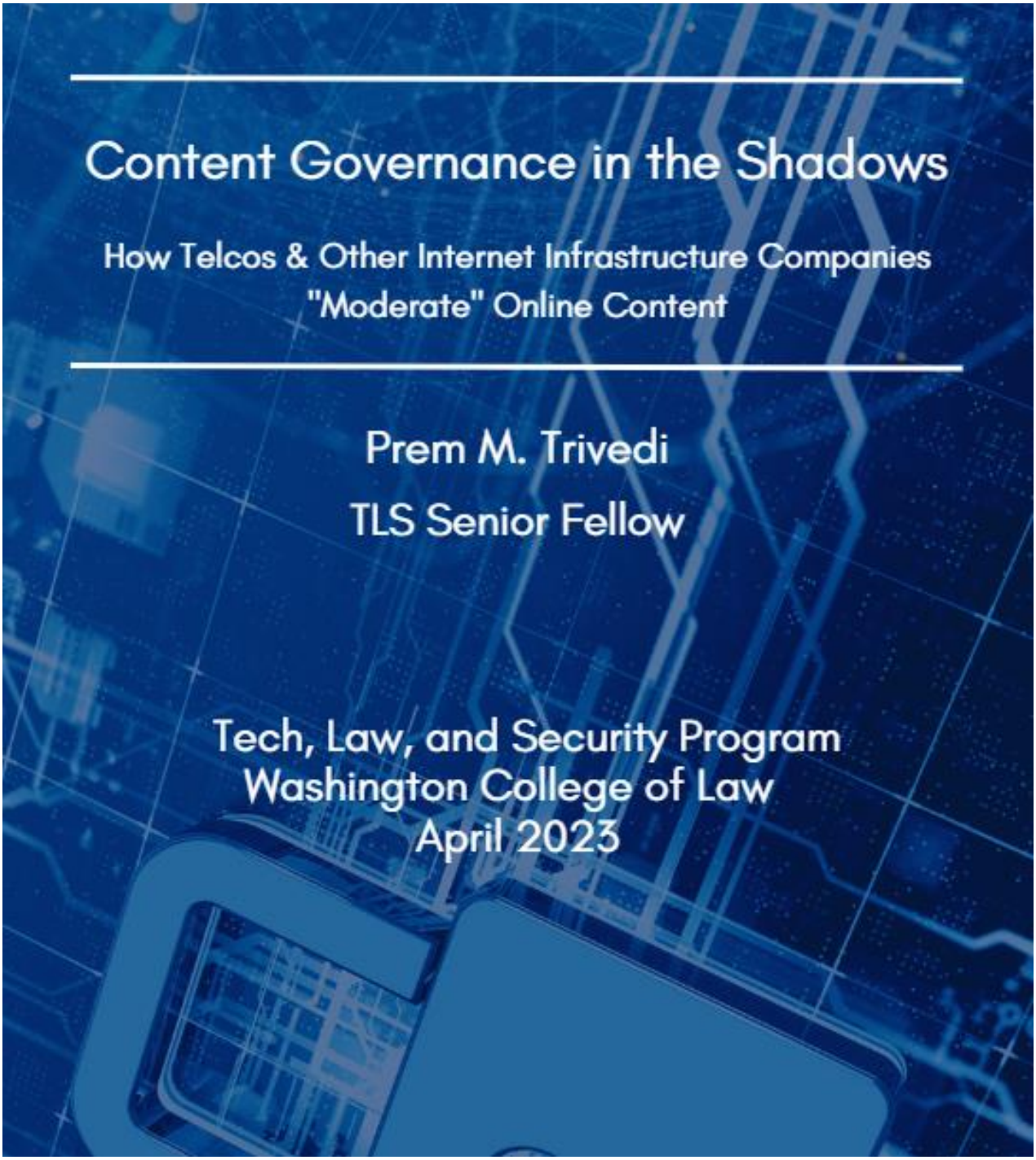
2023

Content Governance in the Shadows: How Telcos & Other Internet Infrastructure Companies "Moderate" Online Content

Prem M. Trivedi
American University Washington College of Law

AMERICAN UNIVERSITY
WASHINGTON COLLEGE OF LAW

TECH, LAW & SECURITY PROGRAM



Content Governance in the Shadows

How Telcos & Other Internet Infrastructure Companies "Moderate" Online Content

Prem M. Trivedi

TLS Senior Fellow

Tech, Law, and Security Program
Washington College of Law
April 2023

ADL CENTER FOR TECHNOLOGY & SOCIETY

The Tech, Law & Security Program (TLS)

TLS is a rapidly expanding initiative at the American University Washington College of Law that tackles the challenges and opportunities posed by emerging technologies — offering innovative solutions, engaging our students, and training the leaders of tomorrow.

This paper is part of a series of scholarship and commentary from TLS's Harmful Content Online project. The project was launched to (1) examine how actors throughout the internet ecosystem — beyond the major social media platforms — engage in content moderation and (2) help craft relevant, rights-protective regulations and standards. For more information about current TLS initiatives, please visit our website at: https://www.wcl.american.edu/impact/initiatives-programs/techlaw/. This research project has received funding from the Center for Technology and Society at the Anti-Defamation League (ADL).

Center for Technology & Society

Launched in 2017, ADL's Center for Technology and Society (CTS) is a research-driven advocacy center that works to end the proliferation of antisemitism and all forms of hate and harassment online. Our team partners with industry, civil society, government and targeted communities to expose these harms, hold tech companies accountable and fight for just, equitable online spaces.

Anti-Defamation League

ADL is the leading anti-hate organization in the world. Founded in 1913, its timeless mission is "to stop the defamation of the Jewish people and to secure justice and fair treatment to all." Today, ADL continues to fight all forms of antisemitism and bias, using innovation and partnerships to drive impact. A global leader in combating antisemitism, countering extremism and battling bigotry wherever and whenever it happens, ADL works to protect democracy and ensure a just and inclusive society for all.

Support

This report was made possible with generous support from the Center for Technology & Society at the Anti-Defamation League. The content of this report reflects the views of its author alone. TLS maintains strict intellectual independence and sole editorial discretion and control over its intellectual property, ideas, projects, publications, events, and other research activities.

Contents

Acknowledgments

Sarah Dean and Dinesh Napal provided invaluable research assistance and insightful comments. Gary Corn and Laura Draper of the Tech, Law & Security Program provided constructive inputs on this paper throughout its evolution. I am indebted to TLS colleagues Alex Joel and Paul Rosenzweig, to John Leitner, and to my family for their thoughtful reviews. I also owe a great deal to Jim Dempsey and Sharon Bradford Franklin, two ideal guides at the start of my career in technology law and policy. Any errors or omissions are mine alone.

This paper is dedicated to my parents, my wife, and our two children.

I. Executive Summary

This paper addresses two significant policy challenges in companies' online content governance (usually called "content moderation") activities: (1) the lack of information we have about most internet infrastructure providers' actual practices and (2) the lack of a standardized decision-making framework that different entities in the online content ecosystem can apply to their respective contexts. These are related problems, and the second problem is in part a function of the first. Section II of this paper discusses these challenges in greater depth.

Section III explains why I think that the term content *moderation* overly narrows our focus on the application layer of the internet and obscures the need to focus on important content *governance* decisions that shape how people access the Internet and how content is hosted, routed, and delivered. Section IV outlines the rationale for a global case study on telecommunications companies that operate at the internet's "access layer." In short, telcos around the world engage in systemic content governance activities at a scale that poses significant consequences for human rights, but their decisions at the individual subscriber level are far less understood than their efforts to restrict internet access within certain geographic boundaries or to block content. I argue that the findings from a diverse set of telecommunications companies should help distill broader lessons relevant to all non-application-layer infrastructure companies. These lessons might be particularly valuable for companies in the early stages of establishing policy and compliance programs. Section V details the methodology for this case study, in which a TLS research team surveyed forty-two telecommunications companies' subscriber-level agreements as a window into telcos' content governance rules. The list of companies surveyed appears in the Appendix.

