WHO CONTROLS ONLINE FREE SPEECH

State Regulation and Platform Power in Online Moderation

By
Yilan Li

Submitted to Central European University - Private University
Legal Studies Department/Central European University/Comparative Constitutional Law

In partial fulfilment of the requirements for the degree of Master of Laws in Comparative Constitutional Law

Supervisor: Tommaso Soave

Vienna, Austria
2025

COPYRIGHT NOTICE

[1]

[1] Icon by Font Awesome.

AUTHOR'S DECLARATION

I, the undersigned, Yilan Li, candidate for the LLM degree in Comparative Constitutional Law declare herewith that the present thesis titled "Who Control Online Free Speech" is exclusively my own work, based on my research, and only such external information as properly credited in notes and bibliography.
I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright.

I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Vienna, 15 June 2025

Yilan Li

ABSTRACT

This thesis examines how online speech is governed through three competing models—classical liberalism, notice-and-takedown regulation, and platform self-regulation—framed through the theoretical lens of the free speech triangle, which conceptualizes the interaction between states, platforms, and users. The first part analyzes the U.S. model grounded in classical liberalism, where Section 230 shields platforms from liability and state interference. The second explores the notice-and-takedown model, exemplified by the European Union's Digital Services Act and national laws such as Germany's NetzDG and France's SREN law, which impose legal obligations on platforms to remove illegal or harmful content. The third focuses on Meta's Oversight Board as an experiment in digital constitutionalism, in which the platform applies international human rights norms to its own moderation practices. While each model aims to balance expression and harm, they present distinct risks of state overreach, private censorship, and regulatory fragility. Through comparative constitutional analysis, this thesis evaluates how different legal frameworks mediate the contested space of online free speech.

Keywords: free speech, content moderation, Section 230, Digital Services Act, Meta Oversight Board, platform governance, digital constitutionalism

ACKNOWLEDGMENTS

Writing about freedom speech during the turbulent onset of Trump's second term and amid the rightward shift in Vienna was a particularly weighty undertaking, especially as someone from a country where free speech remains heavily restricted.

I am deeply grateful for the intellectually open, supportive, and diverse environment at Central European University, which allowed me to reflect on freedom of speech from a place of genuine dialogue. I owe special thanks to my supervisor, Tommaso Soave, whose guidance was rigorous and inspiring. I am also sincerely thankful to the other faculty members at the Legal Studies Department; this past year has been the most intellectually demanding yet fulfilling chapter of my academic journey.

I am also especially grateful to Zhenxing (Benjamin) Tan, the best Chinese cook I know. I sincerely hope that when it's his turn to write a philosophy thesis next year, it won't be so exhausting that it ends with an ambulance ride again.

Finally, I would like to thank my loving family in China for their unwavering support: mom (Qin Guo), dad (Xibin Li), grandma (Huilan Tang), grandpa (Shiyi Guo).

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION

The "marketplace of ideas" is a rationale for the First Amendment's Free Speech Clause, allowing "truth" to be tested among various ideas "in the competition of the market".[2] And debates in this marketplace should be "uninhibited, robust, and wide-open".[3] Under such a theory, editorial discretion, the right not to speak, and other doctrines protecting free speech emerged.

Nowadays, online social media platforms have become the major avenue for the exchange of views. As of February 2025, 5.24 billion individuals worldwide were Social Media users, which amounted to 63.9 percent of the global population.[4] Early in *Reno v. ACLU* (1997), Judge Stevens recognized the Internet's function as "vast democratic forums".[5] However, social media platforms are not neutral marketplaces. These platforms implement content moderation policies that enable them to remove harmful or illegal content. While such moderation is often framed as an act of corporate social responsibility, it also serves a distinctly capitalist function: maximizing profit by capturing and retaining user attention.[6] Beyond merely ensuring that users feel safe, platforms may algorithmically promote or suppress content or even engage in subtle forms of digital manipulation to optimize engagement.

Therefore, content moderation by social media platforms becomes a crucial factor in shaping the modern landscape of freedom of speech. In principle, under the framework of classical liberalism, the Free Speech Clause of the First Amendment applies only to state actors, as

[2] *Abrams v United States* (1919) 250 US 616 (Supreme Court).
[3] *New York Times Co v Sullivan* (1964) 376 US 254 (Supreme Court).
[4] 'Internet and Social Media Users in the World 2025' (*Statista*) <https://www.statista.com/statistics/617136/digital-population-worldwide/> accessed 18 April 2025.
[5] *Reno v ACLU* (1997) 521 US 844 (Supreme Court).
[6] Kate Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2017) 131 Harvard Law Review 1598, 1627.

articulated by the State Action Doctrine.[7] Consequently, social media platforms—as private entities—are not bound by the First Amendment and are thus entitled to exercise content moderation according to their standards, under the doctrine of so-called editorial discretion. However, due to the increasingly public nature of social media platforms and the scale of their content moderation powers, they are increasingly viewed as *de facto* "governors" or "special-purpose sovereigns".[8] In this view, platforms serve public functions and resemble quasi-municipal corporations and therefore should bear certain constitutional obligations to protect freedom of expression.[9]

Indeed, with the rise of social media platforms' private governance, free speech regulation has shifted from the 20th-century dualist model between two players, the state and the speaker, to a triangle model.[10] The triangle model of free speech regulation conceptualizes the interaction between public authorities, digital infrastructure providers, and speakers as a dynamic system in which states pressure platforms to moderate content, platforms govern user speech through private rules, and users influence both through protest, compliance, or migration.[11]

Under this situation, different approaches to regulating online moderation have emerged. Some still adhere to the principle of the "marketplace of ideas," uphold platform autonomy and permit only minimal state intervention—the most famous example being Section 230 of the Communications Decency Act of 1996.[12] Others hold intermediaries liable and effectively compel them to engage in collateral censorship, as embedded in the EU's Digital Services Act

[7] 'State Action Doctrine' (*Oxford Constitutions*) <https://oxcon.ouplaw.com/display/10.1093/law-mpeccol/law-mpeccol-e473> accessed 18 April 2025.
[8] Jack M Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation' (Social Science Research Network, 9 September 2017) 36 <https://papers.ssrn.com/abstract=3038939> accessed 18 April 2025.
[9] 'Quasi-Municipal Corporation' (*LII / Legal Information Institute*) <https://www.law.cornell.edu/wex/quasi-municipal_corporation> accessed 18 April 2025.
[10] Jack M Balkin, 'Free Speech Is a Triangle Essays' (2018) 118 Columbia Law Review 2011, 2013.
[11] *ibid 2015.*
[12] Communications Decency Act § 230 1996 (USC).

(DSA) and the E-Commerce Directive.[13] Within this system, Germany's Network Enforcement Act (NetzDG) is the most well-known example, followed by similar policy initiatives introduced by lawmakers in Australia, Austria, Brazil, France, India, Nigeria, Singapore, and Russia.[14] Moreover, some forms of private governance have gone a step further by establishing quasi-judicial entities, such as internal courts that apply international human rights law to determine whether a platform's moderation decision is "legal".

This thesis does not seek to provide a global best solution to balancing free speech and harmony in the online community. Instead, this article concentrates on a critical analysis of the advantages and limitations inherent in governance approaches corresponding to each vertex of the free speech triangle model. To develop this central idea, Chapter Three will analyze how these theories manifest in practice through three regulatory models: (1) the U.S. model of platform immunity under Section 230; (2) the European notice-and-takedown approach, and (3) Meta's experiment in digital constitutionalism via the Oversight Board.

Since free speech is influenced by multiple elements, discussing content moderation in Mainland China, Iran or Russia is rather different from the United States or Western Europe. In authoritarian regimes, content moderation is systematically subordinated to the state's political imperatives, rendering platforms either direct instruments of censorship or heavily constrained intermediaries.[15] Therefore, this article will only cover case studies of "countries with freedom of speech" according to Freedom House's annual Freedom in the World report.[16]

[13] Directive 2000/31/EC on Electronic Commerce (E-Commerce Directive) 2000 (OJ L 178); Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act) 2022 (OJ L 277).
[14] 'Countries and Territories' (*Freedom House*) <https://freedomhouse.org/country/scores> accessed 22 April 2025; Netzwerkdurchsetzungsgesetz (NetzDG) 2017 (BGBl I).
[15] 'Freedom on the Net' (*Freedom House*, 16 October 2024) 14 <https://freedomhouse.org/report/freedom-net> accessed 24 April 2025.
[16] 'Countries and Territories' (n 14).

CHAPTER TWO: THEORETICAL FOUNDATIONS

1. The Rise of "Lawless" Social Media Platforms

In 2015, Carr and Hayes proposed a definition of social media platforms as "Internet-based, disentrained, and persistent channels of mass personal communication facilitating perceptions of interactions among users, deriving value primarily from user-generated content."[17] Social media platforms provide services that host and organize users' shared content and social interactions and moderate the content and activity of users for different interests including profiting.[18]

While social media platforms derive substantial economic value from user-generated content and the user attention it attracts, this very content also poses significant liabilities.[19] Given that these platforms are distrained—open to participation from individuals across the globe—there is an inherent risk of hosting content that may be considered harmful or unlawful. This includes, but is not limited to, hate speech, terrorist propaganda, and child sexual abuse material, all of which are prohibited in many jurisdictions. As a result, public authorities increasingly impose legal obligations on platforms to comply with domestic laws regarding illegal content, and failure to do so may lead to lawsuits, regulatory sanctions, or substantial financial penalties. To mitigate legal liabilities and manage risks associated with harmful or unlawful content, social media platforms have developed content moderation systems and terms of service.

[17] Caleb T Carr and Rebecca A and Hayes, 'Social Media: Defining, Developing, and Divining' (2015) 23 Atlant ic Journal of Communication 46, 49.
[18] Barrie Sander, 'Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human R ights-Based Approach to Content Moderation' (2020) 43 Fordham International Law Journal 939, 944.
[19] Sarah T Roberts, 'Digital Detritus: "Error" and the Logic of Opacity in Social Media Content Moderation' [20 18] First Monday <https://firstmonday.org/ojs/index.php/fm/article/view/8283> accessed 24 April 2025.

The practice of online content moderation has undergone a significant transformation—from being community-led to company-led.[20] In the early days of the internet, content on online forums was often curated and monitored by community administrators, who were themselves members of the user community. However, with the large-scale commercialization of the Internet and the rise of social media platforms as profit-driven enterprises, moderation responsibilities have shifted toward what is now referred to as "commercial platform moderation".[21] Under this model, moderation is carried out by the platform companies themselves, often through a combination of algorithmic filtering, outsourced moderation labor, and internal policy enforcement teams.

Pozen criticized social media platforms' content moderation system as "authoritarian constitutionalism".[22] Social media platform companies can arbitrarily establish and amend their terms of service without transparency or democratic legitimacy and hold the power to interpret and enforce them.[23] Even though terms of service are literally contracts between the social media platform company and the user, terms of service are non-negotiable, rejecting them means social media exclusion. Therefore, social media content moderation policies are *de facto* "constitutional instruments regulating the exercise of fundamental rights online."[24] If there is

[20] Robert Gorwa, Reuben Binns and Christian Katzenbach, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020) 7 Big Data & Society 2053951719897945, 2.

[21] Edoardo Celeste and others, *The Content Governance Dilemma: Digital Constitutionalism, Social Media and the Search for a Global Standard* (Springer International Publishing 2023) 9 <https://link.springer.com/10.1007/978-3-031-32924-1> accessed 22 April 2025.

[22] David E Pozen, '7. AUTHORITARIAN CONSTITUTIONALISM IN FACEBOOKLAND', *The Perilous Public Square* (Columbia University Press 2020) <https://www.degruyterbrill.com/document/doi/10.7312/poze19712-008/html> accessed 25 April 2025.

[23] Celeste and others (n 21) 14; Giovanni De Gregorio, 'From Constitutional Freedoms to the Power of the Platforms: Protecting Fundamental Rights Online in the Algorithmic Society' (Social Science Research Network, 2018) <https://papers.ssrn.com/abstract=3365106> accessed 25 April 2025.

[24] CELESTE AND OTHERS (N 21) 15.

no intervention and oversight from an external body, social media platforms would be a "lawless" space where platform companies have almost absolute power.[25]

Intuitively, this appears in tension with the critical role that social media platforms play in the sphere of free speech. As Justice Kennedy noted, social media has become "the most powerful mechanisms available to a private citizen to make his or her voice heard" which has hosted massive public expression.[26] Against this backdrop, it is worth questioning whether the private-public distinction can genuinely permit platforms to remain beyond the reach of democratically voted laws when they decide what they will distribute and what will they moderate. Therefore, there comes an online content moderation governance dilemma between governance by platforms and governance of platforms.[27]

2. The Laissez-Faire Logic of the First Amendment

The First Amendment of the United States Constitution provides:

"Congress shall make no law respecting…prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press…"[28]

The First Amendment protects speech only from interference by public authorities. This design is largely rooted in classical liberal theory. Classical liberal theory posits that individuals are independent moral agents capable of making up their minds.[29] Moreover, classical liberal theory has an inherent distrust of public authorities. In the context of free speech, this position means that the state must respect individuals' decision of what to say as well as their choice

[25] Nicolas P Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* (Cambridge University Press 2019) 11.
[26] *Packingham v North Carolina* (2017) 582 US 98 (Supreme Court) 8.
[27] Tarleton Gillespie, 'Regulation of and by Platforms' in Jean Burgess, Alice Marwick and Thomas Poell, *The S AGE Handbook of Social Media* (SAGE Publications Ltd 2018) <https://sk.sagepub.com/reference/the-sage-hand book-of-social-media/i2081.xml> accessed 27 April 2025.
[28] U.S. Constitution amend. I 1791 (US Const).
[29] Immanuel Kant, *Groundwork for the Metaphysics of Morals* (Allen W Wood ed, Yale University Press 2002).

about the speech to which they wish to be exposed and should not override individual preferences.[30] As John Mill wrote in *On Liberty*, truth is best discovered not through state regulation, but through private interaction: "If it is not fully, frequently, and fearlessly discussed, it will be held as a dead dogma, not a living truth".[31] As Isaiah Berlin observed, errors in identifying individuals' supposed true interests can have deeply illiberal consequences, empowering third parties to impose coercion under the guise of promoting those interests.[32] Therefore, Frederick Schauer argued "freedom of speech should be best characterized as an absence of governmental interference", which is closest to Isaiah Berlin's definition of "negative liberty".

The classical liberal theory root largely shapes the American conception of free speech and renders the emergence of the state action doctrine. The state action doctrine means the First Amendment limits the public authorities' ability to regulate private speech, while private action is not usually subject to First Amendment constraints. The First Amendment also prohibits the public authorities from coercing private actors to take actions that suppress other private entities' speech.[33] Private censorship is always rejected by the courts in the United States.[34] Moreover, recent cases such as *Knight First Amendment Institute v. Trump* (2019) have refined this public-private distinction in the field of the state action doctrine, ruling that actions by government officials to block access to forums or delete past communications on private social media platforms constitute impermissible state action interference.[35]

[30] Christopher Yoo, 'Technologies of Control and the Future of the First Amendment' (2011) 53 William & Mary Law Review 747, 750.
[31] John Stuart Mill and Gertrude Himmelfarb, *On Liberty* (1st edition, Longman 1998) 58.
[32] Isaiah Berlin, *Liberty: Incorporating Four Essays on Liberty* (Henry Hardy ed, 2nd edition, Oxford University Press 2002) 197.
[33] 'National Rifle Ass'n of America v. Vullo' (*Harvard Law Review*, 11 November 2024) <https://harvardlawreview.org/?p=16794> accessed 28 April 2025.
[34] Eric Barendt, *Freedom of Speech* (2nd edition, Oxford University Press 2005) 152.
[35] *Knight First Amendment Inst at Columbia Univ v Trump* 928 F3d 226 (Second Circuit Court of Appeals).

