



NUTS & BOLTS *of* **USER-GENERATED CONTENT**

The International Intermediary Liability Framework

A primer for policymakers on the laws surrounding user-generated content and intermediary liability on the Internet.

JULY 2018

presented by:



Engine



**CHARLES
KOCH
INSTITUTE**

INTRODUCTION

Over the past two decades, the growth of the Internet means that every person with a connected device can post their entire lives online. And many do. Every second, Internet users across the world are posting their own content online, everything from their favorite restaurant to how-to videos on creating custom 3D printed objects. This boom in speech has enabled e-commerce to flourish, established trust between strangers, and fueled creativity in ways never imagined possible. To understand what drove the innovation that led to the success of the Internet, we seek to unpack the nuts and bolts of user-generated content and how it is regulated, moderated, and limited by laws and technology.

Innovative platforms and startups continue to invent new ways for users to connect, share, and create. Despite all of these tremendous successes, regulators cannot ignore the fact that not all content is beneficial to society. Policymakers must consider the need to promote free speech and drive innovation while ensuring user safety. Technology is part of the answer to many of these problems, but most content moderation still requires humans to provide and understand context. Therefore, we seek to provide a complete picture of both the legal and technological questions and solutions.

This report explores the current legal framework for intermediary liability and the safe harbors that helped propel the Internet economy, including Section 230 of the Communications Decency Act and the Digital Millennium Copyright Act. It will consider how platforms moderate content and the tools and technology currently available for platforms large and small. Finally, it will look at the international landscape and how other countries are looking to regulate user-generated content.

CONTENTS



Introduction	1
Internet Safe Harbors	3
Tools for Content Moderation	5
Issues for Policymakers to Consider	7
User-generated Content Worldwide	9
International Framework	11
Conclusion	12
About Engine	14

INTERNET SAFE HARBORS

There are two laws that govern user-generated content and are commonly referred to as the “Internet Safe Harbors”:

1 **Section 230 of The Communications Decency Act**, which applies broadly to content posted to Internet platforms

2 The **Digital Millennium Copyright Act**, which applies to copyrighted works

How the DIGITAL MILLENNIUM COPYRIGHT ACT Handles User-Generated Content

To address the copyright infringement online, Congress passed the Digital Millennium Copyright Act (DMCA) in 1998, which created rules by which websites must handle claims of user copyright infringement. Congress recognized that the Internet could be a powerful tool to allow creators to share content online but, also could be used to facilitate copyright infringement. Because existing copyright law could have subjected websites to massive liability for user infringements, even if they weren't aware of those infringements, Congress sought to balance the interests of copyright holders in stopping infringement and allowing online platforms to grow without fear of unforeseen liability. To address this tension, Congress created the DMCA's safe harbor regime, which provides legal protections for platforms that remove content in response to notifications of infringement.

The DMCA establishes that a website can be held secondarily liable for its users' infringements if it has actual knowledge of specific infringements and does not act quickly to disable

access to them. To help clarify when a website has actual knowledge of infringements, the DMCA also created a “notice-and-takedown” procedure by which copyright owners can formally put websites on notice of alleged infringements. If a website receives a legitimate DMCA takedown notice, it must expeditiously disable access to the accused material or face potential liability for infringement. Since penalties for secondary copyright infringement can amount to as much as \$150,000 per work infringed, websites have strong incentives to quickly process every legitimate takedown notice they receive.

What is Notice-and-Takedown?

- The DMCA **provides a mechanism** for copyright owners to put websites on notice of infringements by sending a takedown notice.
- Websites must **disable access to disputed content they are notified of** or risk potentially statutory damages.

How SECTION 230 of THE COMMUNICATIONS DECENCY ACT shaped the Internet

Often described as the “26 words that created the Internet,” Section 230 of the Communications Decency Act, passed in 1996, enabled online platforms to host user-generated content without being held legally responsible for the speech of their users. Section 230 shields websites from liability for user-generated speech and gives platforms breathing room to find and remove

objectionable content without fear of burdensome litigation. For a startup, Section 230 guarantees that a website that gives users a forum to express themselves freely will not face ruinous legal liability if a bad actor says something illegal. This has allowed tens of thousands of startups to build online platforms where users can create, post, and share media of all kinds.

