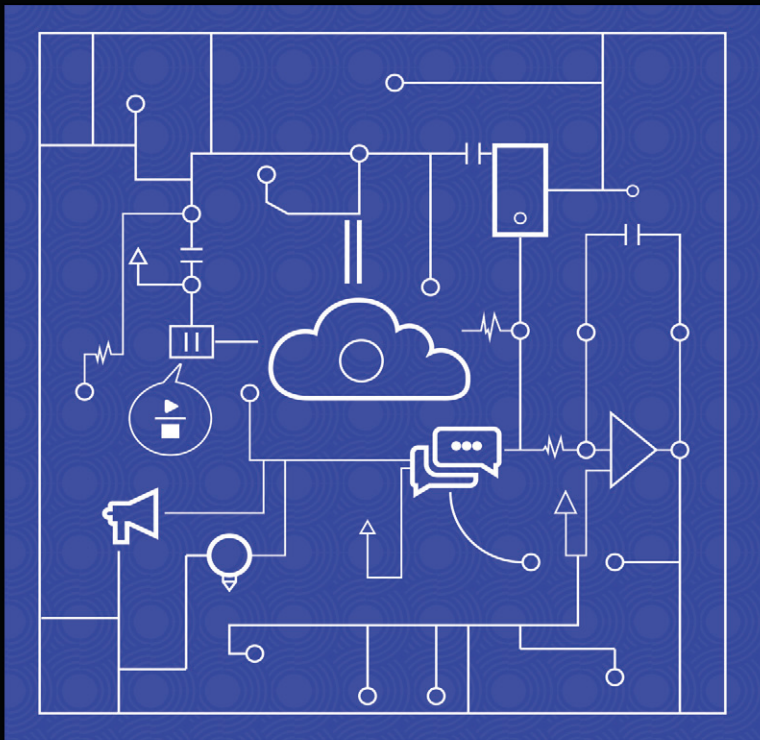


# Platform Governance and Social Justice

*Governing Hate Speech On Social Media*



*Paloma Viejo Otero*

**DIGITAL ACTIVISM AND SOCIETY**



# PLATFORM GOVERNANCE AND SOCIAL JUSTICE

*In an age where digital platforms wield unprecedented influence over our lives, Platform Governance and Social Justice presents a compelling case of critical inquiry and theoretical innovation. In this groundbreaking work, Paloma Viejo Otero delves deep into the intricate web of power dynamics and ideological frameworks that underpin platform governance and moderation systems.*

*At its core, the strength of this book lies in its original theoretical approach, which skilfully unpacks the genealogy of platform governance. By tracing the historical roots and ideological underpinnings of these systems, Viejo Otero offers readers a profound understanding of how platforms have come to control and steer online discourse and user behaviour.*

*Platform Governance and Social Justice prises open and questions the power structures and systems of control that platforms have designed and implemented. Through meticulous analysis and incisive critique, the book exposes the often-hidden mechanisms through which platforms shape our digital experiences and influence societal norms.*

*What sets this book apart is its comprehensive examination of the digital platform ecosystem, considering both mainstream and 'Alt Tech' platforms. In essence, Platform Governance and Social Justice offers a timely and indispensable contribution to the field. With its rigorous scholarship, nuanced insights, and thought-provoking analysis, it constitutes an original contribution to ongoing debates and will inspire further research in the realms of platform governance and critical digital media studies. Viejo Otero has crafted an important work that not only sheds light on the present state of digital platforms but also offers a roadmap for envisioning more equitable forms of digital governance. The book will be indispensable for both advanced undergraduate and postgraduate students as well as for researchers in platform governance.*

—Professor Dr Eugenia Siapera,  
Information and Communication Studies,  
University College Dublin, Ireland

*We have been waiting for this book for years. The debate on the role and responsibility of platform for hate speech has been waging widely and wildly. This book finally lays the foundations to hold this debate in informed ways. What are our concepts of hate speech, and how do they relate to freedom, security and justice? And how have platforms been able to shape and shift our understanding of these concepts? Viejo Otero unfolds this trajectory with vivid language and a sharp eye for the complexities and power structures involved in this debate. This book, with its typology of approaches of hate speech regulation, will be the key reference for research on hate speech and platform governance in the coming years.*

—Professor Dr Christian Katzenbach,  
ZeMKI, University of Bremen, Germany

# DIGITAL ACTIVISM AND SOCIETY: POLITICS, ECONOMY AND CULTURE IN NETWORK COMMUNICATION

*The Digital Activism and Society: Politics, Economy and Culture in Network Communication* series focuses on the political use of digital everyday-networked media by corporations, governments, international organisations (Digital Politics), as well as civil society actors, NGOs, activists, social movements and dissidents (Digital Activism) attempting to recruit, organise and fund their operations, through information communication technologies.

