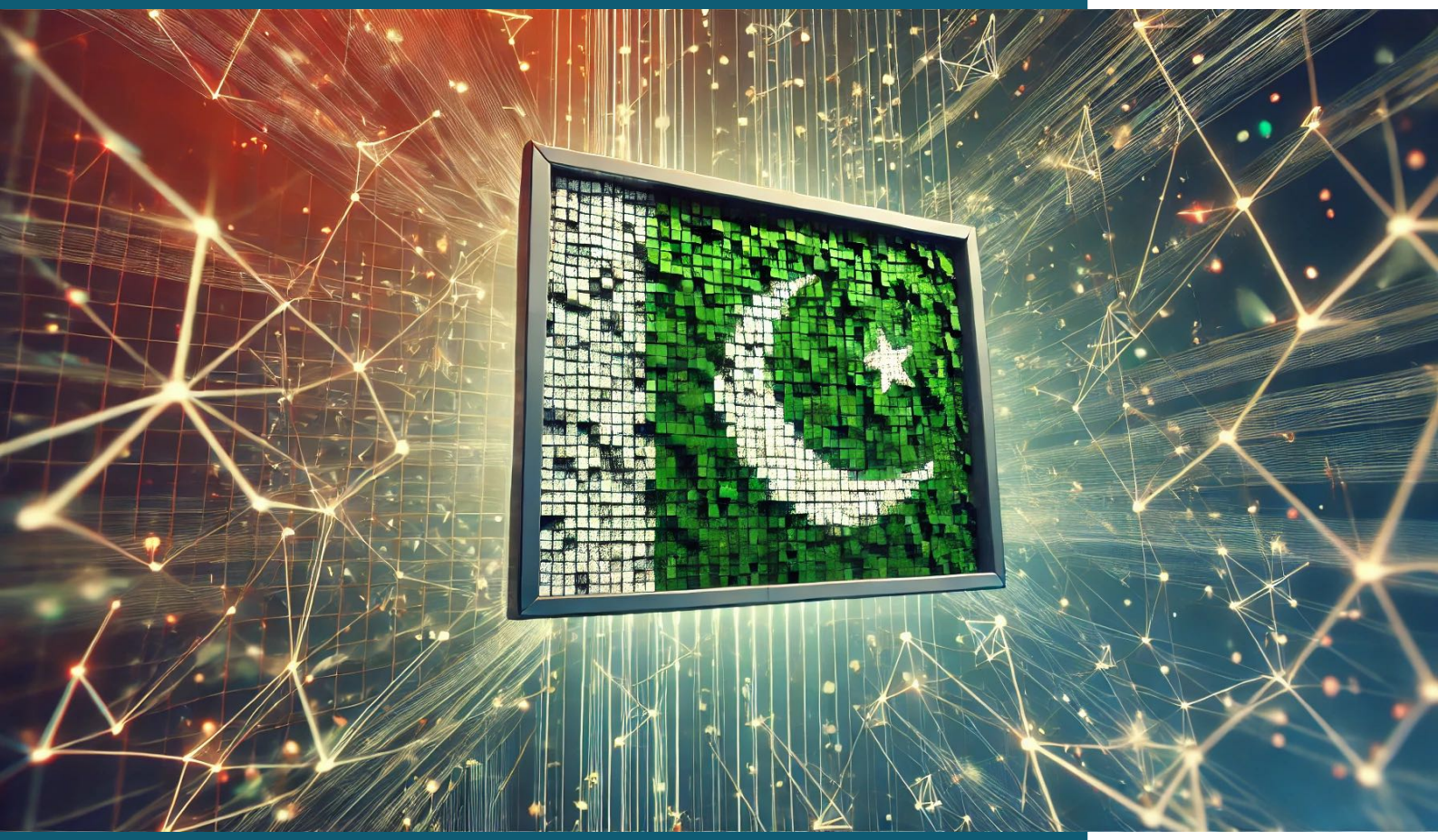

Global Perspectives on Content Regulation:
Lessons for Pakistan

BOLO BHI
Advocacy - Policy - Research

samuelson-glushko
tech.law.policy
clinic

About Bolo Bhi

Bolo Bhi is a civil society organisation geared towards advocacy, policy, and research in the areas of digital rights and civic responsibility. This encompasses the right to information, free speech, and privacy online, so that the internet can be realized as a free and representative space for civic and political engagement for all segments of society, including marginalized communities and genders. Bolo Bhi believes that an informed citizenry with the knowledge, skills, tools and disposition towards civic engagement is integral for effective government transparency and accountability. For more information, visit our website at **bolobhi.org**.

About TLPC

Based at the University of Colorado Law School, the Samuelson-Glushko Technology Law & Policy Clinic (TLPC) advocates for technology laws and policies that advance the public at the international, national, and local levels. The Clinic addresses a wide range of issues, from civil and human rights and intellectual property law. TLPC engages in discussions on technology law and policy by representing clients before administrative, judicial, and legislative bodies, participating in amicus advocacy, conducting significant public policy research, and forming strategic partnerships with advocacy groups, public officials, and multistakeholder groups to effect change.

Date of Publication: December 2024

Cover image art was created using OpenAI's DALL-E image generator

Table of Contents

Foreword

Farieha Aziz[1] and Vivek Krishnamurthy[2]

Since the beginning of 2024, network disruptions and slow internet speeds have become the norm in Pakistan. Various reasons have been cited, including faults in undersea cables, upgrades to a web-management system, and the installation of a national firewall—sometimes referred to interchangeably.