The proliferation of user-generated content happened thanks to the important protections received by online platforms under Section 230. Those protections are a key reason that the United States has been home to the vast majority of top Internet companies. As the pace of innovation accelerates, Section 230 remains as important today as it did when it was passed two decades ago. While some large Internet companies may be in their teenage years, a quick glance at a typical smartphone shows

What Does Section 230 Do?

- It does **establish a uniform regulatory regime**, rather than a 50-state patchwork.
- It does **prevent frivolous litigation**.
- It does **empower platforms to proactively monitor** for objectionable content.

dozens of apps that were invented in the past few years. Startups less than five years old have reinvented the way we share photos, send money, date, order food, and rent our homes. All of these apps rely on user-generated content, and Section 230 has facilitated their growth in multiple ways.

What Does Section 230 NOT Do?

- It does not prevent the Department of Justice from prosecuting **violations of Federal Criminal Law**.
- It does not protect a platform from liability if it **develops illegal content**.
- It does not apply to **intellectual property law**.



The Good Samaritan Doctrine

Section 230(c)(1) of the Communications Decency Act states that websites aren't liable for third party content. But Congress also wanted to create an incentive for platforms to monitor and remove anti-social content. To encourage platforms to engage in proactive content moderation, Congress also included 230(c)(2), which states that websites should not face liability for filtering the content they consider offensive.

What does Section 230 mean for startups?

- **Fewer lawyers:** not facing a lawsuit every time a user posts illegal content
- **Fewer content moderators:** can launch a website without a team of human reviewers
- **More innovation:** can create new platforms without excessive regulation
- **More speech:** a small platform can host millions of users, enabling expression

TOOLS FOR CONTENT MODERATION

While section 230 enables small platforms to host vast quantities of content, this comes with a number of added responsibilities and challenges. Internet platforms find themselves having to consider competing interests when trying to manage online conversations with millions of users. As a result, platforms have developed tools and techniques to filter and remove objectionable material. However, these methods are still being refined, and platforms can easily find themselves on the receiving end of criticism that they either remove too much or too little, often at the same time.

How do Content Moderation Tools Work?

Websites can attempt to moderate content by using filtering technologies to identify and screen out certain content uploaded to their platforms. The algorithms built into these technologies usually do this by analyzing physical characteristics of the content in question (e.g. image, sound or text) and/or associated metadata (e.g. file name, size, posting time and other features which are not visible in the content but help to identify whether it is violating a law or not). Then, the software compares this info to existing databases to determine whether the identified content matches material that has been designated as copyright infringing or some other category of objectionable content. In short, filtering tools generally can only tell whether a piece of scanned content is the same as another piece of content identified in a database. That is, they can determine whether an uploaded file is a copy of a particular movie, but they cannot determine whether the use of that file constitutes a copyright infringement. As such, filtering tools are incredibly limited in their ability to moderate user-generated content, since they leave unanswered many critical questions about whether or not to remove particular content. More nuanced content moderation decisions must inevitably be done by human reviewers.

Natural Language Processing

While most filtering tools merely determine whether a piece of content matches another piece of pre-identified material, so-called “natural language processing” (NLP) tools attempt to analyze text data to identify linguistic characteristics and determine whether content is acceptable or not. This is most often done by identifying positive and negative sentiments. Since much of social media is text-based, NLP can be used to filter uploads on social media platforms and identify problematic contents such as hate speech, online bullying, slurs, and disinformation. However, to identify these types of content accurately, implicit features are just as important as the explicit ones.

Context, tone, and dialect all matter.

Consider the following scenarios, would you take it down?



Someone shares an image from a terrorist group’s social media profile with comments criticizing it. Is this **spreading online extremist content**?



Someone uses a slur to describe themselves. Is this **hate speech**?



Someone mentions another social media user and quotes lyrics threatening violence. Is this a **credible threat**?

Automated tools like filtering algorithms are quickly thrown out as an “easy” solution to user-generated content problems. While automated tools can be useful for addressing certain types of illegal content, it is important to recognize the inherent limitations of these solutions.

Technical Limitations of Automated Content Moderation



Misidentifications: Content moderation tools frequently identify legal user-generated content as copyright infringement. Filtering tools are unable to decipher works that could be protected by copyright limitations and exceptions like fair use, parodies and remixes.