Section VI presents key research findings about how telecommunications companies (1) define unlawful content, (2) define lawful but otherwise objectionable content and the standard for acting against it, and (3) define their enforcement options when subscribers violate terms of service or acceptable use policies. Section VII makes recommendations about how telecommunications companies can improve meaningful transparency into their content governance strategies and practices. Section VIII makes recommendations for how all non-application layer internet infrastructure companies can experiment with substantive content governance principles. It discusses the importance of the Global Network Initiative Principles, the Santa Clara Principles, and the Manila Principles, and notes that these foundational and often detailed guidance documents nevertheless do not readily lend themselves to developing corporate decision-making frameworks for hard content governance problems. I conclude Section VIII by presenting a sample decision-making framework that large and small companies could improve upon and tailor to their own contexts. Section IX summarizes the paper's broad lessons and looks ahead.

II. Introduction: Examining the Online Content Governance Decisions of Non-Application Infrastructure Companies

How do telecommunications companies that provide internet access, like Verizon and AT&T in the United States, Telefónica in Spain, and Reliance and Airtel in India, shape the content that people consume online? What about other internet infrastructure companies that provide Content Delivery Network services and protect clients from Distributed Denial of Service (DDoS) attacks? One such company, Cloudflare, was in the news last fall for its controversial decision to block, and thus take offline, right-wing extremist site Kiwi Farms. Cloudflare's decision followed sustained public pressure from activist Clara Sorrenti, who was the subject of a barrage of hateful, anti-transgender speech emanating from users of Kiwi Farms. Cloudflare CEO Matthew Prince insisted that this was an "extraordinary" and "dangerous" decision he would have preferred not to make, explaining that the company's hand was forced because "the process [of reviewing Cloudflare's warnings to law enforcement about possible illegal activity] is moving more slowly than the escalating risk."[1] Prince might have added to his written rationale that there is little available in the way of standardized and specific content governance principles for Cloudflare to apply. This lack of standardization, compounded by the lack of information we have about most internet infrastructure providers' actual content governance activities, is the focus of this paper.

The term "content moderation" typically makes us think of social media companies or messaging applications, for good reason. Most of us engaged in routine online activities are consuming content, creating content, or engaging in transactions via applications and websites. While performing tasks like messaging, interacting on social media, or banking online, we have little occasion to consider or discover the consequential content governance activities taking place behind the scenes. As a result of this understandable bias toward highly visible sites of content moderation, much ink has been spilled on the issue of how social media companies should better define harmful online content, structure their efforts to moderate it, and provide transparency into how they make decisions.

The preoccupation *du jour* is on the chaos Elon Musk has unleashed at Twitter, but debates on content moderation have of course also involved other typical objects of scrutiny like

[1] Matthew Prince, *Blocking Kiwi Farms*, Cloudflare (Sep. 3, 2022), https://blog.cloudflare.com/kiwifarms-blocked/ (also noting that "[W]e need a mechanism when there is an emergency threat to human life for infrastructure providers to work expediently with legal authorities in order to ensure the decisions we make are grounded in due process. Unfortunately, that mechanism does not exist and so we are making this uncomfortable emergency decision alone."); *see also* Casey Newton, *How Cloudflare got Kiwi Farms Wrong,* The Verge (Sep. 6, 2022), https://www.theverge.com/2022/9/6/23339889/cloudflare-kiwi-farms-content-moderation-ddos (describing Cloudflare's decisions as an act of "content moderation" and explaining how Cloudflare responded to Clara Sorrenti's demands that it block Kiwi Farms).

Facebook, Instagram, YouTube, and TikTok.[2] The application layer of the internet is where huge volumes of content are created, shared, and algorithmically promoted or demoted. Governance of content at the application layer of the internet's network stack is hugely important.[3] But a singular focus on content *moderation* means that our field of vision often excludes a range of other important internet infrastructure players making significant content *governance* decisions.[4] In the next section, I explain why I think a broader conceptual focus on content governance is more useful and precise than a focus on content moderation.

Stepping back, there are several reasons that we should care about the ways in which telecommunications companies and other less-scrutinized internet infrastructure players engage in content governance. *First*, this is not a theoretical exercise; it already is an urgent imperative. Non-application layer infrastructure companies already receive, and will only receive more, requests from governments and non-governmental actors to take down or otherwise restrict access to content that someone deems harmful. Much like social media and messaging companies, these companies must confront the same tidal waves of hateful and often dangerous online speech with which society at large is grappling. Religiously and racially motivated hate speech, harassment based on sexual orientation, bullying of minors, and political disinformation campaigns are just a few examples of speech that acutely disrupt the balance of healthy societies. There is an urgent imperative to create safer online spaces[5] while protecting free expression, intellectual exploration, and privacy in ways that further what Julie Cohen calls the "dynamic, emergent subjectivity" of the individual.[6]

Second, we should care precisely because too few researchers are shining a light on the spaces where non-social media internet infrastructure players are making important decisions. It is not these companies' fault that they are operating in the shadows; few policymakers and civil society watchdogs are forcing them to do otherwise. Given the harsh glare of public scrutiny on social media companies' content governance practices, the business incentive for most non-application layer companies is to keep a low profile. Let me be quite clear: I am not suggesting that telcos should "do more" on content governance. I favor applying the principle of least intervention more rigorously as an entity moves farther away from the application layer and toward the access layer.[7] But my main objective here is to confront the challenges

---

[2] *See, e.g.*, Terry Gross, *Chaos reigns at Twitter as Musk manages 'by whims'*, NPR (Dec. 8, 2022), https://www.npr.org/2022/12/08/1141568738/chaos-reigns-at-twitter-as-musk-manages-by-whims.

[3] *See* Steve Bellovin, *How the Internet Works*, slide 23, (May 18, 2016), https://www.cs.columbia.edu/~smb/talks/internet-intro-cls.pdf.

[4] *See generally* LAURA DENARDIS, THE GLOBAL WAR FOR INTERNET GOVERNANCE, CHAPTER ONE (2014).

[5] *See, e.g.*, *Readout of White House Listening Session on Tech Platform Accountability* (Sep. 8, 2022), https://www.whitehouse.gov/briefing-room/statements-releases/2022/09/08/readout-of-white-house-listening-session-on-tech-platform-accountability/.

[6] Julie Cohen, *What Privacy is For*, 126 HARV. LAW REV. 1904, 1905-06 (2013).

[7] *See, e.g.*, Jack Balkin, *Free Speech is a Triangle,* 118 COLUM. L. REV 2011 at 2038-40 (2018), https://columbialawreview.org/wp-content/uploads/2018/11/Balkin-FREE_SPEECH_IS_A_TRIANGLE.pdf; Ben Thompson, *A Framework for Moderation*, Stratechery (Aug. 7, 2019), https://stratechery.com/2019/a-framework-for-moderation/ ("It makes sense to think

posed by our poor understanding of how internet infrastructure companies *actually* engage in content governance.

Staying in the shadows imposes high costs not just on societies, but on internet infrastructure companies themselves. Companies like Cloudflare already face reputational risks and are thinking through, at least on a case-by-case basis, how to shape their approach to content governance in a way that appropriately balances security and democratic values. Both digital rights and online security advocates should more closely scrutinize how internet infrastructure companies make content governance decisions and identify areas where there are accountability shortfalls and risks to good governance.

The Tech, Law, & Security Program at the American University Washington College of Law (TLS) has worked to illuminate the broader internet infrastructure ecosystem that shapes how content is created, routed, delivered, and consumed online. The hope is that doing so will provide helpful guidance to policymakers, companies, and civil society groups experimenting with establishing effective governance standards. In 2021, TLS published a foundational paper that maps out the rich ecosystem of internet infrastructure entities involved in the "online information ecosystem."[8] These actors include companies that provide internet access, companies that route traffic, content delivery networks and web hosts, and entities that provide or facilitate functions like browsing and financial transactions.[9] Many of these players, far less visible to internet users than the applications where we interact and consume content, are nevertheless making consequential governance decisions about content.