However, there is an inherent flaw in this classical liberalism model of free speech. The First Amendment, for example, gained substantive meaning in the 1920s, during and after the First World War, when the United States implemented strict speech control policies.[36] The famous cases at that time, *Schenck v. United States* (1919), *Abrams v. United States* (1919), and *Gitlow v. New York* (1925) were mostly political dissidents resisting the state's suppression of their ability to speak out. During the First Amendment's flourishing in the 1960s and 1970s, the press, a so-called public watchdog" stood at the center of free speech debates, New York Times Co. v. Sullivan (1964), New York Times Co. v. United States (1971) and Miami Herald Publishing Co. v. Tornillo (1974) stand out as cases in point.[37] The same background underlying this period of the First Amendment's development was the premise of informational scarcity, driven by the high cost of being a speaker.[38] In this situation, free speech's main mission is to protect individual speakers from state authorities.

The arrival of the digital age has significantly lowered the cost of speaking, creating massive speech both inexpensive to produce and often of markedly low social value, which is called "cheap speech" by Eugene Volokh.[39] In this new environment, speech is no longer scarce, instead, the audience's attention is much more valuable. The social media platforms that host vast quantities of cheap speech have been granted a disproportionate amount of power to determine which speech is amplified and which is silenced. Motivated by attention-resale business, social media platforms are using their strong recommendation and moderation power to gain profit from the resale of their users' time and attention.[40] The inclusion of the new third player, social media platforms, made free speech shift from a dualist structure between speaker

[36] Tim Wu, 'Is the First Amendment Obsolete?' (2018) 117 Michigan Law Review 547, 551.
[37] Geneva Overholser and Kathleen Hall Jamieson, *The Press* (Oxford University Press 2005) 169.
[38] Wu (n 36) 553.
[39] Eugene Volokh, 'Cheap Speech and What It Will Do Symposium: Emerging Media Technology and the First Amendment' (1994) 104 Yale Law Journal 1805.
[40] Wu (n 36) 556.

and public authorities to a free speech triangle, introduced by Balkin, between public authorities, platforms, and speakers.[41] While the new triangle-free speech environment includes "new school" speech control in which social media platform's moderation power large involved, the First Amendment is exclusively designed and enforced as a tool to combat "old school" speech control, which means the public authorities' silence of political dissents.[42]

The classical liberalism understanding of free speech makes it hard to make laws to regulate private social media platform companies, even though it has massive influence on individual speakers. Supporters of classical liberalism would say, as David Cole noted "that is the price of freedom".[43]

3. Regulating Speech in the Name of Rights: Europe's Regulationism

As previously discussed, classical liberalism adopts an uncompromising approach to free speech, sometimes to an extreme degree. The strict state action doctrine and strong protection over social media platform companies are almost unique to the United States.[44] There are many other states' laws and courts have ordered social media platforms to restrict their user's speech, with the European Union (EU) as an important example.

The European Convention on Human Rights (ECHR) is an international human rights treaty adopted under the auspices of the Council of Europe in 1950. It applies not only to the EU but also to 47 member states including non-EU countries such as Turkey, and Ukraine. The ECHR was designed after the Second World War, which provided horrific examples of how States can

[41] Balkin (n 10).
[42] *ibid 2015.*
[43] David Cole, 'Who Should Regulate Online Speech?' (2024) 71 *The New York Review of Books* <https://www.nybooks.com/articles/2024/03/21/who-should-regulate-online-speech/> accessed 28 April 2025.
[44] SUZOR (N 25) 49.

misuse their sovereign power and deeply violate individuals' autonomy, dignity, and freedom.[45] Against this background, the ECHR aimed to protect individual's fundamental rights from interference by a public authority. Therefore, the ECHR imposes legal obligations on the Convention States, and Article 34 requires individual applications must be directed against one of the Convention States. Only vertical cases between private actors and the State could be admissible.

Article 10 of the (ECHR) provides protection of freedom of speech:

"1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television, or cinema enterprises.

2. The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions, or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary."[46]

The most obvious feature of Article 10 is its detailed list of circumstances under which limitations on expression are permitted, as set out in Article 10(2). Although Article 10 is historically grounded in the concept of negative liberty, its formulation differs from the First Amendment in that it expressly allows for restrictions under a structured proportionality

[45] Claire Loven, '"Verticalised" Cases before the European Court of Human Rights Unravelled: An Analysis of T heir Characteristics and the Court's Approach to Them' (2020) 38 Netherlands Quarterly of Human Rights 246, 247.
[46] Convention for the Protection of Human Rights and Fundamental Freedoms 1953.

framework. The European Court of Human Rights (ECtHR) balances the right to freedom of expression against other protected human rights, such as the right to privacy or dignity. This rights-balancing approach contrasts sharply with that of U.S. courts, which tend to afford broad and often categorical protection to certain forms of speech under the First Amendment.[47] Article 10(2) provides the ECtHR with a proportionality framework to determine whether state interference is "necessary in a democratic society". However, this does not mean that the ECHR is more permissive of state restrictions on freedom of expression. While the ECtHR allows some margin of appreciation, it has interpreted Article 10(2) with a structured and often rigorous proportionality analysis, particularly where political speech or matters of public interest are concerned.

In Handyside v. United Kingdom (1976), the ECtHR stressed that the test of necessity in Article 10(2) is a rigorous one, "neither has it the flexibility of such expression as 'admissible', 'ordinary', 'useful' or 'desirable'" and stated that freedom of speech applies to "those that offend, shock or disturb the State or any sector of the population" due to democratic society's demands of pluralism, tolerance, and broadmindedness.[48] In *Lingens v. Austria* (1986), the ECtHR ruled that the conviction of Austrian journalist Lingens was a breach of the freedom of speech. The ECtHR stated that a state's interference with freedom of speech could be justified only if it was "prescribed by law" and "necessary in a democratic society". Even though the Austrian court's conviction of defamation was prescribed by law and pursued the legitimate aim of protecting the reputation of others, it is not "necessary in a democratic society" since such a conviction would likely deter journalism from contributing to public discussion.[49]

[47] BARENDT (N 34) 66.
[48] *Handyside v United Kingdom* (1976) 24 Eur Ct HR (European Court of Human Rights).
[49] *Lingens v Austria* [1986] European Court of Human Rights App. No. 9815/82, 8 Eur HR Rep 407.

Although the ECHR is formally designed to have a vertical effect, regulating how states treat individuals, it also has a horizontal dimension, through which it offers substantive protection of fundamental rights in relationships between private actors. The ECtHR does so by way of imposing horizontal "positive obligations" on the Convention States, which require them to take action to secure the rights and liberties guaranteed in the Convention in relations between private actors.[50] In *Airey v. Ireland* (1979), the ECtHR first imposed a positive obligation upon the Convention States: "Although the object of Article 8 is protecting the individual against arbitrary interference by the public authorities, in addition to this primarily negative undertaking, there may be positive obligations".[51] In *X. and Y. v Netherlands* (1985), the ECtHR recalled *Airey* and added that "these obligations may involve in the sphere of the relations of individuals between themselves".[52]

The Convention States also burden positive obligations under Article 10 of ECHR. In *Ozgur Gundem v. Turkey* (2000), the ECtHR recalled the importance of freedom of speech as "one of the preconditions for a functioning democracy", and stressed that the exercise of this freedom "does not depend merely on the State's duty not to interfere, but may require positive measures of protection, even in the sphere of relations between individuals".[53] This positive obligation is expanded in *Dink v. Turkey* (2010), where the ECtHR decided that under Article 10, the state has the positive obligations not only to protect freedom of speech but also to "create an environment for public debate that allowed opinions and ideas to be expressed without fear, including those that might offend or even shock".[54] Positive obligations under Article 10 were reiterated in *Uzeyir Jafarov v. Azerbaijan* (2015), *Huseynova v. Azerbaijan* (2015), *Khadija*

[50] Loven (n 45) 247.
[51] *Airey v Ireland* [1979] ECtHR 78103/14.
[52] *X and Y v the Netherlands* [1985] ECtHR 8978/80.
[53] *Ozgur Gundem v Turkey* [2000] ECtHR 20046/16, 21350/16, 26213/16, 51314/16, 54383/16, 57176/16, 58508 /16, 4630/17, 7268/17, 18590/19, 34713/19, 38209/19, 62293/19, 4853/20, 7245/20.
[54] *Dink v Turkey* [2010] ECtHR 2668/07, 6102/08, 30079/08, 7072/09, 7124/09.

Ismayilova v. Azerbaijan (2019), *Haji and Others v. Azerbaijan* (2010), *Gasi and Others v. Serbia* (2023), and *Tagiyeva v. Azerbaijan* (2019). However, the positive obligation under Article 10 has been limited to an important protective principle without a practical effect, and Article 10 is commonly relegated to a subsidiary position within ECHR.[55] Moreover, the ECtHR's massive precedents of allowing restrictions on hate speech also indicate that the ECtHR appears to have accorded diminished weight to Article 10 compared to its strong protection of "offend, shock or disturb" speech in *Handyside*.[56]

Recent ECtHR rulings on content moderation suggest a tendency to recalibrate Article 10 protections in favor of the state's positive obligations to protect other rights, especially under Article 8 (right to respect for private life). A seminal judgment is *Delfi AS v. Estonia* (2015). The ECtHR reached a key conclusion that secondary internet publishers may be held liable for defamatory user comments, even if they expressly state that the comments do not reflect their views and remove the comments promptly upon notification.[57] The ECtHR upheld the Estonian Supreme Court's decision to fine the platform, finding that the interference with freedom of expression under Article 10 was justified because the restriction was limited. [58] Moreover, the ECtHR also mentioned that because of the particular nature of the Internet which "provides an unprecedented platform for the exercise of freedom of expression", the "duties and responsibilities" of online platforms under Article 10 may differ from those of a traditional publisher.[59] After *Delfi*, further ECtHR precedents indicate that the Court supports the view that online platforms have a responsibility to implement a "notice-and-take-down" system and that the Convention States have a positive obligation to penalize platforms that fail to comply with

[55] Katie Pentney, 'States' Positive Obligation to Create a Favourable Environment for Participation in Public Deb ate: A Principle in Search of a Practical Effect?' (2024) 16 Journal of Media Law 146, 176.
[56] Jacob Mchangama and Natalie Alkiviadou, 'Hate Speech and the European Court of Human Rights: Whatever Happened to the Right to Offend, Shock or Disturb?' (2021) 21 Human Rights Law Review 1008, 1009.
[57] Neville Cox, 'Elfi v Stonia: The Liability of Secondary Internet Publishers for Violation of Reputational Right s under the European Convention on Human Rights' (2014) 77 The Modern Law Review 619, 619.
[58] *Delfi as v Estonia* [2015] ECtHR [GC] 64569/09.
[59] *ibid.*

this requirement.[60] In *Sanchez v. France* (2023), a politician was criminally convicted for failing to take prompt action to remove comments posted by others under one of his Facebook posts. The ECtHR observed that a minimum level of subsequent moderation or automatic filtering is desirable to identify unlawful comments as quickly as possible and to ensure their removal within a reasonable time, even in the absence of notification by the injured party.[61] This reasoning was reiterated in *Zöchling v. Austria* (2023), where the ECtHR criticized the Austrian courts for failing to assess whether a notice-and-take-down system could have been implemented, even though the comments in question had been deleted following the applicant's request. The Court found that the Austrian judiciary had failed to properly balance the applicant's right to private life against the platform's freedom of expression under Article 10, thereby violating the State's positive obligations under Article 8.[62]

The ECHR's jurisprudence reflects a hybrid model of free speech: while the right to freedom of expression remains fundamentally a defensive right against the state, the incorporation of the proportionality test and the doctrine of positive obligations has opened space for state interference. The positive obligations arising from other fundamental rights have created an opportunity to impose liability on platforms for their online moderation practices.

[60] Erik Tuchtfeld, 'Be Careful What You Wish For' [2023] Verfassungsblog 2 <https://verfassungsblog.de/be-careful-what-you-wish-for/> accessed 10 May 2025.
[61] *Sanchez v France* [2023] ECtHR [GC] 45581/15.
[62] 'Zochling v. Austria' <https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-226418%22]}> accessed 10 May 2025.

CHAPTER THREE: COMPARATIVE MODELS IN PRACTICE

1. Classical Liberalism Model

1.1. Section 230 and the Boundary of Platform Immunity

As discussed in the previous section, the United States has a strong inclination toward an almost absolute conception of free speech, rooted in the classical liberal foundations of its constitutional order. Based on that classical liberal understanding, the United States built an extraordinary immunity for social media platforms and other providers or users of an interactive computer service, which is Section 230 of the Communications Decency Act of 1996 (Section 230). Jeff Kosseff recognized Section 230 as "twenty-six words that created the internet":

"No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[63]

It further provides "Good Samaritan" protection from civil liability for operators of interactive computer service in Section 230(c)(2)(a):

"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."[64]

The origin of Section 230 traces back to *Stratton Oakmont, Inc. v. Prodigy Services Co.* (1995) (Stratton Oakmont). In this case, Prodigy, an early online content hosting site as well as the poster were charged with defamation. The plaintiff argued that Prodigy should be considered a

[63] Communications Decency Act § 230.
[64] ibid.

publisher of the defamatory material, however, the defendant cited the *Cubby, Inc. v. CompuServe Inc.* (1991), where CompuServe, another online service provider, was found not liable as a publisher for customer-generated content. The court held that Prodigy was liable because it exercised editorial control over the messages it hosted in three ways, which are analogous to what we now refer to as online content moderation: 1) by posting content guidelines for users; 2) by enforcing those guidelines with "Board Leaders"; and 3) by utilizing screening software designed to remove offensive language. Therefore, Prodigy's "conscious choice" has opened it up to a greater liability than CompuServe which makes no such choice, in other words, Prodigy was penalized because it moderated harmful content.[65] The *Stratton Oakmont* ruling set a bad precedent for the Internet industry, forcing companies to face the moderator's dilemma—either to try to moderate perfectly and risk liability or not moderate at all and be free from liability.[66] Ironically, this defamation judgment could make it easier for Internet users to spread defamatory and damaging bulletin board postings because if a company applies a content moderation policy, it could risk losing its status as a "distributor" and be held liable for every hateful utterance of millions of customers.[67] This concern was very important for Congress in 1995, which was drafting the Communications Decency Act of 1996 (CDA) to impose criminal penalties for the online transmission of indecent material to minors. Therefore, in order to strike a balance between protecting minors and preserving freedom of speech on the internet, Congress included Section 230 in the Communications Decency Act (CDA) to prevent more online service providers from facing the same legal pitfalls as Prodigy.[68]

[65] *Stratton Oakmont, Inc v Prodigy Services Co* [1995] NY Sup Ct No. 31063/94, 1995 WL 323710.
[66] Reese Bastian, 'Content Moderation Issues Online: Section 230 Is Not to Blame' (2022) 8 Texas A&M Journa l of Property Law 43, 50.
[67] Jeff Kosseff, *The Twenty-Six Words That Created the Internet* (Cornell University Press 2019) 56.
[68] Jeff Kosseff, 'What Was the Purpose of Section 230? That's a Tough Question Response' (2023) 103 Boston University Law Review 763, 770.

