathway for ensuring the public can evaluate and understand what is being proposed. We believe that this represents a serious risk for human rights, in particular to the freedom of expression.[[29]](#footnote-29)

Several proposed models or ongoing experiments for countering violent extremism (“CVE”) on the internet seek to deal with potentially harmful content through algorithmic “de-prioritization” of the content, among other types of interference. This approach can be counterproductive, since machine learning technologies used to flag content can fail to take the context into account. This approach can threaten the freedom of expression by undermining the free and open dialogue that the internet can enable.

This is particularly concerning for those in war-torn regions of the world, where activists often rely on internet communications to expose human rights violations. Platforms like YouTube allow activists to document these violations. For instance, in 2011, activists used the platform to show that a peaceful protest against President Bashar Assad’s mandate had become bloody. [[30]](#footnote-30) In June of 2017, YouTube began using a machine-learning algorithm to aid in their content removal, and subsequently took down 150,000 videos flagged for violent extremism. Since YouTube implemented this practice, Syrian Archives, an organization that works to preserve open source evidence, reported that YouTube shut down 180 channels with video content connected to Syria.[[31]](#footnote-31) Although this organization was able to recover 400,000 videos important to documenting the situation in Syria, some 150,000 videos remained in jeopardy.

Some organizations have decided to stop using YouTube and instead develop their own video archive capacity. However, given the significant cost of doing so, many organizations will continue to depend on this global platform, especially since YouTube provides activists with free technological tools for real-time uploading, editing, and translation.

Companies should not rely exclusively on automated systems for flagging content since understanding context is crucial for determining whether the content will encourage rather than discourage extremism. Instead, programs for CVE should implement a procedure that combines use of algorithms and human evaluation, and, crucially, is situated within a framework that is grounded in international human rights law and standards.

## Notification, disclosure, and transparency

The internet provides activists with a platform and opportunity to amplify their voices. However, online platforms can restrain users’ freedom of expression via terms of service policies. Notably, even if a user does not violate a specific policy, companies may still make decisions that negatively impact users’ rights, as we showed in our earlier discussion of GoDaddy’s censorship. Third parties may also request that intermediaries take adverse actions against users.

Transparency is the first step toward accountability. Since Google began the practice in 2010, “transparency reporting” has become an expected way for internet platforms to inform the public about their efforts impacting privacy and freedom of expression, especially regarding interaction with governments and other parties. The revelations of mass surveillance by former NSA contractor Edward Snowden, beginning in June 2013, catalyzed the trend, as companies understood that their business models were threatened by the widespread lack of confidence in their services. Following the summer of 2013, there was a sharp rise in the number of companies publishing transparency reports: in 2012, only six companies published transparency reports; in 2014, this number increased to 39. Today, Access Now’s Transparency Reporting Index lists transparency reports released by 68 companies worldwide.[[32]](#footnote-32) Non-financial reporting is a popular and increasingly regulated practice for companies of all types on a host of environmental, social, and governance issues. In line with this trend, transparency reports educate the public about a company’s policies for content regulation, providing an avenue for understanding, exposing, and countering rights violations.[[33]](#footnote-33)

To satisfy their international human rights obligations, companies must first ensure that policies for restricting or taking down content are enforced in an equitable and transparent manner. Companies should protect users’ rights to access information and engage in free expression by notifying them of restrictions or takedowns as expeditiously and transparently as possible. These notifications should specify the action the company has taken, and indicate whether government and/or business partners requested or ordered the restrictions. Companies should also publish their own responses to government requests, describing the nature and scope of their internal review process to the extent that is legally permitted. They should also provide the affected user with appropriate channels for expressing concerns and grievances, and respond promptly to all requests for appeal.

Transparency reports can help companies establish an arm’s length relationship with third parties regarding content and data requests. Yet the format and depth of these reports vary widely. Without a single legal, compliance, or regulatory standard to adhere to, the content of the reports is left to companies to decide. According to Ranking Digital Rights, no company in their 2015 Corporate Accountability Index disclosed “any information whatsoever about the volume and type of user content that is deleted or blocked when enforcing its own terms of service.”[[34]](#footnote-34) Without much reporting by companies on their own rule enforcement, observers are left to compile anecdotal evidence, which often shows inconsistency in corporate behavior. For example, the onlinecensorhip.org project “seeks to encourage social media companies to operate with greater transparency and accountability toward their users as they make decisions that regulate speech.”[[35]](#footnote-35) To continue improving, transparency reports should be expanded to provide a fuller picture of the discretionary, internal processes that tech companies, and any other companies with user-generated content, employ to restrict or promote content and control user access to their platforms, even in the absence of government demands.