This paper focuses on the telecommunications companies operating at the "access" layer of the internet. These companies, which include landline and mobile Internet Service Providers (ISPs), serve as gatekeepers for their subscribers' online content consumption and creation. In a nutshell, telecommunications companies determine who can and cannot get online. As discussed in Section III below, they engage in content governance in three broad ways: they can completely shut down internet access within defined geographies, block their subscribers' access to certain sites or apps, and suspend or ban individual subscribers that they deem to violate laws or their policies.

A simple example helps to illustrate the role of telecommunications companies and other key players in the online content ecosystem. If Anand wants to post a comment on a social media service, he first must get online. To do that, he relies upon his ISP, which could be a mobile

about these positions of the stack very differently: the top of the stack is about broadcasting — reaching as many people as possible — and while you may have the right to say anything you want, there is no right to be heard. Internet service providers, though, are about access — having the opportunity to speak or hear in the first place. In other words, the further down the stack, the more legality should be the sole criteria for moderation; the further up the more discretion and even responsibility there should be for content").

[8] Jenna Ruddock & Justin Sherman, *Widening the Lens on Content Moderation*, Joint PIJIP/TLS Research Paper Series (July 2021), https://digitalcommons.wcl.american.edu/research/69/.

[9] *Id.*

or landline provider, depending on the device and network he chooses to use. That telecommunications company serves as his gateway, or on-ramp, to the public internet. Anand opens his social media app on his phone, types his content, and posts it. While he is doing all of this, his device and the social media provider's servers are exchanging information. Anand's friend Becky follows him on the same social media app. Anand's post — and the other content that appears in her feed — is broken down into component "packets" of data that are directed via routers across physical cables and wires that eventually make their way to the wires that run into Becky's home and to her wireless router. When she unlocks her mobile phone, her internet service provider already has connected her device to the public internet. She opens her social media app and sees the new post from Anand in her feed. Those component packets into which Anand's post were broken have now been reassembled on Becky's device so that when she clicks on his post she finds an intelligible message.

Without belaboring the details, the key point is that the near-instantaneous social exchange between Anand and Becky did not just *happen*. A whole host of internet infrastructure companies, including the ISPs and companies that route traffic, work behind the scenes to facilitate a commonplace transaction that occurs millions, if not billions, of times a day. This largely hidden world beneath the application layer of the internet is a place where important decisions about content governance are made.

This paper focuses on one piece of that puzzle: what do telecommunications companies around the world say about how they act against unlawful or otherwise objectionable content, and what do their claims of authority suggest about how they apply their own rules? This paper's aim is to distill lessons from that case study into the contours of a more general decision-making framework for companies trying to take a structured approach to content governance. Ideally, such a framework should provide people with a common vocabulary to rise above conversations about seemingly incompatible values or interests and discuss what to *do*. Good frameworks are a precondition for good negotiations, which in turn are a precondition for outcomes that improve governance (whether those outcomes are expressed in legislation, regulation, codes of conduct, internal corporate rules, or even computer code.)

III. Focus on "Content Governance," not just "Content Moderation"

Before discussing telecommunications companies in greater detail, it is worth parsing the terms "content moderation" and "content governance," which can produce linguistic and conceptual confusion. "Moderation" tends to conjure up visions of social media companies making individual determinations about whether posts or other user-generated uploads violate laws, company policies, or other norms. One imagines a team of YouTube or Instagram employees deciding whether to take down a video glorifying political violence or apply a warning label about potential misinformation to a user's post about COVID-19. But these individual, human-directed decisions actually comprise a minority of application-layer

content moderation. Analogizing content moderation to judges deciding individual cases fails to describe the reality of what Evelyn Douek calls the "systemic" nature of social media platforms' content moderation, a process in which huge bureaucracies apply rules and standards, usually via automation, to millions of decisions to take down, leave up, label, or otherwise modify content.[10]

What advocates of free expression and safer online spaces alike are concerned with is something broader than moderation at the individual level. They seek to better understand how governments and private entities craft rules that govern the consumption, creation, and accessibility of online content at a social level. These decisions and the rules that enable them are most appropriately called "content governance," which is the term I will use here.[11] Even though companies like telcos, Cloudflare, or the Apple and Google app stores do not moderate content in the way that social media or messaging companies do, they still engage in all sorts of content *governance* decisions that have important consequences for speech, security, and the nature of online discourse and commerce.

Like social media companies, the broader ecosystem of internet infrastructure companies also makes what Douek calls systemic decisions. Their efforts, and the increasing public and private demands on them to systematically govern content using consistent standards, are a manifestation of what Jack Balkin calls "new-school speech regulation … directed at internet infrastructure."[12] This new-school speech regulation stands in contrast to a traditional, "dyadic" model of speech regulation in which nation states were the primary actors that controlled the speech of individuals, associations, and traditional media players.[13] Today, we operate within what Balkin calls a pluralist model of speech regulation that can be simplified into a "speech triangle."[14] Three groups of actors make up each of the three corners of the triangle: (1) nation-states, states, municipalities, and supranational entities like the European Union; (2) internet infrastructure companies (including but not limited to social media companies, ISPs, web-hosting services, Domain Name System registrars, cybersecurity service providers, and payment systems); and (3) "speakers and legacy media, including mass-media organizations, protesters, civil-society organizations, and trolls."[15]

---

[10] Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 526 (2022), https://harvardlawreview.org/2022/12/content-moderation-as-systems-thinking/.

[11] "Content governance" is hardly an original term, but I believe it is underappreciated and underutilized. *See, e.g.*, *26 recommendations on content governance – a guide for lawmakers, regulators, and company policy makers*, Access Now (Mar. 2020), https://www.accessnow.org/wp-content/uploads/2020/03/Recommendations-On-Content-Governance-digital.pdf.

[12] Jack Balkin, *Free Speech is a Triangle,* 118 COLUM. L. REV 2011 (2018), https://columbialawreview.org/wp-content/uploads/2018/11/Balkin-FREE_SPEECH_IS_A_TRIANGLE.pdf.

[13] *Id. at* 2013.

[14] *Id. at* 2014-2015.

[15] *Id.* at 2015.

Each of these three groups tries to exert influence over the others. Importantly, "the internet infrastructure regulates private speakers and legacy media through techniques of *private governance*."[16] Both nation-states and private speakers try to influence internet infrastructure players through a mixture of compulsion (legislation, requests for action through legal process) and persuasion (public campaigns like Clara Sorrenti's, private requests, *etc.*).[17] This paper takes up the question of how infrastructure providers approach the task of private governance. **How do other internet infrastructure companies, scrutinized far less than social media and messaging services, define what is "objectionable" content, identify it, and then make decisions about permitting, prohibiting, or otherwise affecting access to that material?**

This question could lead in many different directions, and this case study starts at the access layer of the internet. Often described as residing at the base of the internet protocol stack, the telecommunications providers that control people's on-ramps to the internet are in a sense logically farthest from the activity at the application and social layers of the internet. To explain this, imagine that Becky opens her Twitter app and sees Anand's tweets in her feed. She agrees with his views and starts a series of re-tweets to amplify his content. She is engaged at the internet's content layer or, as Jonathan Zittrain has dubbed it, its social layer.[18] The application layer of the internet consists quite clearly of the Twitter app on her phone. But beneath these surface-level activities, a whole host of entities have technically enabled her consumption and creation of Twitter content. The most fundamental and first step is her connection to the internet itself. Telecommunications providers operate and control =this on-ramp.

IV. Why Focus on Telecommunications Companies?

As noted above, telecommunications companies ("telcos") serve as critical on-ramps to the internet and engage at significant scale in the kind of systemic moderation (more appropriately, systemic content governance) that Douek describes. There are three broad ways in which telcos affect the creation or consumption of content. First, they can initiate network shutdowns within a particular geography — that is to say, they can simply "turn off" their provision of service in entire areas, rendering the internet inaccessible within the defined geography. Such practices might be unfamiliar to readers in the United States, but they are quite common in other countries. These internet shutdowns are chiefly effected in

[16] *Id.* at 2015.

[17] *Id.* at 2015-16.