However, the CDA faced a lot of criticism. House Speaker Newt Gingrich denounced: "(the CDA) is a violation of free speech and it's a violation of the rights of adults to communicate with each other". Finally, the original version of CDA was challenged to be violating the First Amendment (freedom of speech) and the Fifth Amendment (due process) in *Reno v. ACLU* (1997).[69] The original CDA attempted to regulate both indecency (when available to children) and obscenity in cyberspace, criminalizing sending sexually explicit or patently offensive content to a specific minor and the transmission of "obscene or indecent" materials to persons known to be under 18. The court ruled that the vagueness of a content-based regulation of speech in the CDA creates an "obvious chilling effect of such regulation on free speech", which is a "governmental regulation which is more likely to interfere with the free exchange of ideas than to encourage such exchange", therefore the CDA violates the First Amendment. [70] However, Section 230 of the CDA was not struck down by the Supreme Court. Instead, the Court cited with approval Judge Stewart Dalzell's district court opinion, which emphasized the distinctive nature of Internet communication as "the most participatory form of mass speech yet developed," and argued that it, therefore, deserves "the highest protection from governmental intrusion".[71] The Supreme Court affirmed that the unique architecture of the Internet demands maximal constitutional protection for speech. Therefore, Section 230 became an incredibly strong shield for online intermediaries, which established the ground rule for lawsuits over internet content: a victim can rarely sue the service providers who host the customer-generated content or facilitate communications.[72]

Zeran v. America Online, Inc. (1997), was the first major case in which a court interpreted the newly enacted Section 230. It was later ranked by Eric Goldman as one of "the ten most

[69] Craig Bicknell, 'CDA: From Conception to Supreme Court' *Wired* <https://www.wired.com/1997/03/cda-from-conception-to-supreme-court/> accessed 13 May 2025.
[70] *Reno v Am Civil Liberties Union* (1997) 521 US 844.
[71] *ibid.*
[72] SUZOR (N 25) 45.

important Section 230 rulings" for its expansive construction of the statute's immunity provision.[73] This case involved a cyber-harassment attack on America Online's message boards, the victim sent America Online takedown notices, but it wasn't deleted immediately. The Fourth Circuit interpreted Section 230 expansively and rejected the argument that platforms should bear distributor liability, holding that Section 230 immunity applies regardless of whether the platform receives a takedown notice or has actual knowledge of the unlawful or false nature of third-party content.[74] The court further said that Section 230(c)(1) bars "lawsuits seeking to hold a service provider liable for its exercise of a publisher's traditional editorial functions—such as deciding whether to publish, withdraw, postpone or alter content."[75] This stands in stark contrast to the European Union's notice-and-take-down regime. Further, *Blumenthal v. Drudge* (1998) demonstrated how far courts were willing to go in granting Section 230 immunity. In this case, columnist Matt Drudge issued a false report on AOL, and the court noted that AOL "affirmatively promoted Drudge as a new source of unverified instant gossip".[76] Besides having the authority to edit and remove Drudge's submissions, AOL also paid Drudge $3,000 monthly in royalties for publishing on the website which is his sole consistent source of income.[77] The court found that AOL could not be liable for defamatory statements carried by AOL but written by gossip columnist Matt Drudge. The immunity provision applied even though AOL paid Drudge for the right to make his gossip column available to its subscribers and actively promoted the column as a benefit of subscription.[78]

[73] Eric Goldman, 'The Ten Most Important Section 230 Rulings' (2017) 20 Tulane Journal of Technology and In tellectual Property 1, 2.
[74] *Zeran v America Online, Inc* (1997) 129 F3d 327 (4th Cir); David S Ardia, 'Free Speech Savior or Shield for S coundrels: An Empirical Study of Intermediary Immunity under Section 230 of the Communications Decency Ac t' 465.
[75] *Zeran v. America Online, Inc.* (n 74).
[76] *Blumenthal v Drudge* (1998) 992 F Supp 44 (DDC).
[77] Joshua M Masur, 'A MOST UNCOMMON CARRIER: ONLINE SERVICE PROVIDER IMMUNITY AGAI NST DEFAMATION CLAIMS IN BLUMENTHAL v. DRUDGE' (2000) 40 Jurimetrics 217, 223; *Blumenthal v . Drudge* (n 76).
[78] Michelle J Kane, 'Blumenthal v. Drudge Part VI: Business Law: Section 1: Electronic Commerce: B) Internet Service Provider Liability' (1999) 14 Berkeley Technology Law Journal 483, 483.

Subsequent cases have established the "material contribution test" as an exception to Section 230 immunity, drawing the line at "the crucial distinction between, on the one hand, taking actions (traditional to publishers) that are necessary to the display of unwelcome and actionable content and, on the other hand, responsibility for what makes the displayed content illegal or actionable".[79] As the Ninth Circuit held in *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC* (2008), the Roommates.com designed question lists which "required users to input illegal content", thus it is "a website helps to develop unlawful content, and thus falls outside of Section 230 if it contributes materially to the alleged illegality of the conduct".[80] The material contribution test requires more than merely hosting or publishing third-party speech, it demands a showing that the platform had a hand in shaping the content's illegal character. The Tenth Circuit adopted this reasoning in *FTC v. Accusearch Inc.* (2009), which involved a website that sold information contained in telephone records. The court considered whether the operator could claim Section 230 protection from an FTC enforcement action when the records had been acquired from third parties. It ultimately denied immunity to the company, reasoning that it had "knowingly sought to transform virtually unknown information into a publicly available commodity" and had thereby "specifically encouraged the development of what was offensive about the content".[81] By contrast, courts have emphasized that passive facilitation alone does not constitute a material contribution. In *Jones v. Dirty World Entertainment Recordings LLC* (2014), the plaintiff filed a civil suit in federal court after anonymous users posted photos and allegedly defamatory comments about her and the site's manager added his remarks. Despite over twenty-seven requests from the plaintiff to remove the posts, the manager refused. As applied to the case in hand, the court held that the Defendants did not materially contribute to the defamatory statements against the Plaintiff "simply because

[79] *Kimzey v Yelp!* (2016) 836 F3d 1263 (9th Cir).
[80] *Fair Hous Council of San Fernando Valley v Roommates.com, LLC* (2008) 521 F3d 1157 (9th Cir).
[81] *FTC v Accusearch Inc* [2009] 10th Cir, 570, 1187 F3d.

those posts were selected for publication". It also ruled that their decision not to remove the contents was not a material contribution based on Section 230, which bars liability against website operators for their exercise of traditional editorial functions, such as deciding whether to publish, withdraw, postpone, or alter content. The court reaffirmed that "a material contribution to the alleged illegality of the content does not mean merely taking action that is necessary to the display of allegedly illegal content. Rather, it means responsibility for what makes the displayed content allegedly unlawful".[82] Moreover, in *Force v. Facebook, Inc.* (2019), Facebook was alleged to unlawfully provided a Hamas communication platform that enabled deadly attacks on U.S. citizens and to use its algorithms to "disseminate Hamas' messages directly to its intended audiences".[83] The court did not find Facebook responsible for the content's development because it did not alter information that its users published, and because the algorithms were "content 'neutral'".[84] The court again reaffirmed that "merely arranging and displaying others' content to users of Facebook through such algorithms" does not contribute to a material contribution, and that such algorithmic recommendation is eligible for Section 230 immunity.[85]

Apart from intentional torts, Section 230 immunity also extends to bar negligence claims against platforms, even where they arguably facilitated foreseeable harm.[86] Early in *Doe v. America Online, Inc*. (2001), the court dealt with a negligence claim. The defendant conducted sexual exploitation of children and transportation of sexually explicit material involving a minor and marketed child pornography material on AOL. The plaintiff claimed that AOL breached an alleged duty to exercise reasonable care to ensure that its service was not to be used for the sale

[82] *Jones v Dirty World Entertainment Recordings LLC* (2014) 755 F3d 398 (6th Cir).
[83] *Force v Facebook, Inc* (2019) 934 F3d 53 (2d Cir).
[84] *ibid.*
[85] Max Del Real, 'Breaking Algorithmic Immunity: Why Section 230 Immunity May Not Extend to Recommend ation Algorithms' (2024) 99 Wash. L. Rev. Online 1, 23; *Force v. Facebook, Inc* (n 83).
[86] Michael Rustad and Thomas Koenig, 'The Case for a CDA Section 230 Notice-and-Takedown Duty' (2023) 2 3 Nevada Law Journal 533, 551.

and distribution of child pornography. However, the court reaffirmed *Zeran* and ruled that "holding AOL liable for harm caused by third-party messages on the theory that it 'knew or should have known standard' would discourage it from efforts to screen or block".[87] In *Green v. America Online, Inc.* (2003), the Third Circuit held that AOL could not be held liable for negligently failing to address harmful content, as liability for actions related to "monitoring, screening, and deletion of content from its network" is "specifically proscribed" by Section 230.[88] Similarly, in *Doe v. MySpace* (2008), the court cited *Green* and dismissed claims that the platform bears negligence liability due to its failure to protect a minor user from sexual assault and lack of any safe precautions.[89] In *Daniel v. Armslist* (2019), the court held that a website facilitating firearm sales including a transaction in which a prohibited purchaser later committed murder was shielded by Section 230 from liability. The court cited again *Roomates.com*, stressing that "Section 230 must be interpreted to protect websites not merely from ultimate liability, but from having to fight costly and protracted legal battles."[90]

Section 230 even shields platforms from court orders to take down user-generated content. A court cannot directly order a platform to remove such content because the platform can't be a direct defendant due to Section 230, and a court order in any other lawsuit cannot reach the platform due to Federal Rule of Civil Procedure 65(d)(2).[91] In *Barnes v. Yahoo!, Inc.* (2009), the plaintiff alleged Yahoo!, Inc.'s failure to remove offensive content about the plaintiff posted by a third party, but Section 230(c) immunizes it.[92] In *Blockowicz v. Williams* (2010), the plaintiffs had secured a defamation judgment and injunction against individual users who posted on a forum hosted by www.ripoffreport.com, but they were unable to compel the website

[87] *Doe v America Online, Inc* (2001) 783 So 2d 1010 (Fla).
[88] *Green v American Online, Inc* (2003) 318 F3d 465 (3d Cir).
[89] *Doe v MySpace (2008)* (2008) 528 F3d 413 (5th Cir).
[90] *Daniel v Armslist, LLC* (2019) 926 Wis 710 (NW2d).
[91] Eric Goldman, 'An Overview of the United States' Section 230 Internet Immunity' in Giancarlo Frosio (ed), *Oxford Handbook of Online Intermediary Liability* (Oxford University Press 2020) 170 <https://doi.org/10.1093/oxfordhb/9780198837138.013.8> accessed 21 May 2025.
[92] *Barnes v Yahoo!, Inc* (2009) 570 F3d 1096 (9th Cir).

to remove the defamatory posts.[93] Despite the court's acknowledgment that the content was unlawful, the website—being a third party that did not act in "active concert or participation" with the defendants—was not subject to the injunction.[94] This principle was reaffirmed in *Hassell v. Bird* (2018), where the California Supreme Court similarly denied the enforceability of a takedown injunction against Yelp.[95] The court ruled that Section 230 prohibits platforms from being treated as publishers and that forcing a website to remove defamatory content creates it as such. The court described the plaintiff's decision not to name Yelp as a defendant as "tactical," stating that the plaintiff was attempting to "accomplish indirectly what Congress has n them to achieve directly."[96] These cases reflect 230 legal regimes in which private platforms possess quasi-sovereign discretion over content removal, immune not only from tort claims but also from lawful judicial orders.

Nonetheless, Section 230 does not provide absolute immunity. Section 230(3) says that Section 230 will not apply to: (1) federal criminal laws; (2) intellectual property laws; (3) any state law that is "consistent with" Section 230; (4) communication privacy law; and (5) certain civil actions or state prosecutions where the underlying conduct violates specific federal laws prohibiting sex trafficking.[97] The last exception was added by Allow States and Victims to Fight Online Sex Trafficking Act of 2017 (FOSTA) in 2018 which also created a new federal crime for anyone who "owns, manages, or operates an interactive computer service" with the intent to promote or facilitate prostitution.[98] The launching of FOSTA is largely the Congress' response to *Doe v. Backpage.com, LLC* (2016).[99] In this case, an online advertising provider

[93] *Blockowicz v Williams* (2010) 630 F3d 563 (7th Cir); Connor Moran, 'Injunction Relief: Must Nonparty Websi tes Obey Orders to Remove User Content' (2011) 7 Washington Journal of Law, Technology & Arts 47, 50.
[94] Moran (n 93) 52.
[95] *Hassell v Bird* (2018) 5 Cal5th 522 (Cal).
[96] *ibid.*
[97] Valerie C Brannon, 'Liability for Content Hosts: An Overview of the Communication Decency Act's Section 2 30' 3.
[98] Eric Goldman, 'The Complicated Story of Fosta and Section 230' (2018) 17 First Amendment Law Review 279, 284; Fight Online Sex Trafficking Act of 2017 (FOSTA) 2018 (Pub L No 115–164).
[99] Goldman, 'The Complicated Story of Fosta and Section 230' (n 98) 282.

was protected under Section 230 from liability to a minor victim of sex trafficking, as it performed "traditional publisher functions" rather than creating or developing sex trafficking advertisements, even though the website profited from the sale of such ads.[100] However, many scholars suggest that FOSTA did not achieve its legislative purpose of combatting sex trafficking and has a chilling effect on free speech.[101] In recent cases regarding FOSTA and Section 230, courts have consistently applied the exception narrowly, making it difficult for plaintiffs to succeed. In *J.B. v. G6 Hosp., LLC* (2020), *Doe v. Reddit, Inc.* (2021), and *Doe v. Salesforce, Inc.* (2025), the courts rejected the plaintiffs' claims for failing to prove that the defendants had actual knowledge of, and actively benefitted from, individual instances of trafficking, and then applied the material contribution test to extend Section 230 immunity to the defendants.[102] In *Woodhull Freedom Found. v. United States* (2023), the plaintiffs challenged FOSTA's constitutionality under the First and Fifth Amendments, arguing that the statute is unconstitutionally vague and that its Section 230 provision strikes at the heart of free speech.[103] Although the court upheld FOSTA's constitutionality, it interpreted the statute narrowly and imposed a high threshold of "actual knowledge," making it more difficult for plaintiffs to bring successful claims under FOSTA and thereby overcome Section 230's liability shield.[104]

[100] *Doe v Backpage.com, LLC* (2018) 881 F3d 167 (F3d1st Cir).
[101] Ronni Vogelsang, 'The Failure of FOSTA: Unintended Consequences Outweigh Good Intentions' (2023) 44 University of La Verne Law Review 59, 89; Kendra Albert and others, 'Fosta in Legal Context' (2020) 52 Columbia Human Rights Law Review 1084, 1102.
[102] *Doe v Salesforce, Inc* (2025) 2025 US Dist LEXIS 52346 (ND Tex); *Doe v Reddit, Inc* (2021) 2021 US Dist LEXIS 235993 (CD Cal); *JB v G6 Hospitality, LLC* (2020) 2020 US Dist LEXIS 151213 (ND Cal).
[103] *Woodhull Freedom Found v United States* (2023) 72 F4th 1286 (DC Cir).
[104] Eric Goldman, 'DC Circuit Upholds FOSTA's Constitutionality (By Narrowing It)-Woodhull v. U.S.' (*Technology & Marketing Law Blog*, 23 July 2023) <https://blog.ericgoldman.org/archives/2023/07/dc-circuit-upholds-fostas-constitutionality-by-narrowing-it-woodhull-v-u-s.htm> accessed 23 May 2025.

1.2. Risks

Owing to classical liberalism's idea of freedom of speech, Section 230 has shielded the internet industry from various legal risks in its twenty-nine years of history. The twin goals of Section 230, "promoting the continued development of the Internet and other interactive computer services" and "preserving the vibrant and competitive free market that presently exists for the Internet and other interactive computer services unfettered by Federal or State regulation" have been achieved.[105] As Anupam Chander wrote, Section 230 "proved central to the rise of the new breed of Silicon Valley enterprise":

"What risks did such firms face? By offering platforms for users across the world, Internet enterprises faced the hazard that some users would use these platforms in ways that violated the law, bringing with it the possibility of liability for aiding and abetting that illegal activity. Consider a sampling of the array of claims that might lie against these platforms for the behavior of their users. Yahoo might be liable if someone uses Yahoo Finance to circulate a false rumor about a public company. Match.com could face liability if a conniving user posted defamatory information about another individual. Craigslist might be liable under fair housing statutes if a landlord put up a listing stating that he preferred to rent to people of a particular race. Amazon and Yelp might be liable for defamatory comments written by a few of their legions of reviewers."[106]

However, the emergence of big technology companies and social media platforms also poses a risk to Section 230 and the values it upholds. The technology companies are vastly larger, more powerful, and less vulnerable than were the nascent "online service providers" of two decades ago. The online environment also experienced profound change since 1996. There were only

[105] *Hassell v. Bird* (n 95).
[106] Anupam Chander, 'How Law Made Silicon Valley' (2013) 63 Emory Law Journal 639, 650.

around 44 million internet users 30 years ago, but today that number has surged to over 330 million—an increase of more than 6.5 times.[107] This shift was already acknowledged nearly two decades ago in *Roommates.com*, where the Ninth Circuit noted that "the Internet is no longer a fragile new means of communication that could easily be smothered in the cradle by overzealous enforcement of laws and regulations applicable to brick-and-mortar businesses," and further observed that "the Internet has outgrown its swaddling clothes and no longer needs to be so gently coddled".[108] Even though the court has reminded that there was a need to "be careful not to exceed the scope of the immunity provided by Congress", the activities of those companies that Section 230 immunizes from liability, is immensely different and less obviously about free speech that Congress sought to protect.[109] Online speech moderation is one crucial area where Section 230 provides massive legal freedom to technology companies.[110] However, this *laissez-faire* immunity has led to significant challenges including under-, over-, and biased moderation.