Issues in the data: Most filtering tools merely match fixed characteristics of a sample against a database. Thus, any biases or issues with existing data will be replicated by the filtering tools. One common example of this in the NLP space is variations in the English dialect. Certain accents or ways of speaking tend to be over-represented in data, and therefore speech from minority groups or non-native English speakers may be unfairly targeted by these tools.



Context: Few tools are advanced enough to detect the contextual meaning of uploads e.g. if a post is satirical or quoting another individual/organization, the tool still might flag the content as problematic.



Avoiding detection: It is relatively simple to avoid detection by filtering tools by manipulating the data contained in media files.



Applicability across domain areas: Since these tools are typically “trained” using existing datasets, it is hard to find a solution that will work across different thematic areas, demographic groups or Internet platforms. Individuals express themselves differently in each of these scenarios. The tools which are marketed as generic solutions tend to have low accuracy rates.



Encryption: Filtering tools require access to the underlying content, rendering them unable to process encrypted files. One of the most secure ways to protect data today is encryption, and filtering tools are still unable to analyze encrypted files.

Most recently, the broader conversation about Internet platforms’ responsibilities has zeroed in on Facebook and the role the social media giant played in the 2016 election as an advertising platform and collector of Internet users’ data. While testifying to Congress in April 2018, Facebook CEO Mark Zuckerberg told lawmakers that his company would be hiring an additional 10,000 employees to monitor content on the platform. This underscores how any Internet platform—including one with the deepest pockets and most advanced technological capabilities—has to rely, at the end of the day, on people to make decisions about specific user-generated content should be allowed online. Sophisticated technologies alone, are not enough.

ISSUES FOR POLICYMAKERS TO CONSIDER

Technology Can Only Do So Much

Filtering technologies come with a few inherent limitations and issues for policymakers to keep in mind.

1. **Costs.** These technologies are expensive for Internet platforms to acquire and implement. Their high cost creates a barrier to entry in the Internet platform market and will limit the number of startups that can compete with established platforms.
2. **Accuracy.** Filtering technologies are notoriously inaccurate. Automated processes result in the systematic over-removal of non-objectionable content, and limited source data can create inherent biases against specific types of speech.
3. **Limited Scope.** Filtering technologies only work on a few types of content including text, audio, and video. Existing filtering techniques do not generally work on different media, requiring platforms to purchase or create several different content-specific tools. No such tools even exist for many types of content, such as software and 3d models.

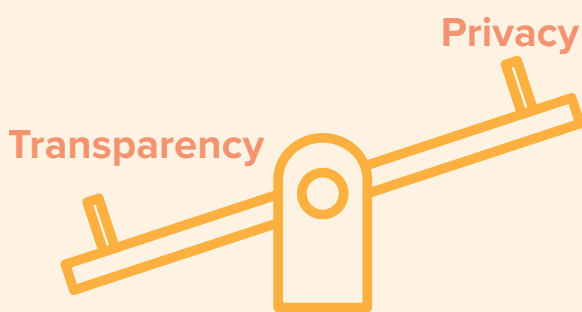
What Conclusion Can We Draw?

Platforms are bad at sorting out the truth 



One person's truth is another's injury

Need to balance:



User-generated content **allows innovators to share and promote ideas** in a new and seamless way.

Permissionless innovation flourishes when entrepreneurs can **collaborate and critique openly on the Internet**.

But there **need to be limits** placed on user-generated content to satisfy competing policy concerns.

Law Enforcement.

Criticism that the Internet makes it easier for criminals to find information they need to commit crimes is not entirely misplaced. However, with billions of people sharing every detail of their life online, law enforcement officers now have more tools than ever before to find and convict criminals. Restricting the speech of billions of law-abiding citizens because there are a few bad actors will only end up making the task of stopping criminals harder.

Civil Society.

Internet platforms and user-generated content has fundamentally transformed the way citizens engage with government, whether we are thinking about the Arab Spring or the #MeToo movement, civil society has a vested interest in protecting a free and open Internet that allows citizens to engage in public activism.

Protecting Users.

Unfortunately, not every piece of user-generated content is harmless or should be promoted. For instances like hate speech, revenge porn, defamation, and others, the policymakers need to consider a legal framework that protects users from anti-social behavior like bullying and disinformation campaigns.