[18] Jonathan L. Zittrain, THE FUTURE OF THE INTERNET AND HOW TO STOP IT at p. 67 (2008) ("At the top is the "application layer," representing the tasks people might want to perform on the network. (Sometimes, above that, we might think of the "content layer," containing actual information exchanged among the network's users, and above that the "social layer," where new behaviors and interactions among people are enabled by the technologies underneath.").

response to a government's invocation of public safety or other law-and-order mandates.[19] A prominent exemplar of this approach is India, which holds the dubious distinction of leading most lists that track numbers of internet shutdowns.[20] Other countries like Turkey also have employed this tool in the name of public safety.[21] Shutdowns are the most heavy-handed form of intervention, since cutting off large geographical areas from internet access has a deeply consequential and large-scale impact on content consumption and generation.

Second, telcos can block or throttle access to particular websites or applications. While not as sweeping as the first category of interventions, decisions to ensure that subsets of subscribers are unable to access certain sites or services are extremely significant. For instance, using the earlier Kiwi Farms example, Verizon or AT&T or Comcast could hypothetically have decided that they would not route any of their subscribers' traffic to Kiwi Farms's site, thus rendering the site inaccessible. A real-world example involves the Indian government's decision to ban TikTok, which was enforced at least in part by orders to Indian ISPs to "filter out Indians' access to TikTok servers."[22]

Third, telcos can suspend or ban subscribers from using their services — that is, they can deny a subscriber use of their "on-ramp" to the internet when they deem that individual to have violated their terms of service or acceptable use policy. Of course, these various acts of moderation could be applied by different companies to different scenarios and might be simultaneously applied in various combinations. A telco has considerable discretion to apply the standards it establishes in its policies. Given a general lack of transparency about how telcos and other internet infrastructure players approach content governance, we know very little about how telcos choose among these various levers. And we know the least about the third category: what happens at the level of individual subscribers? This paper aims to help researchers and practitioners assess how internet infrastructure players like telcos act when measured against their own standards or decision-making frameworks.

Like social media companies, telecommunications companies collectively deal with millions of subscribers, track huge numbers of websites visited and applications used, and may process large volumes of requests from governments to block, filter, or otherwise restrict access to content. In this sense, at least some of their processes for handling issues with objectionable content also have developed into "systemic" content governance. As the preceding discussion about three principal methods of intervention demonstrates, telcos are one of the most

---

[19] *See, e.g.*, Steven Feldstein, *Government Internet Shutdowns Are Changing. How Should Citizens and Democracies Respond?*, Carnegie Endowment for International Peace (Mar. 31, 2022), https://carnegieendowment.org/files/Feldstein_Internet_shutdowns_final.pdf.

[20] *See, e.g.*, *Let the Net Work: Internet Shutdowns in India 2022*, Software Freedom Law Center (Dec. 23, 2022), https://sflc.in/internet-shutdowns-india-2022.

[21] *See* Burak Haylamaz, *Shutting Down the Internet to Shut Down Criticism*, Verfassungsblog (Feb. 17, 2023), https://verfassungsblog.de/turkey-internet-earthquake/.

[22] Justin Sherman, *The problem with India's app bans*, Atlantic Council (Mar. 27, 2023), https://www.atlanticcouncil.org/blogs/southasiasource/the-problem-with-indias-app-bans/.

powerful players in the internet ecosystem, with the ability to affect content dissemination and consumption at a scale and severity that dwarfs even the capabilities of social media companies. For the most part, this tremendous power has *not* been accompanied by commensurate scrutiny on how it is wielded. A recent New America publication by Ranking Digital Rights surveys telcos around the world and highlights their "digital rights deficit," rightly calling attention to their overall track record on free expression.[23]

Ranking Digital Rights notes that telcos' record on free expression remains particularly concerning, in part because "[w]e still know very little about how telco giants process demands from government and private actors to block access to individual IPs and URLs, suspend accounts (for instance, by disabling individual SIM cards), or impose other restrictions."[24] This paper digs deeper into a subset of this question: how do telcos define and address in their legal agreements with subscribers what they deem to be lawful but otherwise objectionable content? This question corresponds most naturally to the third category of interference with content consumption and creation — terminating or suspending a subscriber's internet service.

In sum, I focus on telcos for four key reasons. *First*, telecommunications providers (which encompass both landline Internet Service Providers and mobile network operators) have arguably the broadest powers to restrict access to content. They can block access to websites for all their customers or deny internet access to subscribers whose behavior they deem to violate their terms of use. Particularly when viewed through a global lens, we see that governments frequently target telcos when they want to exert their own content governance powers. A social media company shapes people's content experience on its own application(s); the ambit of a telco's content governance decision could extend to the entire internet. Given this broad ability to shape online content creation and consumption, and the "distance" between the access and content layers of the internet, we might wish and expect to see a principle of least intervention at play.

Second, and *because* of the power of telcos' potential impact, content interventions at the access layer have significant public policy consequences. This power of impact is precisely why government actions to censor online content often target the internet's access layer.

Third, telcos generally operate with broad legal latitude to make decisions about content. In the United States, for example, Section 230 of the Communications Decency Act immunizes ISPs and other "interactive computer services" from liability when they make good-faith

[23] Jessica Dheere, *Missed calls?: It's time telco giants answered for themselves*, Ranking Digital Rights (Dec. 20, 2022), https://rankingdigitalrights.org/tgs22/key-findings/missed-calls-it-s-time-telco-giants-answered-for-themselves.

[24] Jan Rydzak, *Transparency improves on shutdowns, but telcos still weak on free expression*, Ranking Digital Rights (2022), https://rankingdigitalrights.org/tgs22/key-findings/transparency-improves-on-shutdowns-but-telcos-still-weak-on-free-expression.

decisions about restricting access to objectionable content on their services.[25] (My aim here is not to advance any proposals for reforming Section 230, although I second the many thoughtful arguments counseling cautious experimentation and/or opposing an overhaul.[26])

As a policy matter, however, it is valuable to understand how telcos draw lines and define standards within their broad legal authority to moderate content. It is also important to consider this question in the global context in which telecommunications providers operate. Data flows across borders, and many telcos operate in multiple countries. While Section 230 and other U.S. laws and regulations (particularly common carrier regulations) shape the context of telcos' content governance decisions in the United States, the broader policy questions are highly relevant to telcos around the world. Many of them operate with broad legal discretion to affect their subscribers' consumption and creation of content, but many are also deeply susceptible to compulsion and persuasion by state actors.

Fourth, analyzing the public policy implications of telcos' content governance decisions and the variability of their approaches provides insights that are broadly applicable to other entities (like app stores and content delivery networks, to name a few) operating within the broader online information ecosystem.

V. Methodology

To delve deeper into the question of how telcos approach content governance decisions at the subscriber level — the third category of content moderation noted above — a TLS research team examined the terms of service (TOS) and acceptable use policies (AUPs) of 42 telecommunications providers from around the world. To standardize analysis of each telco's policies, our team approached each company's terms of service and/or AUPs with three questions.

1. How does the telco reference illegal online activity as a benchmark?
2. How does it define its authority or discretion to determine what is legal but otherwise objectionable content? And which factors, if any, does it evaluate in making this decision?

---

[25] 47 U.S.C. § 230(c)(2) ("No provider or user of an interactive computer service shall be held liable on account of … any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected").

[26] *See, e.g.*, Caitlin Vogus, Emma Llansó, Samir Jain, *CDT and Technologists File SCOTUS Brief Urging Court To Hold that Section 230 Applies to Recommendations of Content,* Center for Democracy & Technology (Jan. 18, 2023); Quinta Jurecic, Hany Farid, Daphne Keller, Alan Z. Rozenshtein, and Benjamin Wittes, Gonzalez v. Google *and the Fate of Section 230*, Brookings Institution panel (Feb. 14, 2023), video at https://www.brookings.edu/events/gonzalez-v-google-and-the-fate-of-section-230/; Emma Llansó, *Clearing Up Misinformation about Section 230*, Center for Democracy & Technology (Jul. 11, 2019), https://cdt.org/insights/clearing-up-misinformation-about-section-230/.