Government officials and regulators have provided contradictory accounts, neither fully acknowledging the deployments nor clarifying their exact nature. Nevertheless, their narrative consistently emphasizes that firewalls exist in many countries and that content regulation, including restrictions on speech about state institutions and public officials, is a global norm. The justification for these measures is typically framed in terms of national security, counterterrorism, and the fight against "fake news."

Earlier in 2024, Bolo Bhi and the Samuelson-Glushko Technology Law and Policy Clinic (TLPC) at the University of Colorado Law School worked together to examine Pakistan's approach to regulating disinformation. This included analyzing changes to Pakistan's laws in comparison with international approaches and standards. *(Read: How Not to Regulate Disinformation: The 2024 General Elections and the Misregulation of Disinformation.)*

This term, TLPC students Jordan Chen, Santana Andazola, Neven Grigic, and Natalie Phillips explored content regulation and speech restrictions in the context of claims that such practices are common worldwide—even in the US and EU. The project explored whether Pakistan's restrictions—implemented through both technological measures and legislative

[1] Co-Founder, Bolo Bhi.

[2] Associate Professor of Law and Director, Samuelson-Glushko Technology Law & Policy Clinic, University of Colorado Law School.

changes—align with those in other jurisdictions. The resulting memoranda, which are compiled in this document, provide insights into what actually happens in other regions and highlight how these practices differ from Pakistan's approach.

How the U.S. Detects and Removes Unlawful Content from the Internet

Jordan Chen[3]

1. Introduction

This memorandum examines how unlawful online content in the United States (U.S.) is detected and removed.

Following recent internet disruptions in Pakistan, government officials there have claimed that the U.S. government possesses vast authorities to remove unlawful content from the internet. This claim is false.

Due to the strong free speech protections provided for in the First Amendment to the U.S. Constitution, there are only a select few categories of online content that are considered unlawful. Furthermore, the detection and removal of illegal online content in the U.S. is dominated by private efforts, typically by the platforms/providers, who must follow applicable statutes in doing so.

This memorandum outlines the detection and removal process for three main types of online content that are actionable under U.S. law: copyright infringement, child sexual abuse materials (CSAM), and tortious content (e.g. defamatory content or content that invades privacy). Copyright and CSAM are subject to takedown procedures operated by platforms, while tortious content may only be removed upon the issuance of a court order.

2. Discussion

U.S. law recognizes three main kinds of unlawful online content: copyright infringement, CSAM, and tortious content.

---

[3] J.D. Candidate, Class of 2026, University of Colorado Law School.

For copyright infringement, under the Digital Millenium Copyright Act, there are notice and takedown procedures available as a remedy when a platform is found to host such content.[4] For CSAM, once providers gain knowledge of such content, they must take it down and report it to the National Center for Missing and Exploited Children (NCMEC).[5]

There are no such mechanisms for tortious content, by contrast. Removal of this content requires one to file a lawsuit, prove in court that the content is tortious, and obtain a court order for its removal.[6] Because of Section 230 of the Communications Decency Act, providers aren't subject to court orders to take down content that has been judged to be tortious.[7] Instead, to remove such content, courts must order the original person who posted it to take it down.[8]

To be sure, most large online providers have implemented ways for users to report content that violates the law or their own terms of service for removal.[9] But such mechanisms are not mandated by U.S. law.

2.1. Intellectual Property:

U.S. law provides a self-help system for alleged victims of copyright violations. Only the copyright owner or someone authorized to act on their behalf may seek the removal of infringing content.[10] The process begins with the copyright owner (or their agent) sending a notice to a provider

---

[4] *Digital Millennium Copyright Act (DMCA),* Pub. L. No. 105-304, tit. II, 112 Stat. 2860 *(1998)* (codified at 17 U.S.C. §512).

[5] 18 U.S.C. §§ 2258A, 2258B.

[6] *Removing Defamation From Search Engine Results*, Katz Law Group, P.C., https://www.katzlawgroup.com/removing-defamation-from-search-engine-results (2024).

[7] *Hassell v. Bird*, 5 Cal. 5th 522 (2018)

[8] *Id*. at 547.

[9] *See Report Content for Legal Reasons*, Google, https://support.google.com/legal/answer/3110420?hl=en (2024).

[10] *Can I send a DMCA Takedown Notice*, Copyright Alliance, https://copyrightalliance.org/faqs/can-i-send-a-dmca-takedown-notice/ (2024).

requesting them to take down the infringing material.[11] To avoid copyright liability, the provider must promptly take down the material and forward the notice to the person that posted the material.[12] Should that person file a counter-notice contesting the copyright infringement claims, the complainant has 14 days to bring a lawsuit against the alleged infringer.[13] If the complainant does nothing, the platform must then restore the material.[14]

2.2. Child Sexual Exploitation Material (CSAM)

It is illegal under U.S. law to produce, knowingly distribute, or knowingly receive CSAM,[15] however the work of removing CSAM from the internet is done in the U.S. by internet platforms, rather than the government itself.

When a provider gains actual knowledge of CSAM on their services, the law requires them to report it to the National Center for Missing & Exploited Children (NCMEC)—a government funded non-profit.[16] If a provider with "actual knowledge" of CSAM fails to report it "intentionally" or "recklessly, with actual malice," the provider may be held criminally and civilly liable.[17] Additionally, providers must ensure that depictions of CSAM are permanently destroyed upon a valid request from a law enforcement agency.[18]

Notably, U.S. law does not require providers to proactively detect CSAM; they are only required to report it to NCMEC when they have become aware of it. Even so, many providers have voluntarily implemented

---

[11] *What is the DMCA Notice and Takedown Process*, Copyright Alliance, https://copyrightalliance.org/faqs/what-is-dmca-takedown-notice-process/ (2024).
[12] *Id.*
[13] Shelly Garcia, *DMCA Takedown Notices: What They Are and How to Respond*, NOLO, https://www.nolo.com/legal-encyclopedia/responding-dmca-takedown-notice.html#_Toc119581340 (2024).
[14] *Id.*
[15] 18 U.S.C. §§ 2251-2252A.
[16] 18 U.S.C. § 2258A.
[17] 18 U.S.C.S. § 2258B.
[18] *Id*.

mechanisms for users to report CSAM to the providers, law enforcement, or NCMEC.[19]

3. Conclusion

Based on the above, it is clear that claims by Pakistan government officials regarding the powers of the U.S. government to remove illegal content from the internet are categorically false.

