ALEXANDER PIRANG

Freedom of Expression in the Platform Society

Internet und Gesellschaft

Mohr Siebeck

Internet und Gesellschaft

Schriften des Alexander von Humboldt Institut
für Internet und Gesellschaft

Herausgegeben von

Jeanette Hofmann, Matthias C. Kettemann,
Björn Scheuermann, Thomas Schildhauer
und Wolfgang Schulz

38

Alexander Pirang

Freedom of expression in the platform society

The right to freedom of expression as a constraint on public authorities' power to govern online speech through platform companies

Mohr Siebeck

Alexander Pirang, born 1989; law studies at Bucerius Law School, Hamburg and in Saint Petersburg; legal clerkship at the Berlin Court of Appeal; research associate at the Alexander von Humboldt Institute for Internet and Society; non-resident fellow with the Global Public Policy Institute; since 2021 policy officer at the Federal Ministry for Economic Affairs and Climate Action; 2023 doctorate (University of Hamburg).

Open Access gefördert durch den Fachinformationsdienst (FID) interdisziplinäre Rechtsforschung in Berlin. / Open Access funded by the Fachinformationsdienst (FID) interdisziplinäre Rechtsforschung in Berlin.

ISBN 978-3-16-163962-3 / eISBN 978-3-16-163963-0
DOI 10.1628/978-3-16-163963-0

ISSN 2199-0344 / eISSN 2569-4081 (Internet und Gesellschaft)

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliographie; detailed bibliographic data are available at *https://dnb.dnb.de*.

Published by Mohr Siebeck Tübingen, Germany, 2024. www.mohrsiebeck.com

The book was typeset using Times typeface and printed on non-aging paper by Laupp & Göbel in Gomaringen. It was bound by Nädele in Nehren.

Printed in Germany.

Vorwort (Acknowledgements)

Die vorliegende Arbeit wurde von der Fakultät für Rechtswissenschaft der Universität Hamburg im August 2023 als Dissertation angenommen. Rechtsprechung und Literatur konnten bis zur Einreichung im Januar 2023 umfassend berücksichtigt werden. Bis Februar 2024 veröffentlichte Literatur wurde zum Teil ergänzend aufgenommen.

Herzlich danken möchte ich zuallererst Professor Dr. Wolfgang Schulz, der mich mit wertvollen Denkanstößen wiederholt auf den richtigen Weg gebracht und mit seiner humorvollen und wertschätzenden Art für notwendige Motivationsschübe gesorgt hat. Einen besseren Doktorvater hätte ich mir nicht wünschen können.

Großer Dank gebührt auch Professor Dr. Hans-Heinrich Trute, der die Mühe des Zweitgutachtens auf sich genommen hat. Den Herausgeber:innen Prof. Dr. Jeanette Hofmann, Prof. Dr. Matthias C. Kettemann, Prof. Dr. Björn Scheuermann, Prof. Dr. Thomas Schildhauer und Prof. Dr. Wolfgang Schulz danke ich für die Aufnahme der Arbeit in die Schriftenreihe Internet und Gesellschaft. Dem Fachinformationsdienst für internationale und interdisziplinäre Rechtsforschung danke ich für die großzügige Förderung der Open-Access-Publikation.

Die Arbeit ist in substanziellen Teilen am Alexander von Humboldt Institut für Internet und Gesellschaft entstanden. Das interdisziplinäre und inspirierende Umfeld am Institut hat mich dazu ermutigt, in meiner Dissertation auch über den juristischen Tellerrand hinaus zu blicken.

Dass das mitunter einsame Unterfangen einer Doktorarbeit durch anregende Gespräche aufgelockert wurde, habe ich meinen Freund:innen zu verdanken. Thorsten Benner und Klaas Eller haben mich entscheidend darin bestärkt, das Projekt Doktorarbeit anzugehen. Tom Langerhans hat die Promotionsphase auf seine geradlinige und kluge Art am engsten begleitet.

Meinen Eltern danke ich für Ihre unbedingte Loyalität und ihren Glauben an mich.