3. Which consequences or remedies does it enumerate in response to violations of their terms or policies?

The research team did not set out to conduct an exhaustive global survey, but hoped more modestly to avoid a North America-centric focus and to reasonably reflect the diversity of approaches around the world.[27] As noted above, while the legal constraints and due process requirements imposed upon U.S. telcos are considerable and carry far more weight than those placed, for instance, on Content Delivery Networks or app stores, a global examination of telcos tells a different story. The research team also attempted to locate useful data on how companies enforce their terms of service, but there is little systematic or reliable public reporting to be found. As a result, we focused on an area where systematic analysis *was* possible.

This paper focuses on this tool in telcos' kit, rather than on shutdowns or blocked apps or sites because subscriber-level content governance is arguably the least studied of the telcos' levers. Telcos, news reports, commentary, and scholarship say very little – apart from when they turn content over to governments for national security reasons – about how telcos make content governance decisions at the subscriber level.[28] At the moment, we can derive imperfect insights from the only written and publicly available policies that govern the companies' relationship with individual subscribers. Section VI below tries to facilitate a better understanding of (a) the written rules and standards that telcos use to define lawful but otherwise objectionable content and (b) the range of actions or remedies they might take in response to such content.

I fully recognize that these legal terms do not and should not encompass the totality of how telcos think about their subscribers' potentially objectionable activities, but they represent one of the few available starting points for systematic and comparative analysis. Our colleagues Kathleen Stoughton and Paul Rosenzweig noted in a recent essay on internet infrastructure companies' transparency reporting that "46 out of the 56 companies that provided transparency reports at all (82 percent of the total transparency reports) did so without any meaningful disclosure about their content moderation practices."[29] This paper devotes significant attention to the problems posed by this information gap and proposes some concrete ways to begin bridging it.

---

[27] Our search was limited to terms and policies that were available in English.

[28] For instance, Telefonica's and Verizon's transparency reporting are often cited as models of excellence. While their documents — especially Telefónica's — are thorough and illuminating, neither report provides insight into actions taken at the subscriber level in response to violations of terms of service. Like most transparency reports, they address *governmental* requests to access, modify, or remove content. While Telefónica and Verizon simply may not take any subscriber-level content governance actions at their own discretion, we cannot be certain of this theory without confirmation. We also cannot assume that every telco around the world displays fealty to the principle of least intervention at the access layer.

[29] Kathleen Stoughton & Paul Rosenzweig, *Toward Greater Content Moderation Strategy Reporting* (Oct. 6, 2022), Lawfare, https://www.lawfareblog.com/toward-greater-content-moderation-transparency-reporting.

VI. Findings

The research team's review of TOS and AUPs revealed a patchwork of divergent approaches to setting standards for how telcos determine what is "otherwise objectionable" content and act against it. In this section, we summarize key trends and notable observations organized by the three areas of inquiry detailed above.

Finding 1: Unlawful Content

Almost all telcos' TOS and/or AUPs include references to unlawful conduct on services they provide. Mentions of illegal conduct are even more common than the widespread references to offensive, obscene, or harmful content. This is unsurprising. It is broadly in line with the principle of least intervention — that is, the expectation that those providers furthest from users' actions online (*i.e.*, those operating at the access layer), should rely most on the law — rather than their own discretion — for interventions.[30] It also reflects the reality that each telco recognizes the need to comply with laws in various jurisdictions that bear on unlawful speech and censorship. Some policies invoke the terms "illegal" or "unlawful" without further elaboration, while others provide illustrative lists of illegal behavior (incitement to violence, terrorist activities, child pornography, *etc.*).

Finding 2: Standard for defining "otherwise objectionable" content

For some companies, this definition is quite limited in scope, while at other times companies provide significant detail and enumerate what appears to be a comprehensive list.[31] Many telcos also claim broad authority to determine what is objectionable content or a violation of their terms.[32] For example, one provider claimed "sole discretion" to determine whether its services had been used in a way that violated a policy or customer agreement and paired this discretion with the right to "take any responsive actions it deems appropriate."[33] Some companies also referenced reputational risk or harm as a possible justification for

[30] Ben Thompson, *A Framework for Moderation*, Stratechery (Aug. 7, 2019), https://stratechery.com/2019/a-framework-for-moderation/.

[31] Contrast, for example, Sky Broadband's, Telefonica's and BT's more limited approach with the much more comprehensive approach in Vodafone's and Airtel's policies.

[32] *See, e.g.,* Airtel Terms of Service (reserving Airtel's right to act against subscriber conduct that "is an impersonation of another person, grossly harmful, harassing, blasphemous[,] defamatory, obscene, pornographic, paedophilic, libellous, invasive of another's privacy, hateful, or racially, ethnically objectionable, disparaging, relating or encouraging money laundering or gambling, or otherwise unlawful in any manner whatever," as well as content that "[t]hreatens the unity, integrity, defence, security or sovereignty of India or seditious, friendly relations with foreign states, or public order or causes incitement to the commission of any cognizable offence or prevents investigation of any offence or is insulting to any other nation or violates any other provision of law.")

[33] Viasat's AUP asserts broad discretion: "If the Services are used in a way that Viasat, in its sole discretion, believes violate this Policy or a customer agreement, Viasat may take any responsive actions it deems appropriate, including refusing to transmit or post, or removing or blocking, any information or materials, in whole or in part."

intervention, although these references were not common. It is possible that many companies consider that they have made implicit reference to the concept of reputational risk within prohibitions on illegal, "defamatory" or "obscene" content.[34]

In general, we observed two broad approaches to defining objectionable content: telcos either (1) mention objectionable content or a similar term without describing it, or (2) link the concept of objectionable content to some illustrative description. When defining the standard for determining what constitutes such content, telcos' approaches fall into three categories: (1) claim sole discretion to define what is objectionable, without providing criteria or factors considered in this determination; (2) claim to apply a definition of objectionable content but with little to no articulation of the criteria for analysis; or (3) claim to apply a definition of objectionable content along with some criteria for analysis. Our survey revealed many companies that would fit into each of these categories and presented us with several difficult cases that seemed to present mixed approaches.[35]

At the outset of this project, the team expected that the question of how telcos set standards for defining objectionable but legal content would be the most revealing area of analysis. The team's findings bear this out, with the wide variety of approaches suggesting that this is an area that merits greater scrutiny and future research. It is worth noting that we did not find sufficient evidence that these varied approaches correspond neatly to companies headquartered or operating in certain jurisdictions. But we stop short of concluding that geography or jurisdiction plays *no* role in shaping telcos' approach to defining and acting on objectionable content; it seems likely that location matters to at least some extent. More exhaustive surveys of telcos within regions, including analysis of terms not available in English and of smaller companies, might better uncover noteworthy patterns within countries and regions.

A clarification on terms: we have used "otherwise objectionable" to refer to content that does not violate the law but against which telcos still reserve the right to act. Most policies do not use this exact term, but instead employ descriptors like offensive, obscene, or harmful. Importantly, almost all the policies surveyed reference such a category of content. These references are typically (although not always) accompanied by a provision that expressly provides the telco with some degree of discretion to determine which content meets this definition.

---

[34] *See, e.g.*, Telefonica's, Virgin Media's, and Arelion's policies.

[35] I did attempt to neatly categorize which companies' policies fell within each of these categories, but ultimately decided that this exercise couldn't be completed at an acceptable level of rigor and fairness. For instance, a truly fair categorization would be informed by discussion with representatives of each telco to understand the context behind the inevitable ambiguity in legal language and the drafters' intent. Such an undertaking was beyond my resources, but it might be a fruitful area for future research.

Finding 3: Consequences for violations (enforcement)
Again, our limited textual analysis and a general lack of transparency mean that we cannot comment on telcos' actual practices. TLS's research uncovered very little information on how providers enforce their subscriber-level agreements. The paper is therefore necessarily limited to analyzing how telcos *articulate* their right to suspend or terminate services; it cannot credibly speculate on whether and how these subscriber-level enforcement actions take place. Further clarity about actual practice is essential.