1.2.1. Under-Moderation

In the development of case law regarding Section 230, there have been hundreds of cases expanding the protection of Section 230 from intermediary liability for under-moderation and relatively very few minimizing it. This has allowed social media platforms now enjoy broad immunity from intermediary liability even when they have encouraged posting illegal content, benefited from hosting illegal content, or enabled illegal activity by the way of their policies and website design.[111] As Danielle Keats Citron and Benjamin Wittes said: "blanket immunity

[107] 'Internet Users - United States - Telecommunications' <https://www.indexmundi.com/united_states/internet-users.html?utm_source=chatgpt.com> accessed 28 May 2025.
[108] *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC* (n 80).
[109] Danielle Citron and Benjamin Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity' (2017) 86 Fordham Law Review 401, 411; *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC* (n 80).
[110] Eric Goldman, 'Content Moderation Remedies' (2021) 28 Michigan Technology Law Review 1, 5.
[111] Reese Bastian, 'Content Moderation Issues Online: Section 230 Is Not to Blame' (2022) 8 Texas A&M Journal of Property Law 43, 54.

gives platforms a license to solicit illegal activity…Site operators have no reason to take down material that is defamatory or invasive of privacy. They have no incentive to respond to clear instances of criminality or tortious behavior".[112]

With the exponential expansion of the internet's aggregative capacity, the harms inflicted by online defamatory, privacy-invasive, or threatening content have become profoundly more severe, often inflicting long-lasting damage on individual victims' reputations, security, and psychological well-being. A review of recent cases in which Section 230 has shielded platforms from liability for under-moderation reveals substantial harm suffered by victims:

In *Herrick v. Grindr, LLC* (2019), the plaintiff filed a lawsuit against Grindr after his ex-boyfriend used the platform to impersonate him and direct over 1,400 men to his home and workplace over several months. Herrick alleged that despite reporting the abuse to Grindr over 100 times, the company failed to take any meaningful action, instead responding with automated messages and allowing the fake profiles to persist.[113] In *Doe v. Mindgeek USA Inc.* (2021), Pornhub hosted child pornography such as children being raped and assaulted, and offered playlists and tags including phrases such as "less than 18", and "young boys".[114] In *M.H. v. Omegle.com, LLC* (2024), an 11-year-old girl was matched with a predator who exploited children into performing sexual acts over live web feeds while recording.[115] In *Doe v. Grindr Inc.* (2025), a 15-year-old boy was with four adult men and raped by them on consecutive days.[116] And the Ninth Circuit said that Grindr's statement that it provides a "safe and secure environment for its users" was just "a description of its moderation policy".[117]

[112] Citron and Wittes (n 109) 414.
[113] *Herrick v Grindr, LLC* (2019) 765 F App'x 586; Kira Geary, 'Section 230 of the Communications Decency A ct, Product Liability, and a Proposal for Preventing Dating-App Harassment' (2021) 125 Penn State Law Review 503 <https://elibrary.law.psu.edu/pslr/vol125/iss2/4>.
[114] *Doe v MindGeek USA Inc* (2021) 558 F Supp 3d 828 (CD Cal).
[115] *MH v Omegle.com, LLC* (2024) 122 F 4th 1266 (11th Cir).
[116] *Doe v Grindr Inc* (2025) 128 F4th 1148 (9th Cir).
[117] *ibid.*

1.2.2. Over-Moderation and Biased Moderation

Around the world technology companies are always enlisted by governments to aid in efforts to stem the tide of nasty content, to protect privacy, and to police propaganda and misinformation. Technology companies frequently cooperate with governments to remove unwanted speech.[118] For example, YouTube has complied with requests from Thai authorities to create technical measures to block videos that allegedly insult the monarchy, according to the country's *lèse-majesté* laws.[119] In 2006, Google launched a censored search engine in China, complying with government demands to block content that might stoke democratic resistance.[120] In March 2023, Twitter began enforcing country-specific content restrictions in Turkey, removing access to posts that criticized the government's inadequate response to the devastating earthquakes.[121] To mitigate risks and protect the community, content moderation is now an essential component of social media platforms' Terms of Service.[122]

Section 230 allows platforms to block or remove third-party material "whether or not such material is constitutionally protected". Platforms may choose to over-moderate by censoring content for a variety of reasons, and this leads to the loss of online freedom of speech.[123] In Eric Goldman's research, he collected and analyzed 62 cases regarding content removal or account termination decisions, and found that "Internet services have won essentially all of the lawsuits to date brought by terminated or removed users".[124] In these cases, Section 230, the state action

[118] Hannah Bloch-Wehba, 'Global Platform Governance: Private Power in the Shadow of the State' (2019) 72 SMU Law Review 27, 29.
[119] Ryan Singel, 'YouTube Agrees To Help Government Censors' *Wired* <https://www.wired.com/2007/04/youtube-agrees-/> accessed 28 May 2025; Suzor (n 25) 29.
[120] WIRED Staff, 'Google Bends to China's Will' *Wired* <https://www.wired.com/2006/01/google-bends-to-chinas-will/> accessed 28 May 2025.
[121] 'Turkey: Freedom on the Net 2023 Country Report' (*Freedom House*) <https://freedomhouse.org/country/turkey/freedom-net/2023> accessed 28 May 2025.
[122] Bloch-Wehba (n 118) 29.
[123] Bastian (n 111) 52.
[124] Eric Goldman and Jess Miers, 'Online Account Terminations/Content Removals and the Benefits of Internet Services Enforcing Their House Rules' (Social Science Research Network, 1 August 2021) 192 <https://papers.ssrn.com/abstract=3911509> accessed 28 May 2025.

doctrine of the First Amendment, and contractual arguments acted as strong defenses, thereby reinforcing the platforms' broad discretion in moderation.

Another significant challenge posed by the broad immunity granted under Section 230 is the risk of biased moderation, whereby certain viewpoints are disproportionately removed—an outcome that undermines the integrity of public discourse. Allegations of such bias in content moderation have been raised by both liberal and conservative groups across various social media platforms.[125]

The possibility of social media platforms conducting biased online moderation has made state governments enforce must-carry laws. Two landmark cases, *NetChoice, LLC v. Paxton* (2022) and *Moody v. NetChoice, LLC* (2022), both currently before the U.S. Supreme Court, center on the constitutionality of such must-carry laws enacted by Texas and Florida, respectively.[126] These laws aim to regulate "social media platforms" and restrict them from "censoring" or otherwise disfavoring posts—including deleting, altering, labeling, or deprioritizing them—based on their content or source.[127] In *Paxton*, the Fifth Circuit upheld Texas' must-carry law, rejecting the argument that social media platforms possess a First Amendment right to exercise editorial discretion over user content because platforms do not act as traditional speakers when they host user-generated. The court further argued that Texas' must-carry law "did not chill speech but chilled censorship", and it "advanced an important governmental interest in protecting the free exchange of ideas and information".[128] In sharp contrast, the Eleventh

[125] Justin T Huang, Jangwon Choi and Yuqin Wan, 'Politically Biased Moderation Drives Echo Chamber Format ion: An Analysis of User-Driven Content Removals on Reddit' (Social Science Research Network, 17 October 2 024) <https://papers.ssrn.com/abstract=4990476> accessed 29 May 2025; Kari Paul, 'Meta Struggles with Moder ation in Hebrew, According to Ex-Employee and Internal Documents' *The Guardian* (15 August 2024) <https:// www.theguardian.com/technology/article/2024/aug/15/meta-content-moderation-hebrew> accessed 29 May 202 5; Kari Paul, 'Meta Struggles with Moderation in Hebrew, According to Ex-Employee and Internal Documents' *The Guardian* (15 August 2024) <https://www.theguardian.com/technology/article/2024/aug/15/meta-content-m oderation-hebrew> accessed 29 May 2025.
[126] *NetChoice, LLC v Paxton* (2022) 49 F4th 439 (5th Cir); *Moody v NetChoice, LLC* (2022) 34 F4th 1196 (11th Cir).
[127] *Moody v. NetChoice, LLC* (n 126).
[128] *NetChoice, LLC v. Paxton* (n 126).

Circuit in *Moody*, the court struck down key provisions of Florida's must-carry law, holding that the State's restrictions on content moderation trigger First Amendment scrutiny under this Court's cases protecting editorial discretion.[129]

Winning in a biased censorship case as the plaintiff is just as hard as in a removal or account termination decision case. State action doctrine and Section 230 can defend almost every challenge: bias against Muslims and Jews in *Elansari v. Meta Inc,* (2024) and *Klayman v. Zuckerberg* (2014); bias against conservative opinion in *Freedom Watch, Inc. v. Google, Inc.* (2019), *Freedom Watch v. Google* (2020) and *Prager University v. Google LLC* (2020); and bias against pro-democracy opinions in *Sun v. China Press, Inc.* (2022).[130] Although courts may assert that there is no definitive evidence of bias within social media platforms' moderation systems, the case of *Zhang v. Baidu.com Inc.* (2014) presents a notable exception. In this case, the bias was evident and enforced under directives from the Chinese government. Nevertheless, a U.S. federal court held that Baidu's decision to de-index pro-democracy political content was protected under the First Amendment as an exercise of editorial discretion.[131]

2. Notice-and-Takedown Model

Compared to the United States' nearly absolute Section 230 immunity, other jurisdictions choose a more balanced approach to make sure legal accountability doesn't stifle technological progress.

[129] *Moody v. NetChoice, LLC* (n 126).
[130] *Sun v China Press, Inc* (2022) 2022 WL 1718792 (ND Cal); *Prager Univ v Google LLC* (2020) 951 F3d 991 (9th Cir); *Republican Nat'l Comm v Google LLC* (2023) 2023 WL 5731496 (ND Cal); *Freedom Watch, Inc v Goo gle, Inc* (2020) 816 F APP'x 497 (DC Cir); *Klayman v Zuckerberg* (2014) 753 F3d 1354 (DC Cir); *Elansari v Me ta Inc,* (2024) 2022 WL 16948246 (3d Cir).
[131] *Zhang v Baidu.com Inc* (2014) 10 F Supp 3d 433 (SDNY); Eric Goldman, 'Of Course The First Amendment P rotects Baidu's Search Engine, Even When It Censors Pro-Democracy Results (Forbes Cross-Post)' (*Technology & Marketing Law Blog*, 10 April 2014) <https://blog.ericgoldman.org/archives/2014/04/of-course-the-first-amen dment-protects-baidus-search-engine-even-when-it-censors-pro-democracy-results-forbes-cross-post.htm> acces sed 29 May 2025.

2.1. From the E-Commerce Directive to the Digital Services Act

2.1.1. E-Commerce Directive: Articles 14-15

The EU built the safe harbor for internet intermediaries with the E-Commerce Directive (2000, ECD), which exempts intermediaries providing mere conduit, caching, and hosting services from secondary liability under certain conditions in Article 14.[132] In particular, host providers are only exempted as long as they do not know that they are hosting illegal content or activities. Intermediaries must terminate or prevent illegalities when ordered by competent authorities, but cannot be subject to general obligations to monitor and seek information as prohibited in Article 15.[133] Thus, the ECD set out the principle of knowledge-and-take-down and the principle of no-monitoring obligation, under which hosting providers are only liable if they have actual knowledge of the illegal activity and fail to expeditiously remove or disable access to the content.[134]

The Court of Justice of the European Union (CJEU) has interpreted the principle of "knowledge-and-take-down" through a series of key cases. In *Google France SARL and Google Inc. v. Louis Vuitton Malletier SA* (2010), the Court held that Internet service providers may claim immunity only if they are "of a mere technical, automatic and passive nature", which implies that that service provider "has neither knowledge of nor control over the information which is transmitted or stored".[135] Moreover, the CJEU noticed that the mere fact that the service is paid, that Google sets the payment terms, or provides general information to customers does not, by itself, remove Google's liability exemption under ECD In contrast, in *L'Oréal SA v. eBay International* AG (2011), the Court emphasized that eBay, in the context

[132] Directive 2000/31/EC on Electronic Commerce (E-Commerce Directive).
[133] Dr Giovanni Sartor, 'Providers Liability: From the eCommerce Directive to the Future' 16.
[134] Giancarlo Frosio and Christophe Geiger, 'Taking Fundamental Rights Seriously in the Digital Services Act's Platform Liability Regime' (2023) 29 European Law Journal 31, 38.
[135] Google France SARL and Google Inc. v. Louis Vuitton Malletier SA (ECJ 2010).

of sponsored links leading to infringing listings, not only stored the content but also exercised control over its promotion and presentation, therefore has actual knowledge, and constitutes an "active role".[136] However, a similar allegation against eBay in the United States—that it facilitated and advertised the sale of counterfeit "Tiffany" goods was dismissed on the grounds of Section 230 immunity.[137] Moreover, In *Sotiris Papasavvas v O Fileleftheros Dimosia Etaireia Ltd and Others* (2014), the CJEU ruled that an online newspaper publishing company can't enjoy immunity under Article 14 because "it has knowledge of the information posted and exercises control over that information" and "remunerated by income generated by commercial advertisements posted on that website".[138] Similarly, in *Delfi* online newspaper site, was liable. This sharply contrasts with the immunity granted under Section 230, where online news platforms such as AOL, blogs, or gossip sites enjoy near-absolute protection, as seen in cases like *Blumenthal*, *Zeran*, and *Batzel*.

Article 15 of the ECD acts as a structural complement to Article 14, reinforcing the protection of free speech online, and prohibits Member States from imposing general obligations to monitor hosted content. This restriction has been interpreted by the CJEU in cases such as *SABAM v. Netlog NV* (201and *Frank Peterson v. Google LLC and, YouTube LLC* (2021), to preclude *ex-ante* content filtering systems. However, Article 15 doesn't prohibit the platform itself from filtering or moderating.[139]

In contrast to Section 230's immunity from the enforcement of court orders, the CJEU in *Eva Glawischnig-Piesczek v. Facebook Ireland Limited* (2019), affirmed that Article 15 does not preclude Member States from ordering a hosting provider to remove or block access to

[136] *L'Oréal SA and Others v eBay International AG and Others* (ECJ).
[137] *Tiffany (NJ) Inc v eBay Inc* (2010) 600 F3d 93 (2d Cir).
[138] *Sotiris Papasavvas v O Fileleftheros Dimosia Etaireia Ltd and Others* [2014] ECJ Case C-291/13.
[139] *Frank Peterson v Google LLC and YouTube LLC* [2021] CJEU C-682/18, ECLI:EU:C:2021:503 ECLI; *Belgis che Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV* [2012] CJEU C-360/10, E CLI:EU:C:2012:85 ECLI.

information that is "identical or equivalent to" content previously declared unlawful including on a global scale.[140]

2.1.2. Digital Services Act

The Digital Services Act (2022, DSA) builds upon and modernizes the core principles of the ECD.[141] It establishes an updated regulatory framework that maintains the conditional exemption from liability for providers of intermediary services, while introducing enhanced due diligence obligations, particularly for very large online platforms (VLOPs) and search engines (VLOSEs).

While DSA reaffirmed conditional liability immunity based on the "knowledge-and-take-down" principle in Article 6, the introduction of Article 16 institutionalizes a "notice and action" mechanism that effectively lowers the threshold for "knowledge": "Notices referred to in this Article shall be considered to give rise to actual knowledge or awareness for Article 6". In doing so, the DSA shifts the operational standard of platform liability closer to a "notice-and-takedown" model.[142] In this model, any individual or entity (Article 16) and "trusted flaggers" including private, non-governmental entities, or public entities (Article 22) can notify the platform. Moreover, Article 9 allows Member States' national judicial or administrative authorities to issue orders to act against illegal content, failing to obey would result in a loss of conditional immunity.