Der größte Dank gebührt meiner Frau Anna, die mir auch nach der Geburt unserer Kinder die Freiräume, die eine zeitintensive wissenschaftliche Arbeit erfordert, ermöglicht hat. Ohne diese, keinesfalls selbstverständliche Unterstützung

und ihre wortreichen Ermahnungen, irgendwann zum Ende zu kommen, hätte die Arbeit nicht gelingen können.

Meine Töchter Aglaia und Flora führen mir tagtäglich vor Augen, wie wichtig es ist, mit Neugier und Interesse auf die Welt zu blicken. Ihnen ist diese Arbeit gewidmet.

Berlin im März 2024 *Alexander Pirang*

Chapter breakdown

Table of contents

List of figures and tables

All figures and tables were developed by the author.

List of figures

List of tables

List of abbreviations

AG	Advocate General of the Court of Justice of the European Union
ARIO	ILC Articles on the Responsibility of International Organizations
ARSIWA	ILC Articles on Responsibility of States for Internationally Wrongful Acts
AVMSD	Audiovisual Services Directive
CDA	Communication Decency Act
ChFR	Charter of Fundamental Rights of the European Union
CJEU	Court of Justice of the European Union
DMCA	Digital Millennium Copyright Act
DSA	Digital Services Act Regulation
DSMD	Copyright in the Digital Single Market Directive
EC	European Communities
ECD	E-Commerce Directive
ECHR	European Convention on Human Rights
ECtHR	European Court of Human Rights
EU	European Union
GDPR	General Data Protection Regulation
GIFCT	Global Internet Forum to Counter Terrorism
GNI	Global Network Initiative
ICCPR	International Covenant on Civil and Political Rights
ICJ	International Court of Justice
ILC	International Law Commission
InfoSoc Directive	Directive on the harmonisation of certain aspects of copyright and related rights in the information society.
IP	Internet protocol
KoPl-G	Austrian Communication Platforms Act
NetzDG	German Network Enforcement Act
OECD	Organisation for Economic Co-operation and Development
TERREG	Regulation on addressing the dissemination of terrorist content online
TEU	Treaty on European Union
TFEU	Treaty on the Functioning of the European Union
UN	United Nations
UrhDaG	Act on the Copyright Liability of Online Content Sharing Service Providers

Introduction

I. Setting the stage

Platform companies hold unprecedented power in today's platform society.[1] Just take the deplatforming of Donald Trump by X[2] and other platform companies in the days following 6 January 2021.[3] Given the threat of further violence, this was certainly the right call to make – and arguably long overdue, considering Trump's history of spewing hate and misinformation online.[4] Still, questions lingered whether this kind of power should be in the hands of private corporations at all, and whether societies should continue to depend on platform companies "doing the right thing if and when they wish, independently of any rules and democratic accountability."[5] These concerns only became more pressing once Elon Musk reinstated Donald Trump's X account in November 2022, within weeks of taking over the platform, despite earlier promises not to make "major content decisions or account reinstatements"[6] before the formation of a new content moderation council.[7]

---

[1] On the notion of a platform society, see van Dijk, Poell, and de Waal, *The Platform Society*, who refer to "a connective world where platforms have penetrated the heart of societies – disrupting markets and labor relations, transforming social and civic practices, and affecting democratic processes."; see generally Gillespie, *Custodians of the Internet*, 14, who notes that we are today, by and large, speaking from platforms.

[2] Previously Twitter.

[3] On 6 January 2021, Donald Trump incited a mob to storm the US Capitol in an effort to thwart the election of Joe Biden as US President. For an overview of the actions taken by different companies, see Chrichton, "The Deplatforming of President Trump: A Review of an Unprecedented and Historical Week for the Tech Industry."

[4] Cf. Goldberg, "The Scary Power of the Companies That Finally Shut Trump Up"; cf. Kuczerawy, "Does Twitter Trump Trump? A European Perspective."

[5] See Floridi, "Trump, Parler, and Regulating the Infosphere as Our Commons," 1.