Punitive or consequential measures for breach of terms of service/AUPs include: (1) suspension or termination of service (*i.e.*, taking away a subscriber's access to all content) or (2) deletion, removal, or blocking access to user-generated content (*i.e.*, restricting other subscribers' access to content by blocking sites at the DNS level). Some terms mention possible civil action by the telco or referrals to government authority for criminal investigation. Many terms include brief discussions of the processes for investigation and responding to a violation, but they vary in approach. Some companies articulate a process that seems mindful of fairness, requiring that the company first examine the suspected breach of terms and then contact the subscriber suspected of the violation.[36] Others take a stricter approach, reserving the right to suspend or terminate service without notice.[37] Some telcos do not refer to suspension of service and instead refer only to the removal of offending material.[38] Only one provider referenced an obligation to notify the customer so that they could correct the issue and have their service reinstated.

VII. Analysis and Recommendations for Telcos

Our survey reveals a wide variety of approaches in the industry, a general paucity of information about the process for making content governance decisions, and the prevalence of vague substantive standards that preserve broad discretion for companies to decide what is objectionable and how they will respond. This lack of clarity about substantive and procedural standards, coupled with limited public awareness about how telcos actually apply their policies, suggests the potential for significant problems of accountability. While it might be reasonable for telcos to preserve broad discretion to act within their legal agreements with subscribers, the information gap about their actual practice and processes raises concerns and invites speculation.

What would help to navigate through this fog of obscurity and uncertainty? Calls for greater transparency usually headline lists of suggested reforms, and this paper fits that pattern. But it is neither useful to reflexively recommend transparency without defining what it means nor

[36] *See, e.g.*, British Telecommunications Mobile Terms of Service.

[37] *See, e.g.*, NTT, AsahiNet, Mercado Pago, Telenor Terms of Service.

[38] *See, e.g.*, Virgin Media acceptable use policy.

productive to recommend publishing so many metrics that compliance would be overly burdensome to companies or overwhelming to observers. Among others, Daphne Keller has written methodically about the need to be judicious and specific in demanding transparency reforms.[39] Since telcos have generally not been pushed to be forthcoming about either numbers or rationales relevant to decisions that affect individual subscribers' access to services, the first-order need is to encourage meaningful transparency in ways that provide a fuller picture of companies' actual practice.

Stoughton's and Rosenzweig's paper on transparency notes that "[i]n essence, the question that transparency seeks to answer is: To what extent can the organization enhance the public's understanding of that organization's mission, policies, authorities, compliance, activities, programs, and so on? When applied to a specific topic (such as content moderation), the question is: To what extent can the organization provide the public with the information it needs to understand what the organization is doing in terms of that particular issue?"[40] The suggestions that follow proceed in that spirit.

Recommendation 1. Improve transparency about standards and processes, not just numbers.

Although terms of service and acceptable use policies shed some light on how telcos frame their own powers to make content moderation decisions, they do not provide a window into how companies actually apply those written rules. Indeed, legal terms like the ones surveyed here cannot and should not be the source of comprehensive explanations about corporate governance. But addressing this lack of knowledge about how telcos define content moderation problems, apply standards, and remedy their own errors is a prerequisite for identifying thoughtful reforms. To that end, reporting numbers is only one piece of the puzzle. Governments, civil society organizations, and other private companies would benefit from transparency not merely into the number of shutdowns, sites or apps blocked, or the number of subscribers deemed to violate these policies, but also from a clearer picture of the rationales for these decisions and the processes by which they are made.

In fact, many U.S. telcos, including AT&T, Verizon, T-Mobile, and Xfinity, do report some requests to block content and catalog how they have responded. (They do not, however, appear to report on decisions to suspend or ban individual subscribers.) Even with expanded metrics, numerical data would provide a sense of scope and scale of action but shed no light on the "why" and "how" questions. To that end, telecommunications companies should consider:

---

[39] Daphne Keller, *Some Humility About Transparency*, Freeman Spogli Institute for International Studies (Mar. 19, 2021), https://medium.com/freeman-spogli-institute-for-international-studies/some-humility-about-transparency-5814cbbb1a72.

[40] Kathleen Stoughton & Paul Rosenzweig, *Toward Greater Content Moderation Strategy Reporting* (Oct. 6, 2022), Lawfare, https://www.lawfareblog.com/toward-greater-content-moderation-transparency-reporting.

- Publishing short accompaniments to their policies or additional sections within their transparency reports that explain how they decide what is objectionable content and then select a response from a menu of possible options.
- Providing illustrative examples of the types of lawful and otherwise objectionable content that resulted in decisions to suspend or terminate a subscriber. This exhortation to provide "canonical" examples of violating content might seem obvious, but even some highly scrutinized social media companies have not always been forthcoming with these illustrations.[41]

Recommendation 2. Provide a fuller picture of content governance imperatives and available tools.

On a related but broader note, telcos should consider updating their transparency reporting or developing other public resources to present a comprehensive framing of their activities that affect content consumption and creation. Although this paper focuses on terms of service and acceptable use policies, it notes at the outset that telcos can (1) initiate network shutdowns; (2) block individual sites or applications; and (3) make determinations to suspend or deny access to individual subscribers. While the processes for and factors considered when making each of these decisions might well be different, all of them fit within a broader umbrella of content governance decisions taken in response to social or political challenges. Identifying the types of problems to which telcos are responding and then contextualizing the available tools in the tool kit would go a long way toward structuring meaningful discussion.[42]

VIII. Recommendations for Non-Application-Layer Internet Infrastructure Companies

Our analysis of telcos' public policies suggests lessons with broader applicability to non-application layer entities that occupy different positions within the online information ecosystem. What systematic thinking guides content governance decisions for internet infrastructure players like app stores, web hosts, and content delivery networks? The recommendations that follow are designed to suggest paths toward more forward-looking, structured, and systematic thinking about online content governance.

---

[41] *See* Yasmin Green et. al, *Evidence-Based Misinformation Interventions: Challenges and Opportunities for Measurement and Collaboration*, Carnegie Endowment for International Peace (Dec. 2022), https://carnegieendowment.org/files/202212-Wanless_et_al_Misinfo_Interventions2.pdf ("Platforms, for their part, could consider publishing canonical lists or examples of content that meet their own definitions of problematic content and behavior for post-hoc research. This would incentivize the study of interventions that are optimized directly for platform response and would enable a more informed, mutual conversation with external researchers. Facebook and Twitter have done this as part of their efforts to publish information about coordinated activity from inauthentic accounts.")

[42] It is worth nothing that Telefónica's 2021 transparency report does an admirable job of this on a per-country basis.

Recommendation 3. Experiment with Applying Substantive Principles for Infrastructure-Level Content Governance.

Jonathan Zittrain, among others, foresees an "inexorable push" towards infrastructure-level content governance.[43] That push already is well underway. Our exploration of telcos' public policies highlights the risks that attend inconsistent and largely unaccountable approaches to defining and acting against objectionable content. It is premature to suggest a decisive set of governing principles or frameworks to guide internet infrastructure players' approach to content moderation or governance. And it might make little sense for all infrastructure players, operating in diverse contexts, to adopt a highly specific common code of conduct. But it would be a mistake to move too far in the opposite direction and conclude that *no* worthwhile guidance frameworks can be written with at least some applicability across the online content ecosystem.

The GNI Principles

Indeed, several prominent telcos (including some surveyed for this paper) are making efforts in this vein, as evidenced in part by their decision to adopt the Global Network Initiative (GNI) principles.[44] The GNI aims to establish a "global standard for human rights in the information and communications technology (ICT) sector."[45] The following telcos are GNI member companies: British Telecom, Nokia, Orange, Telenor, Telia, Verizon, and Vodafone. Other notable company signatories include Cloudflare, the DDoS mitigation and CDN company mentioned at the start of this paper. GNI's website explains that its "members work together in two mutually supporting ways."[46] The Principles and Implementation Guidelines "provide an evolving framework for responsible company decision making in support of freedom of expression and privacy rights," with the idea that increased company participation will help these norms solidify into global standards. GNI member companies periodically participate in an independent assessment to determine their progress in implementing the principles.[47]

The GNI principles represent a constructive step designed to establish a system of community accountability in the ICT sector that shapes norms and defines increasingly specific standards. The Implementation Guidelines are quite detailed and, in many ways, lend themselves to an actionable company checklist, but they are heavily weighted toward how companies should implement laws and demand accountability from *governments*. In that sense, they reflect the principle of least intervention that should be most rigorous at the access layer but do not address the reality that most internet infrastructure companies face

---

[43] Jonathan Zittrain, *The Inexorable Push For Infrastructure Moderation*, techdirt (Sep. 24, 2021), https://www.techdirt.com/2021/09/24/inexorable-push-infrastructure-moderation/.