In terms of the obligations of platforms to take action concerning content posted by users, the DSA mentions three types of content:

[140] *Eva Glawischnig-Piesczek v Facebook Ireland Limited* [2019] CJEU C-18/18, ECLI:EU:C:2019:821 ECLI.
[141] Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act).
[142] Dawn Nunziato, 'The Digital Services Act and the Brussels Effect on Platform Content Moderation' (2023) 2 4 Chicago Journal of International Law 3 <https://chicagounbound.uchicago.edu/cjil/vol24/iss1/6>.

(a) Illegal content (Article 3(h)), any information that, in itself or relation to an activity, is not in compliance with Union law or the law of any Member State which complies with Union law;

(b) Content that is incompatible with the terms and conditions (Article 3(t));

(c) Legal but harmful content.[143]

With regard to illegal content, as discussed above, platforms are under a positive obligation to implement moderation measures to remove or disable access in compliance with the law. However, for content that is not illegal but violates a platform's terms and conditions, platforms are free to decide what they consider to be harmful in their terms and conditions and may moderate such content at their discretion. This wide discretion carries the risk that platforms may act arbitrarily, which could lead to unjustified restrictions on users' freedom of speech. To reduce this risk, Article 14 of the DSA provides two safeguards. First, platforms must clearly explain their content moderation rules and procedures in their terms and conditions. This information must be publicly available, easy to access, and written in clear language (Article 14(1)). Second, platforms must act with care, fairness, and balance when moderating content, and take into account the fundamental rights of users (Article 14(4)).[144] In addition, Article 15 requires platforms to publish regular transparency reports showing how and why they remove or restrict content.

Though DSA does not ban harmful but legal content, it places special obligations on VLOPs and VLOSEs to address the risks that such content may pose. According to Article 34, these platforms must assess systemic risks annually. This includes examining how their content moderation systems, recommender algorithms, and advertising practices may contribute to the

[143] Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act).
[144] 'User Content Moderation under the Digital Services Act – 10 Key Takeaways – Legal Developments' <https://www.legal500.com/developments/thought-leadership/user-content-moderation-under-the-digital-services-act-10-key-takeaways/> accessed 3 June 2025.

spread of both illegal and harmful (but legal) content. If such risks are found, Article 35 requires them to take appropriate and proportionate measures to reduce the risks. In addition, Article 37 obliges these platforms to undergo independent audits to evaluate how well they manage these risks. This mechanism has been praised by supporters as contributing to the development of Internet constitutionalism, "signaling a shift from private ordering solutions rooted in a liberal economic framework to a democratic model focused on the protection of fundamental rights"[145]

Failure to comply with SDA will result in severe sanctions. Under Article 52, the DSA authorizes fines of up to 6% of a platform's global annual turnover for intentional or negligent violations, alongside daily penalties of up to 5% of daily turnover to compel compliance, and, in extreme cases, the suspension of services within the EU under Article 74.[146]

The EU so far has launched investigations under the DSA into companies including X, AliExpress, Meta's Facebook and Instagram, and TikTok over problems like insufficient consumer protection, disinformation, and addictive algorithms.[147]

[145] Giancarlo Frosio, 'From the E-Commerce Directive to the Digital Services Act' (SSRN, 2024) <https://www.ssrn.com/abstract=4914816> accessed 31 May 2025; Giovanni De Gregorio, 'Digital Constitutionalism across the Atlantic' (2022) 11 Global Constitutionalism 297, 298; Giancarlo Frosio and Christophe Geiger, 'Towards a Digital Constitution' [2024] Verfassungsblog <https://verfassungsblog.de/towards-a-digital-constitution/> accessed 3 June 2025.

[146] Directive 2000/31/EC on Electronic Commerce (E-Commerce Directive).

[147] Adam Satariano, 'E.U. Prepares Major Penalties Against Elon Musk's X' *The New York Times* (3 April 2025) <https://www.nytimes.com/2025/04/03/technology/eu-penalties-x-elon-musk.html> accessed 4 June 2025; 'Commission Preliminarily Finds TikTok's Ad Repository in Breach of the Digital Services Act | Shaping Europe's Digital Future' <https://digital-strategy.ec.europa.eu/en/news/commission-preliminarily-finds-tiktoks-ad-repository-breach-digital-services-act> accessed 4 June 2025; 'Commission Opens Formal Proceedings against Facebook and Instagram under the Digital Services Act | Shaping Europe's Digital Future' <https://digital-strategy.ec.europa.eu/en/news/commission-opens-formal-proceedings-against-facebook-and-instagram-under-digital-services-act> accessed 4 June 2025; 'Commission Opens Formal Proceedings against AliExpress under the Digital Services Act | Shaping Europe's Digital Future' <https://digital-strategy.ec.europa.eu/en/news/commission-opens-formal-proceedings-against-aliexpress-under-digital-services-act> accessed 4 June 2025; 'DSA: Commission Opens Formal Proceedings against Meta' (*European Commission - European Commission*) <https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2664> accessed 4 June 2025.

2.2. National Implementations

There have been national laws applying the "notice-and-takedown" model before and after the DSA entered into force.

2.2.1. Germany: NetzDG

Germany's NetzDG (Network Enforcement Act), enacted in 2017, is one of the earliest national laws targeting illegal online content on social media platforms. It applies to large platforms with more than two million users in Germany and requires them to remove "clearly illegal" content within 24 hours of receiving a user complaint, or within seven days if legal analysis is needed. The law covers a range of criminal offenses under the German Criminal Code, including hate speech (§130 StGB), defamation (§§185–187), incitement to violence, threats (§241), and Holocaust denial.[148] Platforms must also provide an accessible reporting system for users, and publish biannual transparency reports. Failure to comply with these obligations may result in fines of up to €50 million.

In 2019, Facebook was fined €2 million for underreporting illegal content complaints in Germany and inaccessible reporting forms for users.[149] In 2022, Telegram was also fined €5 million for failing to develop tools for reporting illegal content and setting up a legal entity in Germany for interactions with the government.[150]

[148] Netzwerkdurchsetzungsgesetz (NetzDG).
[149] 'Facebook Issued with 2 Million Euro Fine' (*Clifford Chance*) <https://www.cliffordchance.com/content/cliffordchance/insights/resources/blogs/talking-tech/en/articles/2019/07/facebook-issued-with-2-million-euro-fine.html> accessed 4 June 2025.
[150] 'Tech & Terrorism: Germany Fines Telegram For Failing To Comply With Online Content Moderation Law' (*Counter Extremism Project*) <https://www.counterextremism.com/press/tech-terrorism-germany-fines-telegram-failing-comply-online-content-moderation-law> accessed 4 June 2025.

2.2.2. France: From LCEN to SREN law

France's first digital economy law, Law No. 2004-575 of 21 June 2004 on Confidence in the Digital Economy (LCEN), which mirrored the language of ECD, established conditional immunity to platforms.[151]

After 2015 the Charlie Hebdo and Hyper Cacher terrorist attacks, the French Ministry of the Interior gained motivation to publish a new decree about online moderation which has been under consideration since 2011.[152] This is Decree No. 2015-125 of 5 February 2015. The 2015 Decree amended Article 6 of LCEN, established a mechanism for administrative blocking of online content without prior judicial review.[153] It authorizes the Minister of the Interior to compile a confidential blacklist of websites or webpages including social media pages that are deemed to "provoke or praise acts of terrorism" or include child pornography. Internet service providers are required to block access to these listed addresses within 24 hours of notification.[154]

Inspired by NetzDG, the French Avia Law (2020) marked a bolder regulatory experiment. Avia Law took a similar "notice-and-action" system by which any user can flag "manifestly illegal" content (among a long pre-set list of offenses including hate speech, sexual harassment, holocaust denial, and revenge porn) and the notified online service provider is required to remove it within 24 hours. Avia Law also took a step further than the 2015 Decree, reducing

[151] Marco Lewis, 'The NetzDG and the Avia Law: How Two Different Legal Systems Created Two Different Outcomes from Similar Laws' (2022) 40 Wisconsin International Law Journal 491, 498; Loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique (LCEN) 2004 (JORF n°143 du 22 juin 2004).

[152] Amar Toor, 'France Can Now Block Suspected Terrorism Websites without a Court Order' (*The Verge*, 9 Feb ruary 2015) <https://www.theverge.com/2015/2/9/8003907/france-terrorist-child-pornography-website-law-cens orship> accessed 7 June 2025.

[153] Décret n° 2015-125 du 5 février 2015 relatif au blocage des sites provoquant des actes de terrorisme ou en faisant l'apologie 2015 (JORF n°0031 du 6 février 2015).

[154] 'Décret N° 2015-125 Du 5 Février 2015 Relatif Au Blocage Des Sites Provoquant à Des Actes de Terrorisme Ou En Faisant l'apologie et Des Sites Diffusant Des Images et Représentations de Mineurs à Caractère Pornogra phique - Légifrance' <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000030195477> accessed 7 June 2025; Amanda Goodman, 'Blocking Pro-Terrorist Websites: A Balance between Individual Liberty and National Secu rity in France' (2016) 22 Southwestern Journal of International Law 209, 229.

the intermediary's deadline to remove terrorist and child pornography content to 1 hour after the receipt of a notification by an administrative authority. Similar to the NetzDG, the Avia Law increased monetary penalties against companies for failing to follow the provisions in the law, with a maximum fine of €250,000.[155] Unlike the NetzDG, the Avia Law was quickly struck down by the French Constitutional Council on 18 June 2020 (Decision n° 2020-801 DC). The Council stated that the Avia Law "infringed upon the freedom of expression and communication in a way that is not appropriate, necessary, or proportionate to the aim pursued".[156] Moreover, the Council stressed that only judicial authorities can impose restrictions that severely affect freedom of expression in such a short time frame, but not restlessly on administrative discretion without judicial oversight.[157] The divergent attitudes of the French and German courts on similar legislation have been thoroughly examined in Marco Lewis's work.[158]

The French *Loi sur la sécurisation et la régulation de l'espace numérique* (SREN Law), enacted in 2024, represents France's most recent attempt to assert national control over digital platforms.[159] While the law aligns with the DSA in certain respects such as transparency and notice-and-takedown mechanism, it goes further by imposing administrative enforcement powers on France's national regulator (ARCOM), including the authority to oversee non-French platforms operating in France.

[155] 'Proposition de loi, n° 1785' <https://www.assemblee-nationale.fr/dyn/15/textes/l15b1785_proposition-loi> accessed 7 June 2025.
[156] 'Décision n° 2020-801 DC du 18 juin 2020 | Conseil constitutionnel' <https://www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm> accessed 7 June 2025.
[157] *ibid.*
[158] Lewis (n 151).
[159] Loi n° 2024-449 du 21 mai 2024 visant à sécuriser et réguler l'espace numérique (loi SREN) 2024 (JORF n°0118 du 22 mai 2024).

2.2.3. UK: Online Safety Act

The UK's Online Safety Act (OSA, 2023) moves beyond a simple notice-and-takedown model, introducing a proactive duty of care. With penalties of up to 10% global turnover, senior liability, and potential service disruptions, the OSA marks one of the world's most stringent internet safety regimes.[160]

In the initial draft of the OSA, a controversial issue emerged regarding the duty of care to include content categorized as "legal but harmful to adults". This was criticized as a risk a regulatory slippery slope toward wider censorship and was deleted in the final draft.[161]

2.3. Risks

2.3.1. The Chilling Effect

The severe penalties associated with the notice-and-takedown mechanism act as a *sword of Damocles* hanging over social media platforms.

Although most notice-and-takedown regimes include similar provisions, Recital 28 of the Digital Services Act explicitly states that "nothing in this Regulation should be construed as an imposition of a general monitoring obligation or a general active fact-finding obligation, or as a general obligation for providers to take proactive measures about illegal content".[162] Nevertheless, the structure of the notice-and-takedown mechanism itself creates strong incentives for platforms to engage in *ex-ante* content filtering to mitigate the regulatory risk of failing to remove illegal content in time. This is explained in Google's response: "The greater the burden that is put upon intermediaries, in terms of liability and the requirement to use

[160] Online Safety Act 2023 2023 (2023 c 50).
[161] Markus Trengove and others, 'A Critical Review of the Online Safety Bill' (2022) 3 Patterns 2 <https://www.cell.com/patterns/abstract/S2666-3899(22)00147-7> accessed 9 June 2025.
[162] Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act).

resources to mediate or judge third party disputes, the greater will be the incentive to remove content without carefully reviewing or otherwise evaluating the veracity of the notices receive".[163]

The scale of online content has made manual curation or community moderation unfeasible, leading platforms to rely heavily on automated systems.[164] For instance, in the first half of 2024, X (formerly Twitter) suspended 5.2 million accounts and removed 10.7 million posts for rule violations, highlighting the sheer volume of enforcement actions.[165] Although most notice-and-takedown regulations require platforms to adopt moderation measures that are proportionate and respectful of users' rights, they fail to establish clear standards for acceptable levels of content misclassification.[166] This regulatory ambiguity, combined with the pressures of large-scale moderation, encourages over-removal of content. As a result, lawful speech may be disproportionately affected, undermining the very rights the regulations seek to protect.

2.3.2. The Brussels Effect

The "Brussels Effect" describes the European Union's unique and often unintentional capacity to shape global regulations through market mechanisms. It specifically refers to instances where global markets internalize and spread EU regulations, influencing both companies and foreign regulators, even without direct EU intervention.[167]

[163] Aleksandra Kuczerawy, 'Intermediary Liability & Freedom of Expression: Recent Developments in the EU N otice & Action Initiative' (2015) 31 Computer Law & Security Review 46.
[164] Robert Gorwa, Reuben Binns and Christian Katzenbach, 'Algorithmic Content Moderation: Technical and Po litical Challenges in the Automation of Platform Governance' (2020) 7 Big Data & Society 2053951719897945, 3.
[165] 'X Report: We Remove More Accounts, Suspend Fewer Users Than Twitter Did' <https://www.mediapost.co m/publications/article/399766/x-report-we-remove-more-accounts-suspend-fewer-u.html> accessed 9 June 2025.
[166] Marcin Rojszczak, 'The Digital Services Act and the Problem of Preventive Blocking of (Clearly) Illegal Con tent' (2023) 3 Institutiones Administrationis – Journal of Administrative Sciences 44, 51.
[167] Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford University Press 2020) 1.

Though the notice-and-takedown mechanism only applies to certain jurisdictions, once platforms implement such changes for certain markets and apply a strict speech regime, they often apply the same moderation policies globally to streamline enforcement to cut fragmented compliance expenses, causing a global spillover.

For example, after terrorist attacks in Paris and Brussels in late 2015, European regulators pressed social media platforms to take down hate speech and terrorism content, otherwise, they will face civil and criminal penalties.[168] The result is the launch of the 2016 EU Code of Conduct on Countering Illegal Hate Speech Online.[169] Although non-binding, it led Facebook, Microsoft, Twitter, YouTube, and others to adopt 24-hour takedown commitments and revise their terms of service to align with EU definitions of hate speech.[170] The Code also prompted platforms to create a shared database of banned extremist content.[171] Therefore, unlike national laws that apply only within a country's borders, regional regulation of content moderation changes terms of service which apply wherever platforms are accessed, thus influencing global freedom of speech.[172]

2.3.3. Regulatory Abuse

Under the notice-and-takedown mechanism, state public authorities are empowered to issue legally binding content removal orders for specific pieces of illegal content and declare one kind of information is illegal via legislation.