Issuing transparency reports does not guarantee that a company operates with respect for human rights online, nor does this mechanism alone present a full understanding of digital surveillance and censorship. However, together with a guide to the legal landscape and insights into corporate terms of service and community guidelines, these reports can provide a fact-based representation of how user rights and interests are mediated online.

## Recommendations

Following is a summary of Access Now’s responses to questions posed by Special Rapporteur Kaye for his forthcoming report, “Study on Content Regulation in the Digital Age,” listing our recommendations to assist companies in managing content while meeting their obligations under international human rights law.

**Corporate processes dealing with content regulation under law and other state measures**

* Companies should implement procedural safeguards to ensure their content management programs do not impose disproportionate and unnecessary restrictions to freedom of expression
* Intermediaries should incorporate human rights law and norms as a core consideration when they assess whether to restrict online content or accounts
* Companies should not be tasked with evaluating the legality of content in the absence of rule-of-law mechanisms
* Company staff participating in content review should be well-trained to consider context and other factors
* Intermediaries should consult outside experts and form advisory councils to aid content reviewer staff
* The staff reviewing content should be provided with sufficient psycho-social, financial, and other support to assist them in coping with evaluating difficult or offensive material
* Companies should ensure that corporate structures are in place to adequately oversee enforcement mechanisms, with Board-level expertise and robust review and audit procedures
* Government flagging of content for online platforms should be reported in a company’s transparency reports, per best practices and indicators like those identified by the Ranking Digital Rights Corporate Accountability Index
* Platforms must ensure that governments do not use flagging to circumvent legal protections for freedom of expression

False news and disinformation

* Governments should commit not to use “troll armies” to shape online discourse, either domestically or abroad
* Platforms should take appropriate measures to address this matter, including engaging in transparent enforcement of terms of service and implementing rights-respecting policies for regulation of content and accounts

Right to be forgotten

* The right to de-list should be exercised within the data protection context and limited to circumstances where the sole objective is the protection of personal data of non-public figures
* The right to de-list should not lead to the deletion of online content; although a web address may be delisted, the content must remain online
* The implementation of the right to de-list should fall solely on the data controller, the entity responsible for determining when, how, and for what purpose personal data should be processed

**Content-related requests from states that are not based on the law or a company’s terms of service**

* If law enforcement agencies are to challenge the legality of an activity, whether it occurs online or off, they should be required to pursue the matter using legal channels, processes, and procedures
* Removal of disputed content should take place only when the content has been specifically adjudicated as being illegal and a court order has been issued

**Global removals**

* Companies should ensure that their takedown and appeal policies are clear and public
* Companies should interpret governments’ jurisdictional authority appropriately narrowly to minimize the adverse impact of takedown orders on the right to freedom of expression
* Companies should avoid defaulting to a process whereby they remove content first and then notify the user of redress procedures

**Individuals at risk**

* Companies should engage with civil society actors and use their input to develop anti-harassment policies and procedures
* Online platforms should commit to allowing the use of pseudonyms in appropriate circumstances
* Platforms should require that complaints against people based on a real name policy are backed by evidence so these complaints are not abused to victimize people
* To protect digital security, platforms should let people confirm their identities without having to submit government-issued identification, which could be leaked, stolen, and used for identity theft
* Intermediaries should also notify users about what happens when they submit their identity information, including providing information about where and how their data will be stored
* Users should have the capacity to protect their identification for real name policy verification by using PGP or encrypted communication
* Companies should institute a robust appeals mechanism to ensure that when people are locked out of their accounts, they have the ability to request a second review, to submit evidence, and to discuss their investigation with staff, particularly in cases where personal safety is a concern

“Doubleswitch” attacks on human rights activists

* Internet platforms should protect their users from “Doubleswitch” attacks by enabling multi-factor authentication and encouraging users to activate it, thereby preventing hackers from gaining control of accounts in the first instance
* The method should use app-based authenticators, given that text-based phone authentication exposes the identity of the users and places them at a greater risk
* Platforms should consult with “Doubleswitch” victims and update their rules in accordance with victims’ input

**Appeals and remedies**

* Companies should implement grievance and remedy mechanisms to address users’ freedom of expression and privacy concerns
* Intermediaries should also constitute multi-stakeholder bodies who could monitor the development and deployment of content management tools