[19] *Report Content for Legal Reasons*, Google, https://support.google.com/legal/answer/3110420?hl=en (2024).

The U.S. Approach to Regulating Content Critical of the Government

Santana Andazola[20]

1. Introduction

Under U.S. law, "anti-state" and "anti-institution" expression is undoubtedly constitutionally protected speech. In *New York Times v. Sullivan*, the U.S. Supreme Court articulated this in declaring that the Sedition Act of 1798[21] violated the First Amendment "because of the restraint it imposed upon criticism of government and public officials."[22]

The United States affords extensive protection to speech and expression, including opinions and criticisms of the government, its officials, and the military. The First Amendment states "Congress shall make no law . . . abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances."[23] This protection applies to all levels of government: federal, state, and local.[24]

Despite the substantial protection that is afforded to expression, the First Amendment is not absolute. The government does have the power to regulate certain categories of expression.[25] The main categories of unprotected or less protected speech are (1) incitement of imminent

---

[20] J.D. Candidate, Class of 2026, University of Colorado Law School.

[21] The Act made it a crime "if any person shall write, print, utter, or publish…any false, scandalous, and malicious writing" about the government. (https://www.archives.gov/milestone-documents/alien-and-sedition-acts).

[22] New York Times Co. v. Sullivan, 376 U.S. 254, 276 (1964); see also New York Times Co. v. United States, 403 U.S. 713 (1971) (holding that the government may not suppress the dissemination of sensitive, and embarrassing governmental/military information).

[23] U.S. Const. amend. I.

[24] Robert C. Power & Mark C. Alexander, A Short & Happy Guide to The First Amendment 2 (2d ed. 2022).

[25] Id. at 8.

lawless action, (2) fighting words, and (3) true threats.[26] Additionally, the government has recognized less protection for libel/defamation, hate speech, and sexually explicit expression.[27]

2. First Amendment Protection of Content Critical of the Government

The First Amendment provides extensive protection for speech criticizing the government, military, its officials, and the actions of those officials. The Supreme Court has recognized that such speech is essential to the functioning of a healthy democracy by protecting the nation's commitment to robust and uninhibited political debate.[28] This was expressed in *New York Times v. Sullivan*, where the Court stated "debate on public issues should be uninhibited, robust, and wide-open, and that it may well include vehement, caustic, and sometimes unpleasantly sharp attacks on government and public officials."[29] There, the Court argued that the right of the people to criticize the government without fear of retribution is the central meaning of the First Amendment.[30]

3. Limits on Free Expression

As mentioned, there are some limits to the protection afforded by the First Amendment. However, the standard to prove that speech falls into an exception and is, therefore, unprotected is very high. For example, one recognized exception to First Amendment protection is incitement to imminent lawless action. This standard was first articulated in *Brandenburg v. Ohio*, in which the Court held that speech advocating unlawful action

---

[26] David L. Hudson Jr., Legal Almanac: The First Amendment: Freedom of Speech § 3:1 (2012).

[27] Power & Alexander, supra, note 5.

[28] Id. at 3-4.

[29] Sullivan, 376 U.S. at 270.

[30] Jerome A. Barron & C. Thomas Dienes, Barron and Dienes's First Amendment Law in a Nutshell 12 (6th ed. 2023).

could only be penalized if it is "directed to inciting or producing imminent lawless action and [be] likely to incite or produce such action."[31] This is a very difficult standard to meet as it requires proving the speaker actually intended to incite lawless action that is both imminent and likely.[32] The other recognized exceptions to First Amendment protections are similarly difficult to prove. For example, a public figure bringing a defamation action must not only prove the usual of defamation,[33] but they must also show that the defamatory statements were made with "actual malice—that is, with knowledge that it was false or with reckless disregard of whether it was false or not."[34] This standard affords strong protections to individuals or organizations criticizing public officials.

4. Conclusion

The United States provides extensive protections to speech criticizing the government and government officials. Indeed, U.S. jurisprudence explicitly protects such anti-government speech as the very heart of the First Amendment. In short, anti-state and anti-institutional propaganda are not recognized as categories of actionable speech, and content critical to the government and military enjoys robust protections under the U.S. Constitution.

---

[31] Brandenburg v. Ohio, 395 U.S. 444, 447 (1969).

[32] Power & Alexander, supra, at 20.

[33] False statement of fact, publication or communication to a third party, and damage caused to the reputation of the subject. (https://www.law.cornell.edu/wex/defamation).

[34] Sullivan at 280.

The U.S. Approach to Regulating Content Critical of the Military

Santana Andazola[35]

1. Introduction

The First Amendment's broad free speech protections safeguard speech content critical of the U.S. military. For any regulation of such speech to be deemed constitutional, the expression must fall into one of the recognized categories of unprotected speech (i.e. true threats, incitement, defamation, etc.) or the government must satisfy the strict scrutiny test. Under this test, the government must show a compelling interest or goal in regulating speech critical of the military and that the regulation of that speech is necessary to achieve that goal, the measure will be struck down as unconstitutional.[36] The government has almost never been able to satisfy the strict scrutiny test with regard to content critical of the military because of the demanding nature of the test.