[168] Liat Clark, 'Facebook and Twitter Must Tackle Hate Speech or Face New Laws' *Wired* <https://www.wired.com/story/us-tech-giants-must-tackle-hate-speech-or-face-legal-action/> accessed 11 June 2025; Mark Scott, 'Europe Presses American Tech Companies to Tackle Hate Speech' *The New York Times* (6 December 2016) <https://www.nytimes.com/2016/12/06/technology/europe-hate-speech-facebook-google-twitter.html> accessed 11 June 2025.
[169] 'Code of Conduct on Countering Illegal Hate Speech Online' (European Commission).
[170] 'European Commission: The EU Code of Conduct on Countering Illegal Hate Speech Online - Cyberviolence - Www.Coe.Int' (*Cyberviolence*) <https://www.coe.int/en/web/cyberviolence/-/european-commission-the-eu-code-of-conduct-on-countering-illegal-hate-speech-online> accessed 11 June 2025.
[171] Danielle Keats Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep' (2018) 93 Notre Dame Law Review 1035, 1044.
[172] *ibid 1038.*

Compared to states that follow a classical liberalism idea of freedom of speech, states applying notice-and-takedown mechanisms usually have a broad sweep of speech restrictions.[173] For example, Holocaust denial and justification of National Socialism are criminalized in most Europe countries. However, in the United States, *Brandenburg v. Ohio* (1969) the Supreme Court allowed the National Socialist Party of America to march in Skokie, where most population is Jewish, because it is a freedom of speech under the First Amendment. Nevertheless, in *R.A.V. v. City of St. Paul* (1992), the Supreme Court struck down a hate speech ordinance prohibiting the display of Nazi symbols, arguing it was viewpoint discriminatory, thus unconstitutional under the First Amendment.[174]

However, when the state itself acts unjustly or suppresses dissent, the notice-and-takedown mechanism can become a tool for accelerating censorship and entrenching authoritarian control. Turkey has used its Internet law (Law No. 5651) to send tens of thousands of takedown requests to silence dissent.[175] Russia established the legal framework for online content regulation including the 2002 Law on Countering Extremism, the 2013 and 2022 "LGBTQ propaganda" laws, the 2016 "Yarovaya Package" anti-terror laws, and criminalizes spreading "fake news" and disrespecting the state. To enforce these laws, Russia accounts for approximately 64% of all global takedown requests received by Google, submitting over 211,000 requests in the last ten years, which amounts to nearly 130 notices per day.[176]

[173] Dawn Nunziato, 'The Digital Services Act and the Brussels Effect on Platform Content Moderation' (2023) 2 4 Chicago Journal of International Law 119 <https://chicagounbound.uchicago.edu/cjil/vol24/iss1/6>.
[174] *RAV v City of St Paul* (1992) 505 US 377 (US Supreme Court).
[175] Burak Haylamaz, 'Türkiye's Freedom of Expression: Progress Made, Challenges Remain | TechPolicy.Press' (*Tech Policy Press*, 28 May 2024) <https://techpolicy.press/turkiyes-freedom-of-expression-progress-made-challenges-remain> accessed 10 June 2025.
[176] 'Content Removal Attempts from Google This Decade' (*Surfshark*) <https://surfshark.com/research/study/google-content-removal-attempts> accessed 10 June 2025.

3. Platform Self-Regulation and the Turn Toward Digital Constitutionalism

In the free speech triangle, new school speech control is not a one-way state-imposed moderation obligation on platforms. Platform self-regulation represents social media platforms' response to the state's requirement of a "horizontal" model of rights, whereby nongovernmental actors are expected to preserve and protect basic human rights.[177]

3.1. Meta's Oversight Board as a Quasi-Judicial Experiment

The Meta's Oversight Board is an audacious experiment in self-regulation by one of the world's most powerful corporations, set up to oversee one of the largest systems of speech regulation in history.[178]

In 2021, amid escalating controversies surrounding Facebook's role in the Cambridge Analytica data scandal, alleged interference in the U.S. elections, and its failures in moderating content related to the Rohingya crisis in Myanmar, Meta established the Oversight Board (OB)—often referred to as the "Supreme Court for Facebook"—as a response to growing demands for transparency, accountability, and legitimacy in platform governance.[179] Legally independent from Meta, the OB is an independent institution to review Facebook and Instagram's content moderation decisions. It issues binding decisions on content moderation decisions on Facebook and Instagram and issues non-binding recommendations regarding platform policies.[180] The OB is not a simple extension of Meta's existing content review process but rather focuses

[177] David Wong and Luciano Floridi, 'Meta's Oversight Board: A Review and Critical Assessment' (2023) 33 M inds and Machines 261, 264.
[178] Evelyn Douek, 'The Meta Oversight Board and the Empty Promise of Legitimacy' [2023] SSRN Electronic J ournal 373 <https://www.ssrn.com/abstract=4565180> accessed 11 June 2025.
[179] Kate Klonick, 'The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression' (2019) 129 Yale Law Journal 2418, 2445; Ruby O'Kane, 'Meta's Private Speech Governance and t he Role of the Oversight Board: Lessons From' 168.
[180] Wong and Floridi (n 177) 262.

particularly on "the impact of removing content in light of human rights norms protecting free expression" balanced against other values such as "authenticity, safety, privacy and dignity".[181]

Therefore, the OB functions as a private, quasi-judicial body, offering users a channel for appeal that exists outside the platform's internal bureaucracy. As such, it addresses some of the structural deficiencies inherent in platform-led content moderation—particularly the tendency toward over-moderation driven by risk aversion ("erring on the side of caution") and the potential bias introduced by commercial incentives, such as maximizing profit and user engagement.[182] It also provides a form of public reasoning, accountability, and transparency, lending legitimacy to platforms, whose moderation decision-making processes have long operated as opaque "black boxes". For example, the OB has recommended that Meta translate its Internal Implementation Standards into the languages relevant to the content being moderated. In three separate cases involving content in Punjabi, Burmese, and Arabic, the OB emphasized the importance of linguistic accessibility to ensure accurate, context-aware, and culturally sensitive moderation.[183] However, platforms often overlook the needs of marginalized language communities, primarily due to cost considerations and operational efficiency.

The normative frameworks guiding the OB's judging are interesting. Besides "Meta's content policies and values" as the basis of decision-making in Article 2 of its Convention, the OB has adopted many public international laws as its ruling basis. The OB has established an International Human Rights Law (IHRL) framework from its earliest decisions, judging Meta's actions according to obligations arising under the United Nations Guiding Principles on

[181] 'Oversight Board Charter' 2 <https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf> accessed 11 June 2025.
[182] O'Kane (n 179) 171.
[183] 'Oversight Board 2022 Annual Report' 24.

Business and Human Rights (UNGPs) and the balancing methodology outlined in the International Covenant on Civil and Political Rights (ICCPR) Article 19.[184]

In its earliest cases, the OB has applied public international laws and United Nations instruments as normative standards. In *Case 2020-004-IG-UA* (Breast Cancer Symptoms and Nudity), the OB listed related standards in its judgment, including Facebook's content policies; Facebook's values; UNGP; ICCPR; ICESCR; CEDAW; and CRC.[185] In *Case 2023-004-FB-MR* (Armenian Prisoners of War Video), the OB also cited the Geneva Convention and the Rome Statute the of International Criminal Court (ICC).

When the OB deals with cases in the context of political unrest and protest, the OB uses public international laws to further examine the context of the content-related region's political situation to decide whether the human rights protection value of Meta should be applied. In *Case 2021-012-FB-UA* (Wampum Belt), the OB emphasized the importance of examining the whole content for contextual cues, and not removing posts based on a decontextualized phrase in isolation.[186] In *Case 2021-010-FB-UA* (Colombia Protests), the OB also stressed that, in contexts where traditional outlets for political expression are restricted, social media serves as a vital space for individuals and journalists to share information about protests. In such settings, applying the newsworthiness allowance permits only limited exceptions for harmful content, justified by the public interest in disseminating politically significant information.[187] Therefore, the OB does a meticulous analysis of the political situation as a background of content moderation.

[184] O'Kane (n 179) 173.
[185] 'Breast Cancer Symptoms and Nudity | Oversight Board' <https://www.oversightboard.com/decision/ig-7thr3si1/> accessed 13 June 2025.
[186] 'Wampum Belt | Oversight Board' <https://www.oversightboard.com/decision/fb-l1lania7/> accessed 13 June 2025.
[187] 'Colombia Protests | Oversight Board' <https://www.oversightboard.com/decision/fb-e5m6qzga/> accessed 13 June 2025.

In *Case 2025-012-FB-UA* (Content Targeting Human Rights Defender in Peru), the OB cited resources from the Office of the United Nations High Commissioner for Human Rights (OHCHR), UN Special Rapporteur, Human Rights Watch, legislative initiatives in Peru. Finally, the OB overturns Meta's decision to leave up content targeting one of Peru's leading human rights defenders and cited the International Center for Not-For-Profit Law as an interpretation of human rights value: "The work of human rights defenders is essential for strengthening democracy and the rule of law…respect for human rights in a democratic society largely depends on effective and adequate guarantees for human rights defenders that enable them to carry out their activities freely".[188]

Besides supporting moderation of content against human rights defenders, the OB has also overturned Meta's decisions to remove content posted by protesters. In *Case 2022-013-FB-UA* (Iran Protest Slogan), the OB found "in the Iranian context" where "the government systematically represses freedom of expression; the digital spaces have become a key forum for dissent". Therefore, the OB urged Meta to support users' voices. The OB cited UN reports, the Islamic Penal Code of Iran, and documentation from the Open Observatory of Network Interference (OONI) to highlight the severe restrictions imposed by the Iranian state on online expression, and to emphasize the role Meta must play in supporting freedom of expression in digital spaces under such conditions.[189] Similarly, in *Case 2021-009-FB-UA* (the "Posts That Include 'From the River to the Sea'" case) the OB assessed the political situation in the Israel-Palestine conflict by reviewing materials from the UN Office for the Coordination of Humanitarian Affairs (OCHA), the ICC, the UN General Assembly.

[188] 'Content Targeting Human Rights Defender in Peru | Oversight Board' <https://www.oversightboard.com/decision/fb-28m1tlxl/> accessed 14 June 2025.
[189] 'Iran Protest Slogan | Oversight Board' <https://www.oversightboard.com/decision/fb-zt6ajs4x/> accessed 14 June 2025.

By relying on international laws and a detailed assessment of the social and political context of online speech, the OB operates in a quasi-judicial capacity by adopting the language, structures, and procedures of public law institutions to regulate online speech. It seeks to legitimize private content governance by embedding public law principles into the architecture of platform regulation, and it has successfully made Meta accept and enforce some of its non-binding policy recommendations. Notably, Meta has unified its Community Standards across Facebook, Instagram, and Threads to ensure greater policy consistency. It has also improved user notification systems, allowing users to receive warnings and complete educational modules before penalties are applied. In response to calls for increased transparency, Meta has launched the "Oversight Board Impact Tracker" to publicly display the status of each recommendation.[190] There are some effective outcomes, for example, after the recommendation made by the OB in *Case 2022-013-FB-UA* was implemented by Meta, content containing the phrase against the Iranian government increased by nearly 30% on Meta's social media platform.[191]

3.2. Risks

3.2.1. Private Power

Though the application of international laws and human rights standards has been praised as a brave try to constitutionalize digital space, the OB's simple application of IHRL and other international laws doesn't answer a question: since these laws are made to restrict state, why do these laws apply to the private sector such as Meta? While using the three-part test (legality, a legitimate aim, and necessity and proportionality) under IHRL for state restrictions on speech into decisions regarding a private sector's content moderation all the time, the OB failed to

[190] 'Oversight Board Recommendations | Transparency Center' <https://transparency.meta.com/oversight/oversight-board-recommendations/> accessed 14 June 2025.
[191] 'Recommendations | Oversight Board' (30 April 2024) <https://www.oversightboard.com/recommendations/> accessed 14 June 2025.

explain this application other than citing Meta's voluntary commitment to respect human right by the UNGP.[192]

Crucially, IHRL, like First Amendment law, is a body of norms intended to constrain public authorities. As criticized by Evelyn Douek: "It cannot simply be transposed from state-based jurisprudence and applied to the practices of private companies without interrogation of the very meaningful differences between these two contexts."[193] The capability of free speech restriction and the risk it brings between nation-state and private platforms are different. Take Albert O. Hirschman's conception in *Exit, Voice, and Loyalty*, the difficulty and cost of exiting or fighting against a social media platform moderating inappropriately are much harder than exiting or fighting against a nation-state suppressing free speech systematically.[194] Even though in *Case 2020-003-FB-UA* (Armenians in Azerbaijan) the OB agreed with the UN Special Rapporteur on freedom of expression that "although companies do not have the obligations of Governments, their impact is of a sort that requires them to assess the same kind of questions about protecting their users' right to freedom of expression", in further cases, the OB failed to clarify what are the unique obligations for companies and how should international laws be chosen and applied in this framework.[195]

3.2.2. State Pressure

The OB's decisions regarding political protest and content from regions experiencing civil unrest often place Meta in direct tension with domestic legal orders, particularly in authoritarian or semi-authoritarian states with restrictive speech regimes.

[192] Douek (n 178) 43.
[193] *ibid 44.*
[194] Albert O Hirschman, *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States* (Harvard University Press 2004).
[195] 'Armenians in Azerbaijan | Oversight Board' <https://www.oversightboard.com/decision/fb-qbjdascv/> accessed 14 June 2025.

Though the OB often emphasized that Meta should refrain from removing certain opinions in the interest of protecting free expression, such decisions can provoke a backlash from authoritarian or semi-authoritarian governments. In response, states may choose to escalate by imposing heavy fines, demanding local content removal, or even blocking access to Facebook entirely.[196] Ironically, this may result in a more repressive environment for online speech than the original platform moderation.

To what extent does the OB's ICC-like approach characterized by contextual assessments of the political environment in sensitive regions meaningfully advance Meta's commitment to protecting freedom of speech? Or, alternatively, is this merely a form of political posturing given Meta's parallel obligation to preserve market access and avoid regulatory retaliation from authoritarian or semi-authoritarian regimes?

3.2.3. Fragile Autonomy

To ensure the OB's institutional independence, Meta established the independent Oversight Board Trust in 2019, initially committing $130 million and later an additional $150 million in 2022. Despite this formal separation, the OB remains structurally embedded within Meta's institutional and financial framework. Its independence is thus contingent, vulnerable to funding limitations, jurisdictional boundaries, ignored recommendations, and political pressures in sensitive markets. For example, in the second half of 2023, approximately 41.8 % of the OB's recommendations were either declined, still awaiting action, or not implemented.[197]

[196] Shannon Bond, 'India's Government Is Telling Facebook, Twitter To Remove Critical Posts' *NPR* (27 April 2021) <https://www.npr.org/2021/04/27/991343032/indias-government-is-telling-facebook-twitter-to-remove-critical-posts> accessed 15 June 2025; 'Meta Faces "Substantial" Fine for Not Complying with Turkey's Gag Orders' (*POLITICO*, 1 April 2025) <https://www.politico.eu/article/meta-turkey-gag-turkish-government-mayor-ekrem-imamoglu/> accessed 15 June 2025; Pjotr Sauer, 'Russia Bans Facebook and Instagram under "Extremism" Law' *The Guardian* (21 March 2022) <https://www.theguardian.com/world/2022/mar/21/russia-bans-facebook-and-instagram-under-extremism-law> accessed 15 June 2025.

[197] 'Meta's Oversight Board Is Unprepared for a Historic 2024 Election Cycle' (*Brookings*) <https://www.brookings.edu/articles/metas-oversight-board-is-unprepared-for-a-historic-2024-election-cycle/> accessed 15 June 2025.

After Donald Trump won the 2025 presidential election, the CEO of Meta, Mark Zuckerberg, made a series of statements about the change of Meta's content moderation policy. The announcement included that would "get rid of fact-checkers", remove restrictions on subjects like immigration and gender; and bring back political content on its platforms.[198] This wasn't the only time Zuckerberg yielded to Trump. In 2018, he met with conservative leaders and launched reviews to ensure "viewpoint neutrality" and ideological diversity in moderation.[199] The relationship between Meta and OB has grown increasingly strained, and it is not hard to imagine that the OB will be marginalized.[200]

[198] Judit Bayer, 'Zuckerberg's Strategy' [2025] Verfassungsblog <https://verfassungsblog.de/zuckerbergs-strategy/> accessed 14 June 2025.
[199] By Haley Samsel, 'In U.S. House, Texas Republicans Grill Zuckerberg More on Whether Facebook Is Censoring Conservatives' (*The Texas Tribune*, 11 April 2018) <https://www.texastribune.org/2018/04/11/house-mark-zuckerberg-grilled-more-texas-republicans-whether-facebook-/> accessed 14 June 2025.
[200] Juha Tuovinen, 'The Meta Oversight Board in the Trump Era' [2025] Verfassungsblog <https://verfassungsblog.de/the-meta-oversight-board-in-the-trump-era/> accessed 14 June 2025.