**Automation and content management**

* Companies should not rely exclusively on automated systems for flagging content since understanding context is crucial for determining whether the content will encourage rather than discourage extremism
* Companies can implement procedures that combine use of algorithms and human evaluation, but platforms should ensure that these programs are situated within a framework that is grounded in international human rights law standards

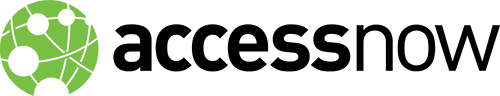
**Notification, disclosure, and transparency**

* Companies must ensure that policies for restricting or taking down content are enforced in an equitable and transparent manner
* Companies should protect users’ rights to access information and engage in free expression by notifying them of restrictions or takedowns as expeditiously and transparently as possible
* These notifications should specify the action the company has taken, and indicate whether government and/or business partners requested or ordered the restrictions
* Companies should also publish their own responses to government requests, describing the nature and scope of their internal review process to the extent that is legally permitted
* Platforms should release information on enforcement of their own terms and policies, provide the affected user with appropriate channels for expressing grievances, and respond promptly to all requests for appeal in line with best practices in human rights remedy

## Conclusion

It is imperative to preserve the open internet as a vehicle for the exercise of the rights to access to information and freedom of expression. The internet provides users with the opportunity to form and share their opinions. Therefore, it is critical to ensure that any restrictions to these rights on the internet meet the “necessary and proportionate” test, conforming to international human rights standards. Governments therefore should not pressure companies to block access, take down content, or suspend user accounts beyond what is required by law. Companies should narrowly interpret governments’ jurisdictional authority in order to minimize the adverse impact on users’ freedom of expression. They should also maintain clear and public takedown and appeal policies to avoid rights-harming takedowns and other restrictive practices. We exhort all UN member states, the Human Rights Council, the Special Rapporteur, and other stakeholders to work towards ending any unnecessary and disproportionate restrictions to the right to freedom of expression.

\*\*\*



*Access Now (*[*www.accessnow.org*](http://www.accessnow.org/)*) is an international organization that defends and extends the digital rights of users at risk around the world. By combining innovative policy, global advocacy, and direct technical support, we fight for open and secure communications for all.*