[6] Elon Musk, Tweet, 28 October 2022, https://twitter.com/elonmusk/status/1586059953311137792 ("Twitter will be forming a content moderation council with widely diverse viewpoints. No major content decisions or account reinstatements will happen before that council convenes.").

[7] Cf. Kopps and Katzenbach, "Turning Back Time for Content Moderation? How Musk's Takeover Is Affecting Twitter's Rules."

Against this background, the normative argument underpinning the thesis is that decisions about what counts as acceptable in the online speech environment should not be taken by platform companies alone. Content moderation, to put it bluntly, is not a private matter.[8] Instead, public actors with democratic legitimacy need to play a role in setting the objectives of online speech governance. The DSA and other regulatory efforts in the EU are therefore welcome in principle. It is also generally defensible that public authorities are coopting platform companies to restrict illegal online content; coping with the sheer amount of online speech could hardly be accomplished without some degree of public-private cooperation.[9]

Yet, we also need to be mindful of the potential pitfalls of such regulatory interventions. "Determining how and where to regulate [online] speech is among the most important, and most delicate, tasks a government may undertake," Michael Karanicolas stressed, seeing that "[i]t requires a careful balancing between removing harmful content while providing space for controversial and challenging ideas to be aired, and deterring dangerous speech while minimizing a broader chilling effect that impacts legitimate areas of debate."[10] Shifting the responsibility to police harmful content to platform companies, as public authorities are increasingly doing, may therefore create its own challenges.[11] Although such regulatory approaches are designed to advance public policy objectives, they may actually reduce the authority of state actors in the long term.[12] As David Kaye pointed out, "the pressure on companies has led to an outsourcing of public roles to private actors, which amounts to an expansion of corporate power instead of constraints on it."[13]

More specifically, delegating the task of online speech control to platform companies carries risks for freedom of expression.[14] As Hannah Bloch-Wehba noted,

> "rather than simply compelling intermediaries to delete specific content, governments are foisting upon platforms increasing responsibility for making *legal determinations* regarding speech – a task that might previously have belonged to a court, administrative agency, or other government body accountable to the public."[15]

---

[8] Eder, "Making Systemic Risk Assessments Work," 8.

[9] Cf. Land, "Against Privatized Censorship," 395.

[10] Karanicolas, "Squaring the Circle Between Freedom of Expression and Platform Law," 177 f.

[11] Cf. Frosio, "From Intermediary Liability to Responsibility."

[12] See Land, "Against Privatized Censorship," 373, fn. 40.

[13] Kaye, *Speech Police: The Global Struggle to Govern the Internet*, 20.

[14] Cf. Husovec, "(Ir)Responsible Legislature? Speech Risks under the EU's Rules on Delegated Digital Enforcement," 3.

[15] Bloch-Wehba, "Global Platform Governance," 31 f.

In this process, online speech governance risks getting warped into something that does not even remotely resemble traditional speech regulation.[16] Jack Balkin warned that users "get no judicial determination of whether their speech is protected or unprotected when companies block, censor, or take down their speech."[17] Instead, he continued, "some nontransparent form of private governance or bureaucracy serves as prosecutor, judge, jury, and executioner."[18] In this context, platform companies have been criticized for their overly vague and opaque content policies and their over-reliance on automation.[19]

It has also been extensively argued that platform companies' motivation to avoid liability, stricter regulation, or administrative fines may incentivize them to err on the side of caution and to systematically restrict legitimate speech.[20] Regulatory efforts to address legitimate concerns about unlawful online speech may thus inadvertently erode users' freedom to impart and receive information and ideas online. In other words, there is a risk of a zero-sum outcome where tackling bad content collaterally hurts good speech.[21]

Moreover, there are concerns – central to this thesis – that delegating speech control responsibilities to platform companies might allow public authorities to bypass their fundamental rights obligations.[22] James Boyle recognized this potential more than two decades ago, when he foresaw that states would rely on privatization to regulate the internet:

> "[O]ne would want to escape from the practical and legal limitations of a sovereign-citizen relationship. Thus, one might seek out private actors involved in providing Internet services who are not quite as mobile as the flitting and frequently anonymous inhabitants of cyberspace. […] By enlisting these nimbler, technologically savvy players as one's private police, one would also gain another advantage: freedom from some of the constitutional and other restraints that would burden the state were it to act directly."[23]

---

[16] Cf. Balkin, "Free Speech Is a Triangle," 2028–32, who describes the result of the connection between public and private ordering in online speech governance as "privatized bureaucracy."