USER-GENERATED CONTENT WORLDWIDE

As more and more people around the world access the Internet, different countries are taking different approaches to regulating user-generated content. Here are some examples of countries that are thinking about intermediary liability differently than the United States.



Canada:

Notice-and-Notice. The Canadian intermediary liability system in the copyright space establishes a notice and notice system which requires that the intermediary forward the notification of infringement to the user that posted the allegedly infringing materials. While this bodes

well for user-generated content generally, the differences between U.S. and Canadian law may be difficult to resolve should NAFTA renegotiations contemplate intermediary safe harbors.

Argentina:

Platform Monitoring Obligations. Argentina's High Court was the first in Latin America to draw a bright-line rule rejecting an obligation for platforms to monitor user-generated content. Setting a strong precedent for free speech online, the Court ruled that Internet platforms must remove content only after a court or other government authority has resolved a legal conflict and ordered removal.



European Union:

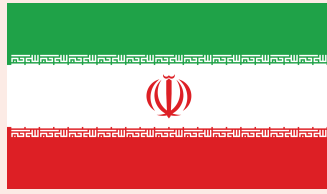
Right to be Forgotten. A 2014 decision from the European Union's Court of Justice imposed new obligations on search engines, which under EU law may also apply to social media. For the EU, the Right to be Forgotten allows individuals to ask platforms to delist certain search results relating to a person's name if the information is "inadequate, irrelevant, or excessive in relation to the purposes of the processing."



Germany:

Hate Speech Laws. Beginning in 2018, Germany started enforcing new laws surrounding hate speech on the Internet. Known as NetzDG, the law requires Internet platforms to adopt a comprehensive complaints structure so posts can be reported to staff and removed within 24 hours in some cases. Questions abound about how to interpret hate speech and some platforms have removed user-generated content en masse in fear of violating the law, harming free speech and expression.





Turkey, Iran, Saudi Arabia:

Legal Blockers. Authoritarian regimes worldwide have laws on the books that allow countries to block, filter, and take down user-generated content for political purposes. These three countries are particularly bold in enshrining in their laws the ability to block content. We see these tensions run high when political unrest brings to light the removal of political speech from Internet platforms.

South Korea:

Trade Deals Export DMCA. South Korean law enshrines many of the user-generated content safe harbor provisions from United States law, reflecting the Korea-US Free Trade Agreement (KORUS). The Intellectual Property chapter from KORUS is thought to be the gold-standard for trade agreements and has been replicated in several other negotiations. Unfortunately, South Korea recently weakened its safe harbor provisions by creating a governmental body to compel platforms to disconnect supposed repeat infringers, depriving many South Koreans of Internet access with no judicial remedy.



China:

The Great Firewall. China remains infamous in the user-generated content regulatory regimes for implementing a combination of legislative actions and technologies to regulate the Internet domestically, subjecting almost all speech online to strict censorship regimes. China also blocks access to selected foreign websites and slows down cross-border Internet traffic. Fear of regulators' disfavor leads many intermediaries to preemptively take down even more content than the law requires.



Thailand:

Lèse majesté. With some of the strictest laws in the world on insulting the King, Thailand has been very aggressive in policing user-generated content that mentions the royal family. Without any real intermediary liability laws on the books, the Thai government is free to demand platforms remove content critical or embarrassing of the King in the name of “national security” and threaten legal action against those that don't.



**For more information on global intermediary liability and user-generated content developments, you can check out the Stanford CIS World Intermediary Liability Map: <http://wilmap.law.stanford.edu/>*

INTERNATIONAL FRAMEWORK

The intermediary liability framework governing online platforms in the United States has tremendous reach worldwide as many of the largest hosts for user-generated content are U.S. companies. As the Internet spreads worldwide, countries are grappling with their own unique policy decisions regarding user-generated content. Countries setting up intermediary liability laws for the first time should consider a few factors surrounding the policies impacting user-generated content.

Growing the Digital Economy.

The Internet enables businesses to reach global audiences in ways never possible before. User-generated content can increase trust, facilitate transactions, and open new markets for anyone with Internet access, regardless of citizenship. Even small businesses can be global, and countries should try to promote the free flow of user-generated content. The U.S. tech sector is the envy of the world and laws like Section 230 and the DMCA have allowed innovation to flourish. Across the globe, policymakers are looking for ways to replicate the successes of Silicon Valley and creating a flexible framework for user-generated content can help them get there.