2. Analysis

The First Amendment protects criticism and dissent—including of the military—to permit debate and public discussion on pressing issues in U.S. society.[37] For example, in *Cohen v. California*, the U.S. Supreme Court held that a vulgar critique of U.S. military policy was protected by the First Amendment.[38] The case considered whether Cohen could be charged for

---

[35] J.D. Candidate, Class of 2026, University of Colorado Law School.

[36] See *R.A.V. v. City of St. Paul, 505 U.S. 377 (U.S.Minn., 1992)* (holding that no new categories of unprotected speech may be added because "the danger of censorship presented…requires that that weapon be employed only where it is necessary to serve the asserted compelling interest."

[37] Robert C. Power & Mark C. Alexander, A Short & Happy Guide to The First Amendment 80 (2d ed. 2022).

[38] *Cohen v. California, 403 U.S. 15 (U.S.Cal. 1971)*.

wearing a jacket that said "Fuck the Draft" outside the Los Angeles County Courthouse as a means "of informing the public of the depth of his feelings against the Vietnam War and the draft."[39] The Court upheld Cohen's right to do so. It explained that allowing the government to restrict this speech, even just the use of the expletive, would "effectively empower a majority to silence dissidents simply as a matter of personal predilections."[40] Indeed, the Court found that protecting offensive material criticizing the military is necessary to achieve the values of the process of open debate and to avoid the government regulating offensive expression as a "guise for banning the expression of unpopular views."[41]

The First Amendment does have limits, and in theory, these limits may allow the government to restrict some expression that is critical of the military. In practice, however, it is extremely difficult for such restrictions to be upheld. For example, in *New York Times Co. v. United States* (the "Pentagon Papers" case), the government attempted to stop newspapers from publishing excerpts from a top-secret Defense Department study of the Vietnam War that had been leaked to the press.[42] The government argued that the publication should be restricted to protect national security.[43] Although the information the New York Times intended to publish included classified military documents whose release could impact national security, the Court held that "the press was protected so that it could bare the secrets of government and inform the people" since "[o]nly a free and unrestrained press can effectively expose deception in government."[44]

The Court specified that the government must clear a very high bar to place restrictions on this kind of expression. It explained that "only

---

[39] *Id.* at 16.

[40] *Id.* at 21.

[41] *Id.* at 25-26.

[42] GEOFFREY R. STONE ET. AL., THE FIRST AMENDMENT 90 (Erwin Chemerinsky et al. eds., 5th ed. 2016).

[43] *Id.*

[44] *New York Times Co. v. U.S., 403 U.S. 713, 717 (U.S.Dist.Col. 1971)*.

governmental allegation and proof that publication must inevitably, directly, and immediately cause" something extremely dangerous to happen – such as imperiling the lives of the sailors of a ship that is already out at sea – would suffice to impose restrictions on the press.[45] In other words, the threat to national security must come directly from the publication of such information and that threat must be immediate and inevitable for the government to regulate it. Even in the interest of national security, the government has a very heavy burden to prove if they want to restrict expression critical to the military.

Another way the government can justify restrictions on content critical to the military is by showing that such expression is a "true threat." A true threat encompasses speech where the speaker "means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals."[46] The government must prove that the speech was an actual threat for the speech to be regulated. However, once again this is exceedingly difficult to prove. In *Watts v. United States*, the Court protected speech critical of the military during a protest against the Vietnam War, despite the speaker saying, "if they ever make me carry a rifle the first man I want to get in my sights is [the President]."[47] The Court found that this was political hyperbole, and therefore, not a true threat – just an offensive method of stating political opposition.[48]

3. Conclusion

The First Amendment permits U.S. residents to engage in robust critiques of the U.S. military, from using offensive language in public places to criticize the military to publishing leaked military documents even when they may endanger national security. Correspondingly, laws or

---

[45] *Id.* at 726-27.
[46] *Virginia v. Black, 538 U.S. 343, 359 (2003).*
[47] *Watts v. United States, 394 U.S. 705 (1969).*
[48] *Id.* at 708.

government practices that specifically target criticisms of the military for criminal or civil penalties would be *per se* unconstitutional.

How the EU Regulates Illegal Content Under the Digital Services Act

Neven Grigic[49]

1. Introduction

Pakistan government officials have claimed that the state censorship of speech critical of governments happens everywhere. This claim is false when it comes to the European Union ("EU"), however.

Through the Digital Services Act ("DSA"),[50] the EU and its Member States aim to protect the fundamental right to free expression enshrined in the Charter of Fundamental Rights of the European Union ("Charter").[51] Neither the EU nor its Member States directly moderate user content online.

Under the DSA, the Member states' competent bodies may order online platforms to remove illegal content. However, these orders must respect due process, the rule of law, and aim only at specific pieces of content that are already posted. Furthermore, transparency reports from online platforms such as Meta (for Facebook) and X reveal that national governments rarely exercise this authority. Lastly, the DSA explicitly prohibits mass surveillance. Apart from this, online platforms carry out content moderation, aimed at identifying and removing illegal content.

---

[49] LL.M Candidate, Class of 2025, University of Colorado Law School.

[50] *The Digital Services Act (the EU Regulation 2022/2065),* EUR-Lex (Sept. 15, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/?toc=OJ%3AL%3A2022%3A277%3ATOC&uri=uriserv%3AOJ.L_.2022.277.01.0001.01.ENG.