CHAPTER FOUR: CONCLUSION

The governance of online speech is increasingly shaped by the dynamic interplay among states, platforms, and users—a triangular relationship in which power can tilt in different directions over time. As this thesis has shown, favoring any one vertex of the free speech triangle inevitably generates distinct advantages and exposes different vulnerabilities. Privileging platforms, the classical liberal understanding of free speech embedded in Section 230 enables swift and scalable moderation but risks corporate arbitrariness; leaning toward the state, the notice-and-takedown model widely used by the European countries may bring democratic accountability but also opens the door to coercive overreach; Meta's Oversight Board represents an attempt to center human rights and empower expressive agency, but it lacks the institutional mechanisms necessary to ensure structural fairness and meaningful accountability. These competing poles of the free speech triangle are not abstract, they shape how speech is structured, discourse unfolds, and rights are experienced.

This thesis does not seek to identify a universally "best" governance model, nor does it assume that such a stable best practice could exist within the inherently fluid and contested space of the free speech triangle. Rather, the aim has been to critically examine how each regulatory configuration manages the tension between liberty and harm, autonomy and accountability. In a system where power is dispersed but not equally distributed, no single solution can offer lasting equilibrium. Instead, the triangle must be understood as a shifting field of influence, where regulatory strategies must continuously adapt to changing sociopolitical, technological, and legal conditions.

While this thesis has focused exclusively on so-called transatlantic "free countries" as defined by Freedom House, it is essential to acknowledge that as of early 2024, over 3.8 billion internet

users reside in these "not free" or "partly free" jurisdictions, accounting for nearly 79% of the global online population.[201] Understanding how content moderation operates in these environments is therefore not peripheral but central to the future of global free speech.

Similar regulatory frameworks do not yield similar outcomes across different political systems. A 2020 report by Jacob Mchangama, Joelle Fiss and Natalie Alkiviadou documented that at least 25 countries have adopted or proposed models similar to the NetzDG. Nine of those countries were ranked "not free" (Honduras, Venezuela, Vietnam, Russia, Belarus, Egypt, Ethiopia, Pakistan and Turkey), twelve were ranked "partly free" (Kenya, India, Singapore, Malaysia, Philippines, Mali, Morocco, Nigeria, Cambodia, Indonesia, Kyrgyzstan and Brazil), and only four ranked "free" (France, UK, Australia and Austria).[202]

Notably, many authoritarian and hybrid regimes have cited regulatory models from liberal democracies to legitimize their own speech restrictions. Although these laws may mirror democratic frameworks in form, their implementation often serves repressive ends, suppressing dissent and consolidating state control.[203] This underscores the risk of legal transplants: rights-based models, when adopted in illiberal contexts, can be repurposed as instruments of censorship.

Any future theory of global online speech governance must remain attentive to the Janus-faced nature of these frameworks: in democratic settings, they may act as guardians of pluralistic expression; in authoritarian contexts, they frequently become tools of repression and control.

[201] 'The Struggle for Trust Online' (*Freedom House*) <https://freedomhouse.org/report/freedom-net/2024/struggle-trust-online> accessed 16 June 2025.
[202] Jacob Mchangama, Natalie Alkiviadou and Jacob Mchangama and Natalie Alkiviadou, 'The Digital Berlin Wall – How Germany (Accidentally) Created a Prototype for Global Online Censorship – Act Two' (*The Future of Free Speech*, 1 October 2020) <https://futurefreespeech.org/the-digital-berlin-wall-how-germany-accidentally-created-a-prototype-for-global-online-censorship-act-two/> accessed 16 June 2025.
[203] Isabelle Canaan, 'NetzDG and the German Precedent for Authoritarian Creep and Authoritarian Learning' (Social Science Research Network, 10 August 2021) <https://papers.ssrn.com/abstract=3908440> accessed 16 June 2025.

BIBLIOGRAPHY

1. Statutes and Treaties

Communications Decency Act § 230 1996 (USC)

Convention for the Protection of Human Rights and Fundamental Freedoms 1953

Décret n° 2015-125 du 5 février 2015 relatif au blocage des sites provoquant des actes de terrorisme ou en faisant l'apologie 2015 (JORF n°0031 du 6 février 2015)

Directive 2000/31/EC on Electronic Commerce (E-Commerce Directive) 2000 (OJ L 178)

Fight Online Sex Trafficking Act of 2017 (FOSTA) 2018 (Pub L No 115–164)

Loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique (LCEN) 2004 (JORF n°143 du 22 juin 2004)

Loi n° 2024-449 du 21 mai 2024 visant à sécuriser et réguler l'espace numérique (loi SREN) 2024 (JORF n°0118 du 22 mai 2024)

Netzwerkdurchsetzungsgesetz (NetzDG) 2017 (BGBl I)

Online Safety Act 2023 2023 (2023 c 50)

Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act) 2022 (OJ L 277)

U.S. Constitution amend. I 1791 (US Const)

2. Cases

Abrams v United States (1919) 250 US 616 (Supreme Court)

Airey v Ireland [1979] ECtHR 78103/14

Barnes v Yahoo!, Inc (2009) 570 F3d 1096 (9th Cir)

Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV [2012] CJEU C-360/10, ECLI:EU:C:2012:85 ECLI

Blockowicz v Williams (2010) 630 F3d 563 (7th Cir)

Blumenthal v Drudge (1998) 992 F Supp 44 (DDC)

Daniel v Armslist, LLC (2019) 926 Wis 710 (NW2d)

Delfi as v Estonia [2015] ECtHR [GC] 64569/09

Dink v Turkey [2010] ECtHR 2668/07, 6102/08, 30079/08, 7072/09, 7124/09

Doe v America Online, Inc (2001) 783 So 2d 1010 (Fla)

Doe v Backpage.com, LLC (2018) 881 F3d 167 (F3d1st Cir)

Doe v Grindr Inc (2025) 128 F4th 1148 (9th Cir)

Doe v MindGeek USA Inc (2021) 558 F Supp 3d 828 (CD Cal)

Doe v MySpace (2008) (2008) 528 F3d 413 (5th Cir)

Doe v Reddit, Inc (2021) 2021 US Dist LEXIS 235993 (CD Cal)

Doe v Salesforce, Inc (2025) 2025 US Dist LEXIS 52346 (ND Tex)

Elansari v Meta Inc, (2024) 2022 WL 16948246 (3d Cir)

Eva Glawischnig-Piesczek v Facebook Ireland Limited [2019] CJEU C-18/18, ECLI:EU:C:2019:821 ECLI

Fair Hous Council of San Fernando Valley v Roommates.com, LLC (2008) 521 F3d 1157 (9th Cir)

Force v Facebook, Inc (2019) 934 F3d 53 (2d Cir)

Frank Peterson v Google LLC and YouTube LLC [2021] CJEU C-682/18, ECLI:EU:C:2021:503 ECLI

Freedom Watch, Inc v Google, Inc (2020) 816 F APP'x 497 (DC Cir)

FTC v Accusearch Inc [2009] 10th Cir, 570, 1187 F3d

Google France SARL and Google Inc v Louis Vuitton Malletier SA [2010] ECJ Joined cases C-236/08 to C-238/08

Green v American Online, Inc (2003) 318 F3d 465 (3d Cir)

Handyside v United Kingdom (1976) 24 Eur Ct HR (European Court of Human Rights)

Hassell v Bird (2018) 5 Cal5th 522 (Cal)

Herrick v Grindr, LLC (2019) 765 F App'x 586

JB v G6 Hospitality, LLC (2020) 2020 US Dist LEXIS 151213 (ND Cal)

Jones v Dirty World Entertainment Recordings LLC (2014) 755 F3d 398 (6th Cir)

Kimzey v Yelp! (2016) 836 F3d 1263 (9th Cir)

Klayman v Zuckerberg (2014) 753 F3d 1354 (DC Cir)

Knight First Amendment Inst at Columbia Univ v Trump 928 F3d 226 (Second Circuit Court of Appeals)

Lingens v Austria [1986] European Court of Human Rights App. No. 9815/82, 8 Eur HR Rep 407

L'Oréal SA and Others v eBay International AG and Others (ECJ)

MH v Omegle.com, LLC (2024) 122 F 4th 1266 (11th Cir)

Moody v NetChoice, LLC (2022) 34 F4th 1196 (11th Cir)

NetChoice, LLC v Paxton (2022) 49 F4th 439 (5th Cir)

New York Times Co v Sullivan (1964) 376 US 254 (Supreme Court)

Ozgur Gundem v Turkey [2000] ECtHR 20046/16, 21350/16, 26213/16, 51314/16, 54383/16, 57176/16, 58508/16, 4630/17, 7268/17, 18590/19, 34713/19, 38209/19, 62293/19, 4853/20, 7245/20

Packingham v North Carolina (2017) 582 US 98 (Supreme Court)

Prager Univ v Google LLC (2020) 951 F3d 991 (9th Cir)

RAV v City of St Paul (1992) 505 US 377 (US Supreme Court)

Reno v ACLU (1997) 521 US 844 (Supreme Court)

Reno v Am Civil Liberties Union (1997) 521 US 844

Republican Nat'l Comm v Google LLC (2023) 2023 WL 5731496 (ND Cal)

Sanchez v France [2023] ECtHR [GC] 45581/15

Sotiris Papasavvas v O Fileleftheros Dimosia Etaireia Ltd and Others [2014] ECJ Case C-291/13

Stratton Oakmont, Inc v Prodigy Services Co [1995] NY Sup Ct No. 31063/94, 1995 WL 323710

Sun v China Press, Inc (2022) 2022 WL 1718792 (ND Cal)

Tiffany (NJ) Inc v eBay Inc (2010) 600 F3d 93 (2d Cir)

Woodhull Freedom Found v United States (2023) 72 F4th 1286 (DC Cir)

X and Y v the Netherlands [1985] ECtHR 8978/80

Zeran v America Online, Inc (1997) 129 F3d 327 (4th Cir)

Zöchling v Austria App no 1510/03 (ECtHR, 7 March 2006)

Zhang v Baidu.com Inc (2014) 10 F Supp 3d 433 (SDNY)

3. Books and Journal Articles

Albert K and others, 'Fosta in Legal Context' (2020) 52 Columbia Human Rights Law Review 1084

Ardia DS, 'Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity under Section 230 of the Communications Decency Act'

'Armenians in Azerbaijan | Oversight Board' <https://www.oversightboard.com/decision/fb-qbjdascv/> accessed 14 June 2025

Balkin JM, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation' (Social Science Research Network, 9 September 2017) <https://papers.ssrn.com/abstract=3038939> accessed 18 April 2025

——, 'Free Speech Is a Triangle Essays' (2018) 118 Columbia Law Review 2011

Barendt E, *Freedom of Speech* (2nd edition, Oxford University Press 2005)

Barrie Sander, 'Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation' (2020) 43 Fordham International Law Journal 939

Bastian R, 'Content Moderation Issues Online: Section 230 Is Not to Blame' (2022) 8 Texas A&M Journal of Property Law 43

——, 'Content Moderation Issues Online: Section 230 Is Not to Blame' (2022) 8 Texas A&M Journal of Property Law 43

Bayer J, 'Zuckerberg's Strategy' [2025] Verfassungsblog <https://verfassungsblog.de/zuckerbergs-strategy/> accessed 14 June 2025

Berlin I, *Liberty: Incorporating Four Essays on Liberty* (Henry Hardy ed, 2nd edition, Oxford University Press 2002)

Bicknell C, 'CDA: From Conception to Supreme Court' *Wired* <https://www.wired.com/1997/03/cda-from-conception-to-supreme-court/> accessed 13 May 2025

Bloch-Wehba H, 'Global Platform Governance: Private Power in the Shadow of the State' (2019) 72 SMU Law Review 27

Bond S, 'India's Government Is Telling Facebook, Twitter To Remove Critical Posts' *NPR* (27 April 2021) <https://www.npr.org/2021/04/27/991343032/indias-government-is-telling-facebook-twitter-to-remove-critical-posts> accessed 15 June 2025

Bradford A, *The Brussels Effect: How the European Union Rules the World* (Oxford University Press 2020)

Brannon VC, 'Liability for Content Hosts: An Overview of the Communication Decency Act's Section 230'

‘Breast Cancer Symptoms and Nudity | Oversight Board’ <https://www.oversightboard.com/decision/ig-7thr3si1/> accessed 13 June 2025

Canaan I, ‘NetzDG and the German Precedent for Authoritarian Creep and Authoritarian Learning’ (Social Science Research Network, 10 August 2021) <https://papers.ssrn.com/abstract=3908440> accessed 16 June 2025

Carr CT and and Hayes RA, ‘Social Media: Defining, Developing, and Divining’ (2015) 23 Atlantic Journal of Communication 46

Celeste E and others, *The Content Governance Dilemma: Digital Constitutionalism, Social Media and the Search for a Global Standard* (Springer International Publishing 2023) <https://link.springer.com/10.1007/978-3-031-32924-1> accessed 22 April 2025

Chander A, ‘How Law Made Silicon Valley’ (2013) 63 Emory Law Journal 639

Citron D and Wittes B, ‘The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity’ (2017) 86 Fordham Law Review 401

Clark L, ‘Facebook and Twitter Must Tackle Hate Speech or Face New Laws’ *Wired* <https://www.wired.com/story/us-tech-giants-must-tackle-hate-speech-or-face-legal-action/> accessed 11 June 2025

‘Code of Conduct on Countering Illegal Hate Speech Online’ (European Commission)

Cole D, ‘Who Should Regulate Online Speech?’ (2024) 71 *The New York Review of Books* <https://www.nybooks.com/articles/2024/03/21/who-should-regulate-online-speech/> accessed 28 April 2025

‘Colombia Protests | Oversight Board’ <https://www.oversightboard.com/decision/fb-e5m6qzga/> accessed 13 June 2025

‘Commission Opens Formal Proceedings against AliExpress under the Digital Services Act | Shaping Europe’s Digital Future’ <https://digital-strategy.ec.europa.eu/en/news/commission-opens-formal-proceedings-against-aliexpress-under-digital-services-act> accessed 4 June 2025

‘Commission Opens Formal Proceedings against Facebook and Instagram under the Digital Services Act | Shaping Europe’s Digital Future’ <https://digital-strategy.ec.europa.eu/en/news/commission-opens-formal-proceedings-against-facebook-and-instagram-under-digital-services-act> accessed 4 June 2025

‘Commission Preliminarily Finds TikTok’s Ad Repository in Breach of the Digital Services Act | Shaping Europe’s Digital Future’ <https://digital-strategy.ec.europa.eu/en/news/commission-preliminarily-finds-tiktoks-ad-repository-breach-digital-services-act> accessed 4 June 2025

‘Content Removal Attempts from Google This Decade’ (*Surfshark*) <https://surfshark.com/research/study/google-content-removal-attempts> accessed 10 June 2025

‘Content Targeting Human Rights Defender in Peru | Oversight Board’ <https://www.oversightboard.com/decision/fb-28m1tlxl/> accessed 14 June 2025

‘Countries and Territories’ (*Freedom House*) <https://freedomhouse.org/country/scores> accessed 22 April 2025

Cox N, ‘Elfi v Stonia: The Liability of Secondary Internet Publishers for Violation of Reputational Rights under the European Convention on Human Rights’ (2014) 77 The Modern Law Review 619

De Gregorio G, ‘From Constitutional Freedoms to the Power of the Platforms: Protecting Fundamental Rights Online in the Algorithmic Society’ (Social Science Research Network, 2018) <https://papers.ssrn.com/abstract=3365106> accessed 25 April 2025

——, ‘Digital Constitutionalism across the Atlantic’ (2022) 11 Global Constitutionalism 297

‘Décision n° 2020-801 DC du 18 juin 2020 | Conseil constitutionnel’ <https://www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm> accessed 7 June 2025

‘Décret N° 2015-125 Du 5 Février 2015 Relatif Au Blocage Des Sites Provoquant à Des Actes de Terrorisme Ou En Faisant l’apologie et Des Sites Diffusant Des Images et Représentations de Mineurs à Caractère Pornographique - Légifrance’ <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000030195477> accessed 7 June 2025