*For more information, contact:*

**Peter Micek | General Counsel |** [**peter@accessnow.org**](mailto:peter@accessnow.org)

**Raman Jit Singh Chima | Policy Director |** [**raman@accessnow.org**](mailto:raman@accessnow.org)

**Special thanks to Denis Nolasco, Legal & Policy Fellow, for coordinating this submission.**

1. About Us. *Access Now*, [www.accessnow.org/about-us/](http://www.accessnow.org/about-us/), (last visited Jan. 04, 2018). [↑](#footnote-ref-1)
2. Freedom on the Net 2016. *Freedom House*, 9 Nov. 2017, freedomhouse.org/report/freedom-net/freedom-net-2016, (last visited Jan. 8, 2018). [↑](#footnote-ref-2)
3. Freedom on the Net 2015: Privatizing Censorship, Eroding Privacy. *| Freedom House*, 28 Oct. 2015, freedomhouse.org/report/freedom-net-2015/freedom-net-2015-privatizing-censorship-eroding-privacy, (last visited Jan.8, 2018). [↑](#footnote-ref-3)
4. Lomas, Natasha. Germany’s social media hate speech law is now in effect. *TechCrunch*, TechCrunch, Oct. 2, 2017, techcrunch.com/2017/10/02/germanys-social-media-hate-speech-law-is-now-in-effect/, (last visited Jan. 8, 2018). [↑](#footnote-ref-4)
5. Deji Olukotun, Peter Micek, & Drew Mitnick, Code for tolerance: How tech companies can respond to hate but respect human rights. *Access Now*, Aug. 17, 2017, [www.accessnow.org/code-tolerance-tech-companies-can-respond-hate-respect-human-rights/](http://www.accessnow.org/code-tolerance-tech-companies-can-respond-hate-respect-human-rights/), (last visited Jan. 8, 2018). [↑](#footnote-ref-5)
6. Kaye, David. Not sure booting daily stormer is controversial. it's abt process & standards. *Twitter*, Twitter, Aug. 16, 2017, twitter.com/davidakaye/status/897963340206493698, (last visited Jan. 8, 2018). [↑](#footnote-ref-6)
7. Benedictus, Leo. Invasion of the troll armies: from Russian Trump Supporters to Turkish state stooges. *The Guardian*, Nov. 6, 2016, [www.theguardian.com/media/2016/nov/06/troll-armies-social-media-trump-russian](http://www.theguardian.com/media/2016/nov/06/troll-armies-social-media-trump-russian), (last visited Jan. 8, 2018). [↑](#footnote-ref-7)
8. White, Megan. How do you solve a problem like troll armies?, *Access Now*, Apr. 23, 2017, [www.accessnow.org/solve-problem-like-troll-armies/](http://www.accessnow.org/solve-problem-like-troll-armies/), (last visited Jan. 8, 2018). [↑](#footnote-ref-8)
9. Allcott, Hunt & Matthew Gentzkow, Social Media and Fake News in the 2016 Election, 31 J. Econ. Persp. 211 ,223 (Spring 2016), <https://web.stanford.edu/~gentzkow/research/fakenews.pdf>, (last visited Jan. 8, 2018). [↑](#footnote-ref-9)
10. Hern, Alex. “ECJ to rule on whether 'right to be forgotten' can stretch beyond EU”, *The Guardian*, July 20, 2017, [www.theguardian.com/technology/2017/jul/20/ecj-ruling-google-right-to-be-forgotten-beyond-eu-france-data-removed](http://www.theguardian.com/technology/2017/jul/20/ecj-ruling-google-right-to-be-forgotten-beyond-eu-france-data-removed), (last visited Jan. 8, 2018). [↑](#footnote-ref-10)
11. Access Now Position Paper: Understanding the “Right to be Forgotten” Globally, Sept. 2016, <https://www.accessnow.org/cms/assets/uploads/2017/09/RTBF_Sep_2016.pdf>, (last visited Jan. 8, 2018). [↑](#footnote-ref-11)
12. Id. [↑](#footnote-ref-12)
13. Mapping Report on Freedom of Expression, States and the Private Sector in the Digital Age. *Freedex*, freedex.org/new-report-on-freedom-of-expression-states-and-the-private-sector-in-the-digital-age/, (last visited Jan. 8, 2018) [↑](#footnote-ref-13)
14. Krahulcova, Lucie, Europol's Internet Referral Unit risks harming rights and feeding extremism. *Access Now*, Nov. 2, 2016, [www.accessnow.org/europols-internet-referral-unit-risks-harming-rights-isolating-extremists/](http://www.accessnow.org/europols-internet-referral-unit-risks-harming-rights-isolating-extremists/), (last visited Jan. 8, 2018). [↑](#footnote-ref-14)
15. Micek, Peter. Access and partners demand DHS investigate GoDaddy domain takedown. *Access Now*, Nov. 7 2015, [www.accessnow.org/access-and-partners-demand-dhs-investigate-godaddy-domain-takedown/](http://www.accessnow.org/access-and-partners-demand-dhs-investigate-godaddy-domain-takedown/), (last visited Jan. 8, 2018) [↑](#footnote-ref-15)
16. Welch, Chris. Google received over 75 million copyright takedown requests in February. *The Verge*, The Verge, Mar. 7, 2016, [www.theverge.com/2016/3/7/11172516/google-takedown-requests-75-million](http://www.theverge.com/2016/3/7/11172516/google-takedown-requests-75-million), (last visited Jan. 8, 2018) [↑](#footnote-ref-16)
17. Google Inc. *v.* Equustek Solutions Inc., 2017 SCC 34. [↑](#footnote-ref-17)