[17] Balkin, 2031.

[18] Balkin, 2031.

[19] See, for instance, Kaye, "Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression."

[20] Council of Europe, "Comparative Study on Blocking, Filtering and Take-down of Illegal Internet Content," 30; Keller, "Who Do You Sue? State and Platform Hybrid Power Over Online Speech," 5; Frosio, "Mapping Online Intermediary Liability," 26.

[21] For a critical take on this narrative, see Woods, "Online Harms: Why We Need a Systems-Based Approach Towards Internet Regulation."

[22] See Jørgensen and Pedersen, "Online Service Providers as Human Rights Arbiters," 180; cf. Land, "Against Privatized Censorship," 395.

[23] Boyle, "Foucault in Cyberspace: Surveillance, Sovereignty, and Hardwired Censors," 197.

As Hannah Bloch-Wehba pointed out, this strategy of compelling platform companies to "instantiate and enforce *public* policy preferences" through *private* ordering converts "what might otherwise be private action into heterodox, hybrid public-private governance arrangements in which state and private power are commingled."[24] The resulting blurring of boundaries between private ordering and public regulation makes it difficult to hold public authorities responsible for adverse effects on freedom of expression.[25] This raises the question whether public authorities can really break free from "pesky constitutional constraints" when using platform companies as private surrogates.[26]

II. Hypothesis and research questions

The research hypothesis is that public authorities have an obligation to abstain from violating users' right to freedom of expression when regulating platform companies' content moderation processes, regardless of whether or not this regulatory activity directly restricts online speech. In particular, this obligation entails that any limitations of users' freedom of expression must be justified. The overall research question that motivates this thesis therefore concerns the function of the right to freedom of expression as a constraint on public power: What limits do Art. 10 ECHR and Art. 11 ChFR impose on public authorities' power to regulate how platform companies moderate online content?

This research question can be broken down into several separate aspects relating, respectively, to content moderation, platform regulation, and platform users' right to freedom of expression. In developing an answer to this query, several sub-questions, which individually address specific aspects of the overall research question, will guide the analysis. With respect to platform companies' *content moderation*, the thesis tackles the following questions:

- What does content moderation entail?
- What role does context play in content moderation?
- To what extent does content moderation impact users' freedom of expression?
- How do different incentives and constraints influence platform companies' decision-making regarding content moderation?

---

[24] Bloch-Wehba, "Global Platform Governance," 30.

[25] Cf. Jørgensen and Pedersen, "Online Service Providers as Human Rights Arbiters," 186 f.

[26] See Boyle, "A Nondelegation Doctrine for the Digital Age?," 10.

As regards *platform regulation*, the thesis aims to address the following questions:

– How has platform regulation evolved in the EU?
– What are the rationales behind platform regulation?
– How does platform regulation interact with other influences on platform companies' decision-making?
– How does platform regulation operate in practice?

Lastly, the thesis explores the following questions regarding users' *right to freedom of expression*:

– To what extent does platform regulation allow public authorities to launder state action through platform companies' private ordering?
– To what extent does the case law of the ECtHR and the CJEU on causation and attribution provide avenues for overcoming these risks?
– How can limitations of users' right to freedom of expression be established where public authorities require platform companies to moderate content?
– To what extent are public authorities obligated to minimize the risks of overblocking in the context of platform regulation in order to justify such limitations?

III. Research scope

In the following, I will outline the research scope with respect to online platforms, content moderation, platform regulation, and the relevant fundamental rights framework.