[44] *GNI Principles on Freedom of Expression and Privacy*, Global Network Initiative (updated May 2017), https://globalnetworkinitiative.org/gni-principles/.

[45] *About GNI*, Global Network Initiative, https://globalnetworkinitiative.org/about-gni/.

[46] *Id.*

[47] *Id.*

increasing pressure to make at least some decisions based on their own corporate standards and rules.

Santa Clara Principles

The Santa Clara principles for content moderation also are a relevant reference point.[48] They were developed by a coalition of civil society organizations and experts and endorsed by twelve major companies mostly but not exclusively operating at the application layer. The Santa Clara principles are in some ways more ambitious and specific than the high-level GNI principles, but they may not be sufficiently applicable to players across the internet ecosystem.

First published in 2018 and then updated in 2020 and 2021, these principles are principally designed for entities operating at the internet's application layer, although the presence of signatories like Cloudflare indicate their relevance to players that fall between the access and application layers. The Santa Clara Principles are organized into two broad categories. The first grouping of foundational principles consists of human rights and due process, understandable rules and procedures, and cultural competence. The second set of operational principles focuses on the transparent reporting of numbers, the process of providing notice to people, and the appeals process.[49] Without a better understanding of how various infrastructure companies apply their written standards to users deemed to violate their policies, it is difficult to opine on how they might adapt operational principles to their own contexts. But considering our review of telcos' policies and their three main methods of affecting content creation and consumption (shutdowns, blocking or throttling apps or sites, or suspending or banning subscribers), the Santa Clara foundational principles have relevant applications to telcos' and other infra companies' activities.

The Manila Principles on Intermediary Liability

The Manila Principles on Intermediary Liability detail standards that governments should meet when making content access or moderation requests of any online intermediary entity.[50] They come closest to articulating a concise decision-making framework for governments and in that sense share much in common with the decision-making template for companies that I propose in Recommendation 4 below.

The stakes for human rights and due process are high not just for telcos, but also for other entities like app stores, content delivery networks, and DDoS mitigation services. Because of the gatekeeping functions these services play, their decisions can profoundly impact internet access, access to content, and/or the ability to speak online. Our review of telcos' terms reinforces the importance of the second Santa Clara principle's call to publish understandable

[48] *The Santa Clara Principles on Transparency and Accountability in Content Moderation* (2018), https://santaclaraprinciples.org/.

[49] *Id.*

[50] *The Manila Principles on Intermediary Liability*, https://manilaprinciples.org/principles.html.

rules and procedures. This shortfall likely extends to many players operating at other layers of the internet. These three sets of principles demonstrate that there is no need for companies to reinvent the wheel, but there is a compelling need to begin translating high-level principles into more concrete frameworks for companies' decision making.

Recommendation 4. Develop a framework and a standard checklist or decision tree to comprehensively guide content governance activities.

What might a more comprehensive framing of content governance look like in a transparency report or other explanatory public document? To start with, companies should think through which rule sets shape — or should shape — their content governance decisions. In most instances, it is likely that two broad sets of rules are at play: on the one hand, the global and sub-national patchwork of laws that companies must navigate and, on the other, the company's own rules for governing content, often designed as a common set of standards that are applied across all the jurisdictions in which it operates.[51]

Walking through the following questions could help guide an internet infrastructure company as it seeks to map its approach to dealing with both unlawful speech as well as speech considered "lawful but awful." The rubric below attempts to simplify and concretize some (but not all) of what appears in the GNI Implementation Guidelines. It misses some of the points captured therein and adds or reframes other points. My rubric should hardly supplant the carefully crafted GNI Principles and Implementation Guidelines; instead, it might serve as a companion guide that helps companies — especially non-GNI members with more limited resources — simplify the process of developing a content governance program. The rubric explicitly acknowledges that two broad bodies of rules shape private infrastructure entities' content governance efforts: (1) the laws imposed by local and national governments and (2) the efforts to deal with content that might be lawful but that the company might nevertheless deem objectionable. The following template is a starting point that should, like open-source code, be improved upon and then tailored by companies to be as useful and specific as possible to the contexts in which they operate.

[51] *See, e.g.*, Chinmayi Arun, *Facebook's Faces*, 135 HARV. L. REV. 236, 239-240 (Mar. 20, 2022), https://harvardlawreview.org/2022/03/facebooks-faces (explaining that social media companies like Meta "us[e] two different systems to regulate content" — one grounded in implementation of local laws and the other a set of "privately ordered 'platform law," or company rules, that Meta calls its Community Standards").

CORPORATE CONTENT GOVERNANCE TEMPLATE

Unlawful Content

1. Define unlawful content using illustrative examples.
2. Enumerate the factors and processes by which our company assesses the validity of government requests through legal process to turn over, take down, block, or otherwise restrict content. Demonstrate with examples the thought process by which we would either decide to comply or refuse to do so.
3. Enumerate the factors and processes by which our company assesses the validity of governmental efforts to *persuade*, rather than compel, us to take the types of actions listed in #2. Demonstrate with examples the thought process by which we would decide to comply or refuse to do so.
4. List and describe the technical mechanisms (blocking, filtering, *etc.*) available to us to comply with what we consider to be a valid governmental request (whether expressed as legal compulsion or extra-legal persuasion). Define which of these mechanisms least intrude on speech, privacy, and other human rights.

Lawful but "Otherwise Objectionable" Content

5. Enumerate the factors or criteria that guide our company's classification of content or behavior as objectionable even when the law does not clearly prohibit the content or conduct in question.
6. Enumerate the factors/criteria and processes by which our standards or criteria help us decide whether we act against this content.
7. List and describe the available technical or policy enforcement mechanisms available to our company.

Notice and Redress

8. Define the processes for notifying affected parties (subscribers, customers, users, governments, *etc.*) of content governance decisions and the process by which those parties can request information about or challenge the decisions we make.
9. Define our approach for dealing with the challenge of security concerns that cannot (completely) be revealed to an affected party. For example, how should we handle acting on classified threat information from a government and our resulting inability to (fully) disclose to a subscriber/customer our rationale for a content governance decision?[52]

[52] *See, e.g.*, Alex Joel, "*Without Confirming or Denying": Opaque Notification and National Security Redress*, Privacy Across Borders (Feb. 2023), https://privacyacrossborders.org/wp-content/uploads/2023/02/Opaque-Notification-and-National-Security-Redress.pdf.

Assessing Accuracy, Efficacy, and Fairness

10. Define our methods to assess the accuracy, efficacy, and fairness of our decisions. Realistically tailor the approach to the company's available resources; internal compliance checks can be costly and risk management programs must be based on an honest assessment of capacity. As the program matures, consider the utility of engaging independent consultants/assessors/auditors and establish processes designed to maximize their impartiality.[53] In particular, we should:

    a. Consider how to compile data that reveals how often we flag and act against unlawful or otherwise objectionable content as we define it.
    b. Define ways to measure (qualitatively and quantitatively) how effective those efforts are at improving the safety of the discursive environment our company regulates.
    c. Design internal reviews to assess the fairness of our decision-making processes and consider establishing a process for external observers to assess our standards and processes.

11. Overall, define and enumerate the least intrusive and most human rights-protective methods of achieving our content governance objectives, whether those objectives flow from legal requirements or decisions to enforce our own policies. (For example, would a telco elect to block access to a few sites that are known purveyors of terrorist content instead of shutting down internet access within a defined geography because a government expresses concerns about public safety and political instability? [54])

Transparency

12. Provide meaningful public transparency about our content governance practices, accounting for available resources and the maturity of our content governance

---

[53] *See., e.g.,* Jim Dempsey, *Enforcement of Cybersecurity Regulations Parts 1 and 2*, Lawfare (Mar. 22, 2023 and Mar. 29, 2023), https://www.lawfareblog.com/enforcement-cybersecurity-regulations-part-1, and https://www.lawfareblog.com/enforcement-cybersecurity-regulations-part-2 (exploring both the virtues and inherent limitations of self-assessment and hired consultants, in contrast to the role that truly independent auditors or assessors can play.)