[51] *Charter of Fundamental Rights of the European Union*, EUR-Lex (Sept. 15, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT.

2. What is the DSA?

The DSA is a regulation that is directly applicable across the EU and that supersedes national laws dealing with the same subject matter.[52] The Member States and the EU supervise and enforce the provisions of the DSA.[53]

The purpose of the DSA is to effectively protect fundamental rights enshrined in the Charter.[54] One of these fundamental rights is the freedom of expression and information enshrined in Article 11, paragraph 1. This freedom includes all forms of expression conveyed by any means of communication, including the Internet. The freedom protects the expression of any content, even if it is offensive. Further, it covers all kinds of expression, including political expression. Interference with the freedom of expression must be justified by being "necessary in a democratic society", which means that there must be a "pressing social need" to interfere. The freedom under Article 11, paragraph 1 of the Charter also includes the right to access information because that right is a prerequisite for the expression of opinion.[55]

The DSA applies to intermediary services offered to users within the EU.[56] Examples of such services are very large online platforms such as Facebook, Instagram, or X (Twitter) ("platforms").

---

[52] *Questions and Answers about the Digital Services Act*, The European Commission (Sept. 15, 2024), https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348.

[53] Article 56 of the DSA.

[54] Article 1, paragraph 1 of the DSA.

[55] *The EU Treaties and the Charter of Fundamental Rights: A Commentary*, 2132-2136 (Kellerbauer, Manuel ed., et al., Oxford University Press, 1st edition 2019).

[56] Article 2, paragraph 1 of the DSA.

3. The Regulation of Illegal Content in the EU

The DSA lacks a substantive definition of 'illegal content'. Instead, it relies on the Union's and Member States' national law which complies with the Union law to define what information is 'illegal content'.[57]

In accordance with the DSA's provisions, the Member States' competent national judicial or administrative authorities may issue orders to the platforms to act against one or more specific items of illegal content. Such orders are issued on the basis of applicable Union law, or national law in compliance with Union law.[58]

These orders must contain a number of elements, some of which will be mentioned. First, orders must have a reference to the legal basis under Union or national law. Second, there should be included a statement of reasons explaining why the information is illegal, by reference to specific provisions of Union law or national law. Third, orders must include sufficient information to enable online intermediaries to identify and locate the illegal content. Fourth, the order must contain information about redress mechanisms available to the platform and impacted users. In addition, the territorial reach of such orders must be limited to what is strictly necessary to achieve its aim.[59]

The DSA also requires platforms to notify the user of such orders that they are received and implemented. The notice must be provided at the latest when the platform gives it effect or, where applicable, at the time specified by the order. Such information provided to the user must include a statement of reasons, the existing possibilities for redress, and a description of the territorial scope of the order.[60] However, the DSA states that the

[57] Article 3, paragraph 1, point (h) of the DSA.
[58] Article 9, paragraph 1 of the DSA.
[59] Article 9, paragraph 2 of the DSA.
[60] Article 9, paragraph 5 of the DSA.

aforementioned conditions and requirements are without prejudice to national civil and criminal procedural law.[61]

The DSA explicitly states that there is no general obligation for the platforms to monitor the information that they transmit or store. They do not need to actively seek facts or circumstances indicating illegal activity.[62] However, platforms may carry out content moderation activities aimed at detecting, identifying, and addressing illegal content.[63]

X's and Meta's required transparency reports under the DSA[64] reveal that state-ordered content removal under the DSA is rare, compared both to their overall user bases and the number of user-flagged illegal content reports they receive under Article 16 of the DSA.[65] This can be seen in the following table.

Platform	Period	Number of Average Active Users in the EU (in millions)	Number of Removal Orders Received (Art. 9)	Number of Notices of Illegal Content Received (Art. 16)
X[66]	2023-10-21 to 2024-03-31	109.2	13 (from three Member States)[67]	238,108

[61] Article 9, paragraph 6 of the DSA.
[62] Article 8 of the DSA.
[63] Article 3, paragraph 1, point (t) of the DSA.
[64] Articles 15, 24, and 42 of the DSA.
[65] Under Article 16 of the DSA, users can report content they believe is illegal.
[66] *DSA Transparency Report – April 2024*, X (Sept. 8, 2024) https://transparency.x.com/dsa-transparency-report.html.
[67] X notes that it has omitted countries and violation types with no legal requests.

Facebook[68]	October 1, 2023, to March 31, 2024	260.7	2,089[69] (1,562 from Germany)	601,863

4. Conclusion

Based on the above analysis of the provisions of the DSA and the operation of its provisions relating to content moderation, it is fair to say that the claims of the Pakistan government officials are false as they pertain to the EU.

[68] *Regulation (EU) 2022/2065 Digital Services Act Transparency Report for Facebook*, *dated April 26, 2024 (updated June 13, 2024)*, Meta (Oct. 5, 2024), pages 2-4, 8-9, and 25, https://transparency.meta.com/sr/dsa-transparency-report-apr2024-facebook.

[69] Meta states that the number of orders refers to "Member States' authority orders to act against illegal content, including under Article 9 of the DSA."

How Governments Use Technology to Regulate Content

Natalie Phillips[70]

Introduction

The government of Pakistan is following the lead of other states that have embraced "digital authoritarianism" by harnessing technological tools to oversee the online activities of its citizens. While these tools have been leveraged by repressive governments to exert control, they have been prohibited in other nations to safeguard privacy, freedom of expression and other fundamental human rights.

This memorandum explores the various technological instruments available to governments for regulating online content and assesses how different countries have implemented these tools to manage and moderate digital discourse within their borders. This memo is divided into three parts. Part One will discuss the tech tools currently available to moderate and censor internet content. Part Two will show examples of how different government models, one authoritarian and one democratic, use tech tools for internet moderation. Finally, Part Three will cover how Pakistan is using these tech tools.