1. Content moderation by online platforms

The thesis focusses on content moderation at the application layer of the tech stack. This calls for a more detailed outline of the "what" (content moderation), the "who" (online platforms), and the "where" (application layer).

a) Online platforms

To start with the "who," the thesis focuses on online platforms for user-generated or user-uploaded content (also referred to as user speech or online speech in the following). Attempts to further define these entities – often vaguely characterized as social media – are made difficult by the ambiguous nature of the term platform, which may also be understood slightly differently in various disci-

plines.[27] It follows that "there is no consensus on a single definition of [online platforms], neither in computer science nor in the economic and legal domain."[28] Moreover, regulatory frameworks such as the DSA and the DSMD advanced several new legal concepts related to the notion of online platforms, which reshape and add to pre-existing legal concepts.[29]

Against this background, the thesis opts for a broad perspective on online platforms. Selectively drawing on existing definitions, online platforms are understood in the thesis as providing three central affordances:

- the technological intermediation of user-generated or -uploaded content,
- the possibility of interactivity among different users and of direct engagement with content,
- the possibility for users to carry out specific activities.[30]

Naturally, the thesis focusses on online platforms dealing with content moderation questions – although this hardly narrows the scope, since content moderation turns up in unexpected places, from knitting forums[31] to porn sites[32] to the Metaverse.[33] The research scope therefore covers a diverse set of actors, including, but not limited to,

- social media platforms characterized by network connections between users (e. g., Facebook, LinkedIn),
- online media sharing platforms allowing the upload or the livestreaming of content, including images and video (e. g., YouTube, Twitch), music (e. g., Spotify, SoundCloud), and text (e. g., Medium),
- discussion forums allowing users to hold conversations (e. g. Reddit),
- messaging platforms allowing users to communicate and share online content privately (e. g., Telegram),

---

[27] For a concise overview of the term and its history, see Gorwa, "What Is Platform Governance?," 3; cf. Schwarz, "Platform Logic: An Interdisciplinary Approach to the Platform-Based Economy," 3; see generally Gillespie, "The Politics of 'Platforms'" (on the strategic use of the platform metaphor).

[28] Bertolini, Episcopo, and Cherciu, "Liability of Online Platforms," 7 f.

[29] Cf. Quintais et al., "Copyright Content Moderation in the EU," 44–51 (with an instructive overview of the EU law terminology).

[30] See DeNardis and Hackl, "Internet Governance by Social Media Platforms," 2; see also Quintais et al., "Copyright Content Moderation in the EU," 52 f. (for an overview of different definitions of online platforms proposed in the literature).

[31] See Copia Institute, "Content Moderation Case Study: Knitting Community Ravelry Bans All Talk Supporting President Trump (2019)."

[32] See, for instance, Meineck and Alfering, "We Went Undercover in xHamster's Unpaid Content Moderation Team."

[33] See Mak, "I Was a Bouncer in the Metaverse."

– matchmaking and e-commerce platforms facilitating the transaction of goods and services (e. g., eBay, Airbnb) or the matching between different sets of users (e. g., Tinder),
– platforms for ratings and reviews of third-party services or offerings of different kinds (e. g., Yelp), and
– platforms allowing collaborative production (e. g., Wikipedia).[34]

For sake of brevity, I will use the term *platform company* to refer to private enterprises offering any kind of platform services and the term *platform* to refer to those services.[35] I will not focus on online search engines, unless relevant provisions in the regulatory frameworks analyzed in the thesis specifically refer to them.[36]

b) Content moderation

As regards the "what," content moderation (also referred to as content governance, self-regulation, and private ordering[37]) is "a broad concept with fuzzy borders."[38] It generally refers to platform companies' practice of setting rules around speech and enforcing them, usually with a mix of human laborers and automated systems at scale.[39] In this vein, Kate Klonick characterized content moderation as the "industry term for a platform's review of user-generated content posted on its site and the corresponding decision to keep it up or take it down."[40]

The thesis adopts a broader perspective, which is not solely focused on content removal.[41] To this end, it draws on James Grimmelmann's conception of content moderation; Grimmelmann argued that the term should be understood as the "governance mechanisms that structure participation in a community to facilitate

---

[34] See Bertolini, Episcopo, and Cherciu, "Liability of Online Platforms," 17; see van Hoboken et al., "Hosting Intermediary Services and Illegal Content Online," 12–14.