Trade.

The proliferation of user-generated content has created new opportunities for cross-border trade and investment. As such, many trade agreements now consider rules surrounding intermediary liability and intellectual property rights. We are seeing a global trend to add these issues to the calculus when discussing how to promote and protect domestic industries and access to new markets. Here are just a few examples of recent trade deals:

- **North American Free Trade Agreement (NAFTA) 2.0:** In this case the United States, Canada, and Mexico all have different approaches to regulating online platforms and the original NAFTA deal (struck in the same year as the enactment of the DMCA and Section 230) does not consider intermediary liability issues.
- **Trans-Pacific Partnership (TPP):** TPP negotiators are considering placing a greater burden on platforms for both intermediary liability and copyright protections. While negotiations for TPP have continued without the U.S., the agreement does not appear to be friendly towards the promotion of user-generated content.
- **Trade in Services Agreement (TiSA):** Current negotiations for TiSA include 23 members of the World Trade Organization and the purpose is to increase the flow of e-commerce and trade in services. TiSA parties are still negotiating to set an innovation framework.

Human Rights and Cultural Priorities.

The way a foreign body prioritizes digital rights is incredibly dependent on the way it prioritizes rights in general. In the U.S., a fundamental right to assembly and free speech creates an inherently high barrier to police users' speech online or force Internet platforms to police their users' speech themselves. But other parts of the world prioritize other rights. For instance, in Europe, the right to privacy often outweighs speech rights. This plays out in the debate over "the right to be forgotten," or the idea that Internet platforms, including search engines, should remove access to content about an individual if that individual has requested it be taken down.

CONCLUSION

In our increasingly connected world, questions about the responsibilities of Internet platforms have serious and widespread consequences. As policymakers and Internet companies think through regulating content online, it's necessary to fully weigh the costs and risks that come with policing user-generated content. In an increasingly digital and connected world, empowering users to share their stories, creations, and innovations, while ensuring the safety of users, is paramount.

Final Thoughts.

- The task of moderating user-generated content boils down to prioritizing protections for users. Should one user's right to speak publicly on a platform outweigh another user's right to privacy? These are not easy questions with straight-forward answers.
- Moderating user-generated content is fundamentally a human-driven task. While algorithms can help detect certain types of content, they cannot discern context or tone. There are many instances that fall into a gray area and require a human perspective.
- Companies of different sizes have different capacities to deal with moderation mandates or legal liabilities. A startup may host millions of pieces of user-generated content but only have a small team to actually sift through all of them. Legal burdens should not restrict innovation.
- The Internet is inherently global, and rules around user-generated content in one country will impact what Internet platforms can offer users in another country.



Engine was created in 2011 by a collection of startup CEOs, early-stage venture investors, and technology policy experts who believe that innovation and entrepreneurship are driven by small startups, competing in open, competitive markets where they can challenge dominant incumbents. We believe that entrepreneurship and innovation have stood at the core of what helps build great societies and economies, and such entrepreneurship and invention has historically been driven by small startups. Working with our ever-growing network of entrepreneurs, startups, venture capitalists, technologists, and technology policy experts across the United States, Engine ensures that the voice of the startup community is heard by policymakers at all levels of government. When startups speak, policymakers listen.

The logo for the Charles Koch Institute features a blue square outline. Inside the square, the words "CHARLES", "KOCH", and "INSTITUTE" are stacked vertically in a blue, sans-serif font.

**CHARLES
KOCH
INSTITUTE**

For more than five decades, Charles Koch's philanthropy has inspired bold new ideas to improve American lives. Inspired by a recognition that free people are capable of extraordinary things, the Charles Koch Institute supports educational programs and dialogue to advance these principles, challenge convention, and eliminate barriers that stifle creativity and progress. We offer educational programs, paid internships, and job placement assistance to students and professionals, and encourage civil discussion about important issues like free speech, foreign policy, and criminal justice reform. In all of our programs, we are dedicated to identifying new perspectives and ideas that help people accomplish great things for themselves and others.



Engine

engine.is  @EngineOrg