[70] J.D. Candidate, Class of 2026, University of Colorado Law School.

1. Tech Tools to Regulate Content

1.1. Firewalls

A firewall is designed to prevent unauthorized content from entering a network system.[71] The term has its origins in the physical walls found in some buildings to prevent fires from spreading. Like a physical wall, an electronic firewall creates a barrier to protect private networks from outside threats.

Many corporations use firewalls to reduce the risk of cyberattacks and to prevent employee access to non-work-related internet content. However, governments have reconfigured firewalls not for the purposes of keeping hackers out, but to block undesirable content within their network. In the case of repressive states, this includes foreign or domestic content critical of the ruling government (among other things).

1.2. Packet Filtering and Deep Packet Inspection (DPI)

Packet filtering is a methodology employed by firewalls to enforce network rules and restrictions. An internet "packet" consists of three components: a header, the payload, and a trailer.[72] The header contains instructions on delivering the data, such as the source and destination network address.[73] The payload carries the actual user data while the trailer is responsible for checking for errors.[74] "Packet filtering" inspects packet headers for IP

---

[71] https://www.cisco.com/site/us/en/learn/topics/security/what-is-a-firewall.htmlhttps://www.cisco.com/site/us/en/learn/topics/security/what-is-a-firewall.html

[72] https://www.techtarget.com/searchnetworking/definition/packet#:~:text=A%20network%20packet%20is%20a,message%20gets%20to%20its%20destination.

[73] https://www.techtarget.com/searchnetworking/definition/packet#:~:text=A%20network%20packet%20is%20a,message%20gets%20to%20its%20destination.

[74] *Id.*

addresses and other protocols, allowing or blocking traffic according to network rules.[75]

1.1.1. Packet Filtering and Virtual Private Networks (VPN):

A common method to circumvent a firewall's packet inspection is through the use of a virtual private network, or VPN. VPNs work by encrypting a user's IP address, allowing access to content usually restricted in their network.[76] Because the IP address is disguised to match the allowable IP address, it will bypass packet filtering protocols and enter the network undetected.

1.1.2. Deep Packet Inspection

Deep Packet Inspection (DPI) is an advanced variation of packet filtering that examines the entire packet content. DPI can be used for beneficial purposes, such as ensuring the quality of video calls during peak internet usage.[77] However, its capabilities can also be harnessed for more restrictive measures, such as blocking.[78]

DPI functions like a checkpoint for data packets. When a packet arrives, DPI will examine both the header and contents, making determinations about whether to allow or block data entry based on predefined network rules.[79]

---

[75] https://www.techtarget.com/searchnetworking/definition/deep-packet-inspection-DPI#:~:text=Deep%20packet%20inspection%20(DPI)%20is,only%20packet%20headers%2C%20cannot%20detect.

[76] https://www.fortinet.com/resources/cyberglossary/how-does-vpn-work

[77] https://www.digitalguardian.com/blog/what-deep-packet-inspection-how-it-works-use-cases-dpi-and-more#:~:text=There%20are%20several%20uses%20for,detection%20systems%20cannot%20adequately%20detect.

[78] *Id.*

[79] https://www.techtarget.com/searchnetworking/definition/deep-packet-inspection-DPI#:~:text=Deep%20packet%20inspection%20(DPI)%20is,only%20packet%20headers%2C%20cannot%20detect.

DPI systems use one of three main techniques for detection that vary in scope. The first is pattern matching, a method limited to known threats that have signatures in the database. It cannot detect new or modified attacks that don't match existing signatures.[80] The second is protocol anomaly, a procedure capable of identifying unknown threats by recognizing unusual behavior in traffic that deviates from standard protocols.[81] This is known as the "default deny" approach, where the network will deny entry to *any* data packet that does not match the signature database, not just those previously identified as threats.[82] Lastly, intrusion prevention systems (IPS) combines elements of both approaches by detecting known threats via signatures and potentially unknown threats through protocol anomalies or behavioral rules.[83]

All use the same detection methodology of examining network traffic to identify security threats based on network rules and rely on regular protocol updates. These methods are only effective against data categorized as "threats" or potential threats by network rules, even though such content may be completely harmless or protected speech.[84]

1.1.3. DPI and VPN:

Unlike traditional packet filtering, DPI is capable of identifying VPN usage. An administrator can create network protocols that detect VPN traffic based on payload signatures and then block or throttle that traffic accordingly.[85] This blocks any attempts to bypass content policies that would otherwise evade the firewall's packet filtering through VPNs. With

---

[80] *Id.*

[81] *Id.*

[82] *Id.*

[83] *Id.*

[84] *Id.*

[85] https://cleanbrowsing.org/help/docs/strategies-to-block-vpn-connections-on-a-network/#:~:text=Deep%20Packet%20Inspection%20(DPI)&text=You%20can%20create%20rules%20to,or%20throttle%20that%20traffic%20accordingly.