Douek E, ‘The Meta Oversight Board and the Empty Promise of Legitimacy’ [2023] SSRN Electronic Journal <https://www.ssrn.com/abstract=4565180> accessed 11 June 2025

‘DSA: Commission Opens Formal Proceedings against Meta’ (*European Commission - European Commission*) <https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2664> accessed 4 June 2025

‘European Commission: The EU Code of Conduct on Countering Illegal Hate Speech Online - Cyberviolence - Www.Coe.Int’ (*Cyberviolence*) <https://www.coe.int/en/web/cyberviolence/-/european-commission-the-eu-code-of-conduct-on-countering-illegal-hate-speech-online> accessed 11 June 2025

‘Facebook Issued with 2 Million Euro Fine’ (*Clifford Chance*) <https://www.cliffordchance.com/content/cliffordchance/insights/resources/blogs/talking-tech/en/articles/2019/07/facebook-issued-with-2-million-euro-fine.html> accessed 4 June 2025

‘Freedom on the Net’ (*Freedom House*, 16 October 2024) <https://freedomhouse.org/report/freedom-net> accessed 24 April 2025

Frosio G, ‘From the E-Commerce Directive to the Digital Services Act’ (SSRN, 2024) <https://www.ssrn.com/abstract=4914816> accessed 31 May 2025

Frosio G and Geiger C, ‘Taking Fundamental Rights Seriously in the Digital Services Act’s Platform Liability Regime’ (2023) 29 European Law Journal 31

——, ‘Towards a Digital Constitution’ [2024] Verfassungsblog <https://verfassungsblog.de/towards-a-digital-constitution/> accessed 3 June 2025

Geary K, ‘Section 230 of the Communications Decency Act, Product Liability, and a Proposal for Preventing Dating-App Harassment’ (2021) 125 Penn State Law Review <https://elibrary.law.psu.edu/pslr/vol125/iss2/4>

Gillespie T, 'Regulation of and by Platforms' in Jean Burgess, Alice Marwick and Thomas Poell, *The SAGE Handbook of Social Media* (SAGE Publications Ltd 2018) <https://sk.sagepub.com/reference/the-sage-handbook-of-social-media/i2081.xml> accessed 27 April 2025

Goldman E, 'Of Course The First Amendment Protects Baidu's Search Engine, Even When It Censors Pro-Democracy Results (Forbes Cross-Post)' (*Technology & Marketing Law Blog*, 10 April 2014) <https://blog.ericgoldman.org/archives/2014/04/of-course-the-first-amendment-protects-baidus-search-engine-even-when-it-censors-pro-democracy-results-forbes-cross-post.htm> accessed 29 May 2025

——, 'The Ten Most Important Section 230 Rulings' (2017) 20 Tulane Journal of Technology and Intellectual Property 1

——, 'The Complicated Story of Fosta and Section 230' (2018) 17 First Amendment Law Review 279

——, 'An Overview of the United States' Section 230 Internet Immunity' in Giancarlo Frosio (ed), *Oxford Handbook of Online Intermediary Liability* (Oxford University Press 2020) <https://doi.org/10.1093/oxfordhb/9780198837138.013.8> accessed 21 May 2025

——, 'Content Moderation Remedies' (2021) 28 Michigan Technology Law Review 1

——, 'DC Circuit Upholds FOSTA's Constitutionality (By Narrowing It)-Woodhull v. U.S.' (*Technology & Marketing Law Blog*, 23 July 2023) <https://blog.ericgoldman.org/archives/2023/07/dc-circuit-upholds-fostas-constitutionality-by-narrowing-it-woodhull-v-u-s.htm> accessed 23 May 2025

Goldman E and Miers J, 'Online Account Terminations/Content Removals and the Benefits of Internet Services Enforcing Their House Rules' (Social Science Research Network, 1 August 2021) <https://papers.ssrn.com/abstract=3911509> accessed 28 May 2025

Goodman A, 'Blocking Pro-Terrorist Websites: A Balance between Individual Liberty and National Security in France' (2016) 22 Southwestern Journal of International Law 209

Gorwa R, Binns R and Katzenbach C, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020) 7 Big Data & Society 2053951719897945

——, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020) 7 Big Data & Society 2053951719897945

Haylamaz B, 'Türkiye's Freedom of Expression: Progress Made, Challenges Remain | TechPolicy.Press' (*Tech Policy Press*, 28 May 2024) <https://techpolicy.press/turkiyes-freedom-of-expression-progress-made-challenges-remain> accessed 10 June 2025

Hirschman AO, *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States* (Harvard University Press 2004)

Huang JT, Choi J and Wan Y, 'Politically Biased Moderation Drives Echo Chamber Formation: An Analysis of User-Driven Content Removals on Reddit' (Social Science Research Network, 17 October 2024) <https://papers.ssrn.com/abstract=4990476> accessed 29 May 2025

‘Internet and Social Media Users in the World 2025’ (*Statista*) <https://www.statista.com/statistics/617136/digital-population-worldwide/> accessed 18 April 2025

‘Internet Users - United States - Telecommunications’ <https://www.indexmundi.com/united_states/internet-users.html?utm_source=chatgpt.com> accessed 28 May 2025

‘Iran Protest Slogan | Oversight Board’ <https://www.oversightboard.com/decision/fb-zt6ajs4x/> accessed 14 June 2025

Kane MJ, ‘Blumenthal v. Drudge Part VI: Business Law: Section 1: Electronic Commerce: B) Internet Service Provider Liability’ (1999) 14 Berkeley Technology Law Journal 483

Kant I, *Groundwork for the Metaphysics of Morals* (Allen W Wood ed, Yale University Press 2002)

Keats Citron D, ‘Extremist Speech, Compelled Conformity, and Censorship Creep’ (2018) 93 Notre Dame Law Review 1035

Klonick K, ‘The New Governors: The People, Rules, and Processes Governing Online Speech’ (2017) 131 Harvard Law Review 1598

——, ‘The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression’ (2019) 129 Yale Law Journal 2418

Kosseff J, *The Twenty-Six Words That Created the Internet* (Cornell University Press 2019)

——, ‘What Was the Purpose of Section 230? That’s a Tough Question Response’ (2023) 103 Boston University Law Review 763

Kuczerawy A, ‘Intermediary Liability & Freedom of Expression: Recent Developments in the EU Notice & Action Initiative’ (2015) 31 Computer Law & Security Review 46

Lewis M, ‘The NetzDG and the Avia Law: How Two Different Legal Systems Created Two Different Outcomes from Similar Laws’ (2022) 40 Wisconsin International Law Journal 491

Loven C, ‘“Verticalised” Cases before the European Court of Human Rights Unravelled: An Analysis of Their Characteristics and the Court’s Approach to Them’ (2020) 38 Netherlands Quarterly of Human Rights 246

Masur JM, ‘A MOST UNCOMMON CARRIER: ONLINE SERVICE PROVIDER IMMUNITY AGAINST DEFAMATION CLAIMS IN BLUMENTHAL v. DRUDGE’ (2000) 40 Jurimetrics 217

Mchangama J and Alkiviadou N, ‘Hate Speech and the European Court of Human Rights: Whatever Happened to the Right to Offend, Shock or Disturb?’ (2021) 21 Human Rights Law Review 1008

Mchangama J, Alkiviadou N and Alkiviadou JM and N, ‘The Digital Berlin Wall – How Germany (Accidentally) Created a Prototype for Global Online Censorship – Act Two’ (*The Future of Free Speech*, 1 October 2020) <https://futurefreespeech.org/the-digital-berlin-wall-

how-germany-accidentally-created-a-prototype-for-global-online-censorship-act-two/> accessed 16 June 2025

'Meta Faces "Substantial" Fine for Not Complying with Turkey's Gag Orders' (*POLITICO*, 1 April 2025) <https://www.politico.eu/article/meta-turkey-gag-turkish-government-mayor-ekrem-imamoglu/> accessed 15 June 2025

'Meta's Oversight Board Is Unprepared for a Historic 2024 Election Cycle' (*Brookings*) <https://www.brookings.edu/articles/metas-oversight-board-is-unprepared-for-a-historic-2024-election-cycle/> accessed 15 June 2025

Mill JS and Himmelfarb G, *On Liberty* (1st edition, Longman 1998)

Moran C, 'Injunction Relief: Must Nonparty Websites Obey Orders to Remove User Content' (2011) 7 Washington Journal of Law, Technology & Arts 47

'National Rifle Ass'n of America v. Vullo' (*Harvard Law Review*, 11 November 2024) <https://harvardlawreview.org/?p=16794> accessed 28 April 2025

Nunziato D, 'The Digital Services Act and the Brussels Effect on Platform Content Moderation' (2023) 24 Chicago Journal of International Law <https://chicagounbound.uchicago.edu/cjil/vol24/iss1/6>

——, 'The Digital Services Act and the Brussels Effect on Platform Content Moderation' (2023) 24 Chicago Journal of International Law <https://chicagounbound.uchicago.edu/cjil/vol24/iss1/6>

O'Kane R, 'Meta's Private Speech Governance and the Role of the Oversight Board: Lessons From'

Overholser G and Jamieson KH, *The Press* (Oxford University Press 2005)

'Oversight Board 2022 Annual Report'

'Oversight Board Charter' <https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf> accessed 11 June 2025

'Oversight Board Recommendations | Transparency Center' <https://transparency.meta.com/oversight/oversight-board-recommendations/> accessed 14 June 2025

Paul K, 'Meta Struggles with Moderation in Hebrew, According to Ex-Employee and Internal Documents' *The Guardian* (15 August 2024) <https://www.theguardian.com/technology/article/2024/aug/15/meta-content-moderation-hebrew> accessed 29 May 2025

——, 'Meta Struggles with Moderation in Hebrew, According to Ex-Employee and Internal Documents' *The Guardian* (15 August 2024) <https://www.theguardian.com/technology/article/2024/aug/15/meta-content-moderation-hebrew> accessed 29 May 2025

Pentney K, ‘States’ Positive Obligation to Create a Favourable Environment for Participation in Public Debate: A Principle in Search of a Practical Effect?’ (2024) 16 Journal of Media Law 146

Pozen DE, ‘7. AUTHORITARIAN CONSTITUTIONALISM IN FACEBOOKLAND’, *The Perilous Public Square* (Columbia University Press 2020) <https://www.degruyterbrill.com/document/doi/10.7312/poze19712-008/html> accessed 25 April 2025

‘Proposition de loi, n° 1785’ <https://www.assemblee-nationale.fr/dyn/15/textes/l15b1785_proposition-loi> accessed 7 June 2025

‘Quasi-Municipal Corporation’ (*LII / Legal Information Institute*) <https://www.law.cornell.edu/wex/quasi-municipal_corporation> accessed 18 April 2025

Real MD, ‘Breaking Algorithmic Immunity: Why Section 230 Immunity May Not Extend to Recommendation Algorithms’ (2024) 99 Wash. L. Rev. Online 1

‘Recommendations | Oversight Board’ (30 April 2024) <https://www.oversightboard.com/recommendations/> accessed 14 June 2025

Roberts ST, ‘Digital Detritus: “Error” and the Logic of Opacity in Social Media Content Moderation’ [2018] First Monday <https://firstmonday.org/ojs/index.php/fm/article/view/8283> accessed 24 April 2025

Rojszczak M, ‘The Digital Services Act and the Problem of Preventive Blocking of (Clearly) Illegal Content’ (2023) 3 Institutiones Administrationis – Journal of Administrative Sciences 44

Rustad M and Koenig T, ‘The Case for a CDA Section 230 Notice-and-Takedown Duty’ (2023) 23 Nevada Law Journal 533

Samsel BH, ‘In U.S. House, Texas Republicans Grill Zuckerberg More on Whether Facebook Is Censoring Conservatives’ (*The Texas Tribune*, 11 April 2018) <https://www.texastribune.org/2018/04/11/house-mark-zuckerberg-grilled-more-texas-republicans-whether-facebook-/> accessed 14 June 2025

Sartor DG, ‘Providers Liability: From the eCommerce Directive to the Future’

Satariano A, ‘E.U. Prepares Major Penalties Against Elon Musk’s X’ *The New York Times* (3 April 2025) <https://www.nytimes.com/2025/04/03/technology/eu-penalties-x-elon-musk.html> accessed 4 June 2025

Sauer P, ‘Russia Bans Facebook and Instagram under “Extremism” Law’ *The Guardian* (21 March 2022) <https://www.theguardian.com/world/2022/mar/21/russia-bans-facebook-and-instagram-under-extremism-law> accessed 15 June 2025

Scott M, ‘Europe Presses American Tech Companies to Tackle Hate Speech’ *The New York Times* (6 December 2016) <https://www.nytimes.com/2016/12/06/technology/europe-hate-speech-facebook-google-twitter.html> accessed 11 June 2025

Singel R, 'YouTube Agrees To Help Government Censors' *Wired* <https://www.wired.com/2007/04/youtube-agrees-/> accessed 28 May 2025

Staff W, 'Google Bends to China's Will' *Wired* <https://www.wired.com/2006/01/google-bends-to-chinas-will/> accessed 28 May 2025

'State Action Doctrine' (*Oxford Constitutions*) <https://oxcon.ouplaw.com/display/10.1093/law-mpeccol/law-mpeccol-e473> accessed 18 April 2025

Suzor NP, *Lawless: The Secret Rules That Govern Our Digital Lives* (Cambridge University Press 2019)

'Tech & Terrorism: Germany Fines Telegram For Failing To Comply With Online Content Moderation Law' (*Counter Extremism Project*) <https://www.counterextremism.com/press/tech-terrorism-germany-fines-telegram-failing-comply-online-content-moderation-law> accessed 4 June 2025

'The Struggle for Trust Online' (*Freedom House*) <https://freedomhouse.org/report/freedom-net/2024/struggle-trust-online> accessed 16 June 2025

Toor A, 'France Can Now Block Suspected Terrorism Websites without a Court Order' (*The Verge*, 9 February 2015) <https://www.theverge.com/2015/2/9/8003907/france-terrorist-child-pornography-website-law-censorship> accessed 7 June 2025

Trengove M and others, 'A Critical Review of the Online Safety Bill' (2022) 3 Patterns <https://www.cell.com/patterns/abstract/S2666-3899(22)00147-7> accessed 9 June 2025

Tuchtfeld E, 'Be Careful What You Wish For' [2023] Verfassungsblog <https://verfassungsblog.de/be-careful-what-you-wish-for/> accessed 10 May 2025

Tuovinen J, 'The Meta Oversight Board in the Trump Era' [2025] Verfassungsblog <https://verfassungsblog.de/the-meta-oversight-board-in-the-trump-era/> accessed 14 June 2025

'Turkey: Freedom on the Net 2023 Country Report' (*Freedom House*) <https://freedomhouse.org/country/turkey/freedom-net/2023> accessed 28 May 2025

'User Content Moderation under the Digital Services Act – 10 Key Takeaways – Legal Developments' <https://www.legal500.com/developments/thought-leadership/user-content-moderation-under-the-digital-services-act-10-key-takeaways/> accessed 3 June 2025

Vogelsang R, 'The Failure of FOSTA: Unintended Consequences Outweigh Good Intentions' (2023) 44 University of La Verne Law Review 59

Volokh E, 'Cheap Speech and What It Will Do Symposium: Emerging Media Technology and the First Amendment' (1994) 104 Yale Law Journal 1805

'Wampum Belt | Oversight Board' <https://www.oversightboard.com/decision/fb-l1lania7/> accessed 13 June 2025

Wong D and Floridi L, 'Meta's Oversight Board: A Review and Critical Assessment' (2023) 33 Minds and Machines 261

Wu T, 'Is the First Amendment Obsolete?' (2018) 117 Michigan Law Review 547

'X Report: We Remove More Accounts, Suspend Fewer Users Than Twitter Did' <https://www.mediapost.com/publications/article/399766/x-report-we-remove-more-accounts-suspend-fewer-u.html> accessed 9 June 2025

Yoo C, 'Technologies of Control and the Future of the First Amendment' (2011) 53 William & Mary Law Review 747