[35] See also Gorwa, "What Is Platform Governance?," 3.

[36] See, for instance, Art. 33 DSA.

[37] See Gorwa, "The Shifting Definition of Platform Governance."

[38] Quintais et al., "Copyright Content Moderation in the EU," 33.

[39] See Gorwa, "The Shifting Definition of Platform Governance."

[40] Klonick, "The Facebook Oversight Board," 2427.

[41] On this broader perspective, see Douek, "Content Moderation as Systems Thinking," 531 ("'Content moderation,' especially but not exclusively at the largest platforms, now includes many more things than it did even a few years ago: increased reliance on automated moderation; sticking labels on posts; partnerships with fact-checkers; greater platform and government collaboration; adding friction to how users share content; giving users affordances to control their own online experience; looking beyond the content of posts to how users behave online to determine what should be removed; tinkering with the underlying dynamics of the very platforms themselves.").

cooperation and prevent abuse."[42] This definition covers the many ways in which platforms influence users' online activities,[43] including the host of design decisions that structure content flows and user interactions.[44] It underlines that individuals' capability for speech, online as offline, is shaped through various and evolving constraints and affordances.[45] This is echoed by the DSA's definition of content moderation, which includes

> "measures taken that affect the availability, visibility, and accessibility of that illegal content or that information, such as demotion, demonetisation, disabling of access to, or removal thereof, or that affect the ability of the recipients of the service to provide that information, such as the termination or suspension of a recipient's account."[46]

Against this background, the term content moderation is used in the thesis to refer to the development of rules regarding online speech, the institutional processes of identifying, adjudicating, and sanctioning content, and, lastly, redress mechanisms that allow users to appeal specific enforcement outcomes.[47] As in the DSA, this includes algorithmic downranking and other restrictions that stop short of removing content, but only to the extent to which they are used as a means to sanction illegal, prohibited, or unwanted online behavior.[48] On the other hand, the thesis will not bring demonetization, as another means to sanction content,[49] into focus. By using the term *content* moderation, I do not mean to exclude restrictions that are not directly applied to individual pieces of content, such as account-level restrictions.

---

[42] Grimmelmann, "The Virtues of Moderation," 47; see also Gorwa, "The Shifting Definition of Platform Governance."

[43] Gorwa, "The Shifting Definition of Platform Governance."

[44] Cf. Gorwa; cf. Gorwa, Binns, and Katzenbach, "Algorithmic Content Moderation," 3.

[45] Cf. Bietti, "Free Speech Is Circular."

[46] Art. 3(t) DSA.

[47] See, for a similar definition, Gillespie et al., "Expanding the Debate About Content Moderation," 2.

[48] De-prioritizing certain pieces of content based on engagement metrics therefore falls outside the scope, whereas demoting content due to its harmful properties is covered by the definition. For a similar discussion on the relation between the concepts of content moderation and content recommending in the DSA context, see Quintais et al., "Copyright Content Moderation in the EU," 35 f.

[49] See Llansó et al., "Artificial Intelligence, Content Moderation, and Freedom of Expression," 18; see generally Caplan and Gillespie, "Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy."

I will focus on *industrial, commercial* content moderation[50] by what the DSA refers to as very large online platforms,[51] which typically relies on complex governance structures with tens of thousands of human reviewers as well as large-scale automated systems.[52] In this sense, content moderation is understood in the thesis "as a project of mass speech administration."[53] This does not only include individual speech restrictions carried out by frontline moderators but also measures at what I refer to as the systems level of content moderation, including design decisions that structure the overall content moderation process.[54]