DPI, a network administrator or government can restrict all the payload content from specific websites or applications (like X).[86]

1.3. Geo-Blocking/Geo-Locking

Geo-blocking is a method of restricting access to online content based on the user's geographic location.[87] Like packet filtering, it identifies a user's IP address and blocks content access based on the location of the device.[88] Geo-blocking is a key tool that technology companies use to comply with government censorship demands within a particular jurisdiction.[89]

1.1.4. How Geo-Blocking Works:

For example, internet content such as YouTube videos may only be allowed in certain countries or restricted by others. Therefore, those in a restricted country are unable to access certain content because their IP addresses are geographically blocked from viewing that content based on their device's location. VPNs are capable of circumventing geo-blocking protocols through disguising a user's IP address as one from a different region or nation.[90]

1.1.5. How Geo-Blocking is Used:

Continuing with the YouTube example, the company can block a video in Pakistan at a request from the government while it remains accessible everywhere else. This allows for tech companies to operate their websites in compliance with local law without burdening the free expression and informational rights of persons beyond that jurisdiction. However, it is

---

[86] *Id.*

[87] https://www.atinternet.com/en/glossary/geoblocking-geographical-blocking/#:~:text=Geoblocking%20is%20a%20practice%20of,authorisation%20or%20denial%20of%20access.

[88] *Id.*

[89] *Id.*

[90] https://www.fortinet.com/resources/cyberglossary/how-does-vpn-work

important to note that companies, not governments, are responsible for taking the action to implement geo-blocking.[91]

1.4. Border Gateway Protocol (BGP)

Border Gateway Protocol (BGP) is a set of rules that determines the way information is routed across networks.[92] BGP is handled by routers that transfer data from one router to its intended destination. It is responsible for finding all available pathways that data could travel and choosing the best route to the router.[93]

Pathways identified by BGP connect autonomous systems controlled by private companies or government agencies to interconnect them, giving rise to the Internet.[94] Through their power over these networks, BGP can be purposefully manipulated by governments and corporations to trigger router errors and force Internet blackouts, effectively disconnecting Internet access to the public.[95] Such routing disruptions can bring all online activities to a halt.

1.5. Hash Matching

Hash matching is a two-fold process to identify and block known illegal content. The process begins by turning content stored or shared by users into "hashes" that are then compared to a database of known illegal content.[96] If a "match" is found between the user content and the illegal

---

[91] https://nordvpn.com/blog/what-is-geoblocking/#:~:text=Geo%2Dblocking%20means%20restricting%20or,abundance%20of%20foreign%20websites%20inaccessible.

[92] https://www.whitehouse.gov/oncd/briefing-room/2024/09/03/press-release-white-house-office-of-the-national-cyber-director-releases-roadmap-to-enhance-internet-routing-security/

[93] https://www.cloudflare.com/learning/security/glossary/what-is-bgp/

[94] *Id.*

[95] https://www.securityweek.com/bgp-flaw-can-be-exploited-for-prolonged-internet-outages/

[96] https://proton.me/blog/online-safety-act-hash-scanning

database, then that content is restricted or blocked.[97] Hashing functions belong in one of two categories: cryptographic or perceptual.[98] Cryptographic will identify exact matches while perceptual will find matches it determines are very similar.[99] Perceptual hashing is most often used for detecting known illegal visual media.[100] This technology is most effective for content that is illegal everywhere, such as child sexual abuse material (CSAM).[101] However, it can be trained to detect content such as images of a protest or anti-government propaganda for the purposes of restricting online speech.[102]

1.6. Artificial Intelligence

Artificial intelligence (AI) models can be trained to moderate or censor content. By training large language models (LLMs) to pick out certain language like "freedom" or "protest", AI creates an "allowlist/blocklist" that filters specific language in and out of a national network.[103] These AI tools serve as powerful censors, capable of identifying and removing posts that criticize authoritarian regimes. This makes AI particularly challenging because it pairs the possibility of very effective content moderation with vast surveillance possibilities. Automated AI systems designed to detect critical speech can stifle protected expression and reinforce biases present

---

[97] *Id.*

[98] https://www.ofcom.org.uk/online-safety/safety-technology/overview-of-perceptual-hashing-technology/#:~:text=Comparing%20such%20a%20hash%20with,Technology%20(PDF%2C%202.1%20MB)

[99] *Id.*

[100] *Id.*

[101] chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/other/perceptual-hashing-technology.pdf?v=328806

[102] *Id.*

[103] https://www.lesswrong.com/posts/oqvsR2LmHWamyKDcj/large-language-models-will-be-great-for-censorship

in their training data, often disproportionately impacting minority and marginalized communities.

1.1.6. How AI is Used:

By making online surveillance easier, faster, cheaper, and more effective, AI has amplified digital repression across the globe.[104] AI-enabled surveillance tools raise due process concerns, threatening the requirement of individualized suspicion by treating everyone as a potential wrongdoer. These surveillance systems scour social media and other websites to identify dissenters. Such tools may be paired with facial scans to track down online protesters.[105] Governments in autocratic countries with histories of human rights violations are more prone to abuse of AI surveillance than liberal democracies.[106]

2. Government Use of Tech Tools to Regulate Content

2.1. China & Digital Authoritarianism

China is the pioneer of the digital authoritarianism model that Pakistan seems to be embracing. China's "Great Firewall" was the first internet censorship system deployed on a national scale.[107] In its early years, the firewall only blocked anti-Communist Party websites within the country.[108] However, China began to implement additional methods to

---

[104] https://www.lesswrong.com/posts/oqvsR2LmHWamyKDcj/large-language-models-will-be-great-for-censorship

[105] https://www.npr.org/2021/01/05/953515627/facial-recognition-and-beyond-journalist-ventures-inside-chinas-surveillance-sta

[106] https://carnegieendowment.org/research/2019/09/the-global-expansion-of-ai-surveillance?lang=en

[107] https://freedomhouse.org/country/china/freedom-world/2024

[108] *Id.*

restrict public internet use, including DPI in its firewall and geo-blocking systems.[109] Today, government censors use AI to scrub the internet for criticism of the government, promptly flagging posts, taking them down, and in some cases, locating the poster to face jail time.[110] AI chatbots produced by China-based companies have refused to engage with user prompts on issues sensitive to the government, such as the autonomy of Taiwan.[111] Further, AI has been deployed to support security cameras with facial recognition technology to track down those promoting pro-democracy sentiments online.[112]