18. Deahl, Dani. Canada's Supreme Court rules Google must block certain search results worldwide. *The Verge*, The Verge, June 28, 2017, www.theverge.com/2017/6/28/15888408/canada-supreme-court-google-block-search-results-equustek-datalink. [↑](#footnote-ref-18)
19. Li, Anqui, & Kim Burton. Don't panic! Download. *Access Now*, May 4, 2017, [www.accessnow.org/a-first-look-at-digital-security/](http://www.accessnow.org/a-first-look-at-digital-security/), (last visited Jan. 8, 2018). [↑](#footnote-ref-19)
20. Keeping the internet safe for women and marginalized communities in India. *Access Now*, 25 May 2017, www.accessnow.org/keeping-internet-safe-women-marginalized-communities-india. [↑](#footnote-ref-20)
21. Appendix to October 5, 2015 Coalition Letter to Facebook. *Electronic Frontier Foundation*, Oct. 5, 2015, www.eff.org/document/appendix-october-5-2015-coalition-letter-facebook. [↑](#footnote-ref-21)
22. Micek, Peter ,Nameless Coalition calls on Facebook to change its real name policy, *Access Now*, Dec. 4, 2015, [www.accessnow.org/nameless-coalition-calls-on-facebook-to-change-its-real-name-policy/](http://www.accessnow.org/nameless-coalition-calls-on-facebook-to-change-its-real-name-policy/), (last visited Jan. 8, 2018). [↑](#footnote-ref-22)
23. Open Letter to Facebook About its Real Names Policy, *Electronic Frontier Foundation*, Oct. 30, 2015, [www.eff.org/document/open-letter-facebook-about-its-real-names-policy](http://www.eff.org/document/open-letter-facebook-about-its-real-names-policy), (last visited Jan. 8, 2018) [↑](#footnote-ref-23)
24. Winchester, Nick. Nameless Coalition' Challenges Facebook's Real-Name Policy. *Newsweek*, May 12, 2017, [www.newsweek.com/facebookreal-name-policytransgenderdrag-queensnative-599142](http://www.newsweek.com/facebookreal-name-policytransgenderdrag-queensnative-599142), (last visited Jan. 8, 2018). [↑](#footnote-ref-24)
25. The “Doubleswitch” social media attack: a threat to advocates in Venezuela and worldwide. *Access Now*, June 19, 2017, [www.accessnow.org/doubleswitch-attack/](http://www.accessnow.org/doubleswitch-attack/), (last visited Jan. 8, 2018). [↑](#footnote-ref-25)
26. Human Rights Council Res. 17/4, Rep. of the Hum. Rts. Council, 17th Sess., June 16, 2011, U.N. Doc. A/HRC/RES/17/4 (July 6, 2011); Special Rep. of the Sec'yGen., Guiding Principles on Business and Human Rights: Implementing the United Nations “Protect, Respect and Remedy” Framework, Hum. Rts. Council, U.N. Doc. A/HRC/17/31 (Mar. 21, 2011) [↑](#footnote-ref-26)
27. Remediation, Grievance Mechanisms and the Corporate Responsibility to Respect Human Rights. *Shift*, [www.shiftproject.org/resources/publications/remediation-grievance-mechanisms-corporate-responsibility-respect-human-rights/](http://www.shiftproject.org/resources/publications/remediation-grievance-mechanisms-corporate-responsibility-respect-human-rights/), (last visited Jan. 8, 2018). [↑](#footnote-ref-27)
28. Peter Micek & Jeff Landale. Forgotten Pillar: The Telco Remedy Plan, *Access Now*. Available at https://s3.amazonaws.com/access.3cdn.net/fd15c4d607cc2cbe39\_0nm6ii982.pdf. [↑](#footnote-ref-28)
29. Masse, Estelle, & Fanny Hidvegi. Artificial Intelligence: what are the issues for digital rights?, *Access Now*, Oct. 24, 2017, [www.accessnow.org/artificial-intelligence-issues-digital-rights/](http://www.accessnow.org/artificial-intelligence-issues-digital-rights/), (last visited Jan. 8, 2018) [↑](#footnote-ref-29)
30. AP. Activists worry YouTube erasing proof of Syria atrocities. *CBS News*, CBS Interactive, Sept. 13, 2017, [www.cbsnews.com/news/youtube-videos-syria-war-activists-human-rights-violations-war-crimes/](http://www.cbsnews.com/news/youtube-videos-syria-war-activists-human-rights-violations-war-crimes/), (last visited Jan. 8, 2018). [↑](#footnote-ref-30)
31. Expanding our work against abuse of our platform. *Official YouTube Blog*, Dec. 4, 2017, youtube.googleblog.com/2017/12/expanding-our-work-against-abuse-of-our.html, (last visited Jan. 8, 2018) [↑](#footnote-ref-31)
32. Transparency Reporting Index. *Access Now*, [www.accessnow.org/transparency-reporting-index/](http://www.accessnow.org/transparency-reporting-index/), (last visited Jan. 8, 2018). [↑](#footnote-ref-32)
33. Micek P., Aydin D.D. (2017). “Non-financial Disclosures in the Tech Sector: Furthering the Trend,” in Taddeo M., Floridi L. (eds) *The Responsibilities of Online Service Providers*. Law, Governance and Technology Series, vol 31. Springer, Cham, <https://link.springer.com/chapter/10.1007/978-3-319-47852-4\_13>. [↑](#footnote-ref-33)
34. Indicators - Ranking Digital Rights. (2015). Retrieved from https://rankingdigitalrights.org/ project-documents/2015-indicators [↑](#footnote-ref-34)
35. Onlinecensorship.org. (2016). Retrieved from https://onlinecensorship.org [↑](#footnote-ref-35)