[54] *See, e.g.,* Nathan Matias, *Choosing Between Content Moderation Interventions*, Freedom to Tinker (May 7, 2019), https://freedom-to-tinker.com/2019/05/07/choosing-between-content-moderation-interventions/; Caitlin Vogus and Emma Llansó, Making Transparency Meaningful: A Framework for Policymakers at p. 21 n. 46, Center for Democracy and Technology (Dec. 2021), https://cdt.org/insights/report-making-transparency-meaningful-a-framework-for-policymakers (Explaining that, in the context of internet platform companies, "[c]ontent moderation is not just a binary decision to either take down content or accounts or allow them to remain on a service; depending on how they have designed their service, intermediaries can take a wide variety of actions against violative content, some of which may not be immediately obvious to the user who posted the content.")

programs.[55] Present as comprehensive a picture of content governance activities as possible.

13. In transparency reports or other public documents, describe and/or provide the following, with illustrative examples:
    a. The imperatives that drive our engagement in content governance;
    b. The rationales and standards we apply for making decisions; and
    c. The technical, legal, and policy tools we can use to enforce our standards and decisions.

These subcategories and specific considerations illustrate the form that a company's decision-making template might take. Internet infrastructure companies should tailor it to their own operational contexts and develop a flowchart or decision tree that guides them in investigating and addressing content suspected of being illegal or otherwise objectionable.

Given the likelihood that many such decisions are currently made *ad hoc*, putting pressure on internet infrastructure players to document their content governance standards and decision-making processes would achieve the dual objectives of advancing public accountability *and* improving companies' internal clarity and consistency. For many entities, a transparency exercise like this might even reveal that they *rarely* exercise their right to limit a subscriber's or client's access to their service based upon their own determination that content is lawful but otherwise objectionable. Such a finding — perhaps evidence that the principle of least intervention is at play — would have significant consequences for public debates. The opposite result would be equally significant.

IX. Looking Forward

We are living in a time of hyper-focus on the relationship between social media companies and democracies' health. This is hardly a misguided preoccupation, but the focus on content moderation and content governance at the internet's application layer has diverted necessary attention from other parts of the online content ecosystem. Absent laws or other pressures that compel greater transparency, non-application-layer infrastructure companies are incentivized to conduct their content governance practices largely in the shadows. This means, as the Cloudflare and Kiwi Farms example demonstrates, that many companies will largely continue to deal with challenging requests or scenarios on a case-by-case basis. The

[55] The following civil society reports are useful references for thinking through what meaningful transparency looks like: Caitlin Vogus and Emma Llansó, *Making Transparency Meaningful: A Framework for Policymakers*, Center for Democracy and Technology (Dec. 2021), https://cdt.org/insights/report-making-transparency-meaningful-a-framework-for-policymakers; Gennie Gebhart, *Who Has Your Back? Censorship Edition 2019*, Electronic Frontier Foundation (Jun. 12, 2019), https://www.eff.org/wp/who-has-your-back-2019#appeals-transparency; Spandana Singh and Kevin Bankston, *The Transparency Reporting Toolkit: Content Takedown Reporting,* New America Open Technology Institute (Oct. 25, 2018), https://www.newamerica.org/oti/reports/transparency-reporting-toolkit-content-takedown-reporting/.

prevailing *ad hoc* approach is inefficient as a matter of good governance and unsustainable from a security and a digital rights perspective.

And if, as some have argued, the era of social media is drawing to a close,[56] we should hardly assume that such a decline will cause governments and private actors to suddenly stop seeking ways to moderate or otherwise govern online content. Instead, because "channels of communications are also channels of control," we should expect governments and other players to try to exert *more* control in other parts of the online internet ecosystem.[57] It is not at all clear that governments, civil society, or companies are prepared for those potential shifts.

Internet infrastructure companies want clarity from governments about rules of the road, but most companies operate in multiple jurisdictions and must navigate an intimidatingly diverse thicket of global and local laws. This means that private players bear a significant share of the burden of standardization. Companies have responded by developing their own private bodies of law, but these internal rules for content governance are largely opaque to the interested public. Civil society, private sector, and governmental collaboration have produced some important principles and implementation guidelines, but they are not easily translated into practical decision-making playbooks for companies that can be grabbed off the shelf and applied to rapidly developing crises.

This paper aims to accelerate the process of simultaneously addressing knowledge gaps and the problem of standardization by reorienting our focus from content "moderation" to content "governance." It presents the results of a survey of 42 telecommunications providers' terms of service and acceptable use policies as a window into the broader online content ecosystem. This paper's case study sheds some light on how these gatekeepers at the access layer of the internet conceive of their authority to define and act against content at the subscriber level that they deem to be lawful but otherwise objectionable. The research presented here reveals a wide variety of approaches to defining that content and the standards for addressing it.

This paper offers recommendations for telcos to improve meaningful transparency and concludes by presenting a template that is potentially useful for any internet infrastructure company crafting its approach to content governance. It is hardly "the answer," but my hope is that is a starting point for structured corporate decision making about content governance that should be critiqued, revised, and then tested against companies' specific operational contexts.

Transparency reforms and thoughtful experimentation with governance principles are challenging and invite potentially unwelcome attention. Companies are unlikely to develop

---

[56] *See, e.g.*, Ian Bogost, *The Age of Social Media is Ending*, The Atlantic (Nov. 10, 2022), https://www.theatlantic.com/technology/archive/2022/11/twitter-facebook-social-media-decline/672074/.

[57] Jonathan L. Zittrain, THE FUTURE OF THE INTERNET AND HOW TO STOP IT at p. 42 (2008).

them without constructive and persistent engagement from civil society. Just as researchers and advocates regularly engage social media companies on a wide range of public policy issues, they should do the same with telcos and other internet infrastructure players. The trends outlined here underscore the need for a more searching examination of how different internet infrastructure players shape the online content ecosystem. That exhortation applies as much to those alarmed by the proliferation of harmful online content as it does to those chiefly worried about the erosion of digital rights in the name of safety and security.

APPENDIX: Telecommunications Providers Surveyed

Company	Country HQ/Principal Zones of Operation	Terms of Service (TOS) or Acceptable Use Policy (AUP)
Comcast	US	AUP
Viasat	US	TOS/AUP
Cox Communications	US	TOS/AUP
Mediacom	US	TOS/AUP
CenturyLink	US	TOS/AUP
Astound Broadband	US	TOS/AUP
Frontier Internet	US	TOS/AUP
AT&T	US	AUP
Time Warner/Spectrum	US	AUP
Verizon	US	AUP
Spectrum	US	AUP
Lumen Technologies Inc. (CenturyLink)	US	AUP
Safaricom	Kenya	TOS
Telkom	South Africa	AUP
WorkOnline	South Africa	AUP
Globacom (Glo) Ltd.	Nigeria	AUP
Vodacom	South Africa	AUP
Reliance Jio	India	TOS
Airtel Secure	India	TOS
NTT Communications	Japan	AUP
AsahiNet	Japan	TOS
NTT	Japan	AUP
SK Telecom	South Korea	TOS
China Telecom	China	AUP
PTCL	Pakistan	AUP
NayaTel	Pakistan	TOS
Telecom Argentina	Argentina	AUP
Telmex	Mexico	TOS
Claro Brasil	Brazil	TOS
Telefónica	Spain/Europe	TOS
Mercado Libre	Argentina/Latin America	TOS
Millicom (International Cellular SA)	Latin America	TOS

Telia Company	Sweden/Europe	TOS/AUP
Arelion	Sweden/Europe	TOS/AUP
Telenor	Norway/Europe	TOS
Virgin Media	UK/Europe	TOS/AUP
British Telecom	UK/Europe	TOS/AUP
Bouygues Telecom	France/Europe	TOS
Sky Broadband	UK/Europe	TOS/AUP
Orange	France/Europe	AUP
Vodafone	UK/Europe	AUP
Deutsche Telekom	Germany	AUP