The Cyberspace Administration of China (CAC) is China's government regulatory body tasked with integrating the country's censorship goals into content recommendation algorithms that tech companies operating within China are required to follow.[113] All social media and AI applications available to the Chinese public must adhere to strict regulations and exclude content deemed illegal or undesirable.[114] Major foreign technology companies like Google were forced to pull out of China for refusing to comply with government demands to filter searches.[115]

2.2. European Union (EU) & Freedom of Expression

In 2022, the EU passed the Digital Services Act (DSA) to mitigate harmful online activities. As explained in detail in a separate memo in this collection, the DSA protects consumers fundamental rights and facilitates technological innovation by setting clear rules for both users and

---

[109] https://www.chinaiplawupdate.com/2022/11/forbidden-china-geo-blocking-americans-from-accessing-supreme-peoples-court-website-and-published-decisions/

[110] *Id.*

[111] https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence

[112] https://www.npr.org/2021/01/05/953515627/facial-recognition-and-beyond-journalist-ventures-inside-chinas-surveillance-sta

[113] *Id.*

[114] *Id.*

[115] https://www.technologyreview.com/2018/12/19/138307/how-google-took-on-china-and-lost/

platforms. This transparency in content moderation has fostered a safe and regulated internet platform for citizens to express views while the government still retains oversight of the systemic risks of disinformation.

In March 2024, the EU passed the Artificial Intelligence Act, banning AI products that are deemed to present an "unacceptable risk" to freedom of expression and public welfare.[116] Research conducted by the EU found that government-imposed restrictions on access to prominent websites and social media platforms were unproductive methods of addressing foreign interference and promoting online safety.[117] The results of the implementation of this law have yet to be seen.

3. Pakistan's Emerging Use of Tech Tools to Regulate Content

The government of Pakistan is actively implementing measures to censor and control internet access, drawing inspiration from models used in countries like China and the UAE. These measures violate the free expression rights of people in Pakistan.

While there has been no official acknowledgment of a comprehensive firewall, reports suggest that the government is developing a web management system capable of filtering online content, blocking specific apps, and monitoring traffic.[118] This system is designed to identify and restrict VPN usage through deep packet inspection and manipulation of the border gateway protocol. The recent disruptions experienced by users, particularly in accessing media on platforms like WhatsApp, hints at the extent of this filtering. Given that users who secured their connection with

---

[116] https://datamatters.sidley.com/2024/03/21/eu-formally-adopts-worlds-first-ai-law/#:~:text=On%20March%2013%2C%202024%2C%20the,of%20legislation%20for%20the%20EU.

[117] https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law

[118] https://www.dawn.com/news/1853742/whats-happening-with-the-internet-in-pakistan

a VPN did not experience issues with such platforms, we can surmise that deliberate throttling rather than technical failures were the cause of these disruptions.[119]

Currently, internet speeds in Pakistan are notably slow, with frequent disruptions impacting economic growth and innovation. The Overseas Investors Chamber of Commerce and Industry (OICCI) has raised concerns that these interruptions hinder foreign direct investment, which is crucial for the country's economic revival.[120] Moreover, freelancers, who contribute over one billion dollars annually to the economy, are disproportionately affected by these measures, facing challenges in their work due to unreliable internet access.[121] As the government seeks to enforce its censorship policies, the implications for business operations could lead to a significant economic downturn, with estimates suggesting potential losses of around $300 million in monthly revenue.[122]

In addition to the increasingly sophisticated monitoring and filtering systems being put in place, the government has established regulatory frameworks that compel individuals and businesses to register their VPNs. While these measures are framed as necessary for cybersecurity, they significantly infringe on citizens' rights to privacy and freedom of expression, undermining the very principles the government is constitutionally obligated to protect.

---

[119] https://www.geo.tv/latest/557762-bandwidth-bandits-internet-regulation-riddle-strikes-vpns-in-pakistan

[120] https://www.geo.tv/latest/557762-bandwidth-bandits-internet-regulation-riddle-strikes-vpns-in-pakistan

[121] https://www.geo.tv/latest/557762-bandwidth-bandits-internet-regulation-riddle-strikes-vpns-in-pakistan

[122] https://www.geo.tv/latest/557762-bandwidth-bandits-internet-regulation-riddle-strikes-vpns-in-pakistan

4. Conclusion

Global internet freedom is currently under threat. Governments employ a range of technological tools to censor online content. Firewalls monitor and regulate network traffic according to government-determined security protocols, which may include blocking pages containing specific keywords or phrases, such as “rights.” China's "Great Firewall" serves as a notable example of national-level firewall implementation. However, not even China is immune to challenges from civilians using VPNs to circumvent its restrictions. In response, governments like China are increasingly investigating methods to block or track VPN usage, promising high fines and jail time for those caught using VPNs. Techniques like packet filtering and deep packet inspection allowing governments to restrict access to certain websites and monitor citizens attempting to connect to unauthorized IP addresses.

While such advanced technologies were once the domain of powerful countries like China, the tools of digital authoritarianism have become accessible to smaller nations as well. The emergence of new technologies to assist in government content moderation, such as artificial intelligence, large language models, and hash-matching techniques, has only made online censorship more accessible to smaller nations and easier for global superpowers. As Pakistan navigates these challenges, the balance between security and civil liberties remains a contentious issue, raising concerns about the potential for a complete erosion of internet freedom in the country.