By contrast, I will not specifically analyze *artisanal* approaches, involving case-by-case review of content by human moderators on a smaller scale (think smaller instances on Mastodon), and *community-reliant* approaches, which commonly combine top-down policy decisions by company staff with a larger group of volunteer human reviewers (think Wikipedia).[55] I will also primarily focus on content moderation at the individual-firm level. That said, it is important to recognize that there are many overlaps between platform companies' content moderation processes, from informal exchange between the major US platform companies,[56] to collective self-regulatory bodies such as the GNI,[57] to the formation of what Evelyn Douek called content cartels,[58] such as the joint GIFCT database for terrorist content.[59]

Lastly, three important areas are excluded from the thesis entirely: Practices of data extraction and accumulation associated with the business models of platform companies fall outside the research scope. The thesis also does not comment on platform affordances that allow users to moderate content themselves.[60]

---

[50] See Caplan, "Content or Context Moderation?," 6; see Roberts, "Digital Refuse: Canadian Garbage, Commercial Content Moderation and the Global Circulation of Social Media's Waste," 6 f.

[51] On the notion of very large online platforms (and very large online search engines), see Art. 33 DSA.

[52] See Caplan, "Content or Context Moderation?," 15–25.

[53] On the notion of content moderation bureaucracies, see Douek, "Content Moderation as Systems Thinking."

[54] Cf. Douek, 545–48 (on the role of design and affordances in content moderation).

[55] Caplan, "Content or Context Moderation?," 15–25.

[56] See Gillespie et al., "Expanding the Debate About Content Moderation," 5.

[57] The GNI was founded in 2008 by a coalition of different platform companies, academics and NGOs with the objective to both coordinate resistance to censorship requests by authoritarian states and to respond to criticisms levied at platforms for accommodating such demands, see Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism*, 241.

[58] Douek, "The Rise of Content Cartels."

[59] See below, Ch. 1, Sect. I.3.a.(2)(a).

[60] On the challenges of allowing political parties to moderate debates in the comments sec-

Lastly, I will not consider online ads as another category of content hosted and moderated by platform companies.

c) Application layer

With respect to the "where," the thesis focuses on content moderation at the top level of the tech stack,[61] also sometimes referred to as the application layer.[62] This top level makes up what users experience as the internet, broadly speaking, and it is where platform companies operate.[63] Measures taken by actors further down the tech stack – such as cloud service providers, content delivery networks, domain registrars, and internet service providers[64] – therefore fall outside the research scope.

Still, it is worth noting that there can be an important overlap between the different levels of the tech stack, which may also impact content moderation decisions taken at the top level.[65] Indeed, we have seen app store providers pressure platform companies and, in some cases, even remove platforms entirely from their ecosystems for a perceived failure to ensure adequate content moderation.[66] These actions often do not follow clear standards or procedures and tend to be highly opaque,[67] even though they may have far-reaching consequences.[68]

2. Platform regulation

Platform regulation is used in the thesis to refer to regulation directed at content *moderation* rather than at specific pieces of content. Put differently, the thesis

---

tions and to hide problematic speech, see generally Kalsnes and Ihlebæk, "Hiding Hate Speech: Political Moderation on Facebook."

[61] For an overview of the tech stack's different levels, see Donovan, "Navigating the Tech Stack: When, Where and How Should We Moderate Content?"

[62] The application layer forms part of the open systems interconnection (OSI) model developed by the International Organization for Standardization, see Cloudflare, What is the OSI model?, https://www.cloudflare.com/learning/ddos/glossary/open-systems-interconnection-model-osi/.

[63] Cf. Donovan, "Navigating the Tech Stack: When, Where and How Should We Moderate Content?"

[64] This list is drawn from Joan Donovan's overview of the different actors at various levels of the tech stack, see Donovan.

[65] Cf. Gillespie et al., "Expanding the Debate About Content Moderation," 6 f.

[66] Cf. Gillespie et al., 7.

[67] For a first hand account of the influence of Apple and Google on Twitter, see Roth, "I Was the Head of Trust and Safety at Twitter. This Is What Could Become of It."

[68] Cf. Kuczerawy, "Freedom of Expression in the Era of Online Gatekeeping," 284 f. (on the need for safeguards against premature measures by actors down the tech stack).

Index