The series publishes books on theories and empirical case studies of digital politics and activism in the specific context of communication networks. Topics covered by the series include but are not limited to:

- the different theoretical and analytical approaches of political communication in digital networks;
- studies of sociopolitical media movements and activism (and ‘hacktivism’);
- transformations of older topics such as inequality, gender, class, power, identity and group belonging;
- strengths and vulnerabilities of social networks.

Series Editor

Dr Athina Karatzogianni

About the Series Editor

Dr Athina Karatzogianni is an Associate Professor at the University of Leicester, UK. Her research focuses on the intersections between digital media theory and political economy, in order to study the use of digital technologies by new sociopolitical formations.

Published Books in This Series

*Digital Materialism: Origins, Philosophies, Prospects* by Baruch Gottlieb

*Nirbhaya, New Media and Digital Gender Activism* by Adrija Dey

*Digital Life on Instagram: New Social Communication of Photography* by Elisa Serafinelli

*Internet Oligopoly: The Corporate Takeover of Our Digital World* by Nikos Smyrnaiois

*Digital Activism and Cyberconflicts in Nigeria: Occupy Nigeria, Boko Haram and MEND* by Shola A. Olabode

*Platform Economics: Rhetoric and Reality in the 'Sharing Economy'* by Cristiano Codagnone

*Communication as Gesture: Media(tion), Meaning, & Movement* by Michael Schandorf

*Digital Media and the Greek Crisis: Cyberconflicts, Discourses and Networks* by Ioanna Ferra and Athina Karatzogianni

*Journalism and Austerity: Digitization and Crisis During the Greek Memoranda* by Christos Kostopoulos

*The Emerald Handbook of Digital Media in Greece: Journalism and Political Communication in Times of Crisis* by Anastasia Veneti and Athina Karatzogianni

*Protest Technologies and Media Revolutions* by Athina Karatzogianni, Michael Schandorf and Ioanna Ferra

*Posthumanism in digital culture: Cyborgs, Gods and Fandom* by Callum T.F. McMillan

*Chinese Social Media: Face, Sociality, and Civility* by Shuhan Chen and Peter Lunt

*Media, Technology and Education in a Post-Truth Society: From Fake News, Datafication and Mass Surveillance to the Death of Trust* by Alex Grech

*3D Printing Cultures, Politics and Hackerspaces* by Leandros Savvides

*Environmental Security in Greece: Perceptions from industry, government, NGOs and the public* by Charis(Harris) Gerosideris

*Fantasy, Neoliberalism and Precariousness: Coping Strategies in the Cultural Industries* by J r my Vachet

*Crisis Communication in China: Strategies taken by the Chinese Government and Online Public Opinion* by Wei Cui

*Digital Politics, Digital Histories, Digital Futures: New Approaches for Historicising, Politicising and Imagining the Digital* by Adi Kuntsman and Liu Xin

*Digital Memory in Brazil: A Fragmented and Elastic Negationist Remembrance of the Dictatorship* by Leda Balbino

*Duty to Revolt: Transnational and Commemorative Aspects of Revolution* by George Souvlis and Athina Karatzogianni

*Organisation and Governance Using Algorithms* by Ioannis Avramopoulos  
*Fractal Leadership* by Athina Karatzogianni and Jacob Matthews

*Future Feminisms: Biolabour, Technofeminist Care, and Transnational Strategies* by Ioanna Ferra, Fenia Ferra, Korinna Patelis and Athina Karatzogianni

*Greece in the 1940s: Occupation and Civil War in Digital Culture, Screen Media, and the Arts* by Ioanna Ferra and Leandros Savvides

Forthcoming Titles

*Visual Misogyny: Platformed Politics of Visual Gendered Hate* by Patricia Prieto-Blanco and Suay Melisa Özkula

*Untangling Platform Power: The Oppositional Affordances of Data Activism* by Venetia Papa

*Massively Marginal: Kuaishou as China's Subaltern Platform* by Dino Ge Zhang, Jian Xu and Gabriele de Seta

*The Road to Neo-Feudalism: Syriza, Melancholy and the Future* by Korinna Patelis

# PLATFORM GOVERNANCE AND SOCIAL JUSTICE

Governing Hate Speech on Social  
Media

BY

**PALOMA VIEJO OTERO**

*ZeMKI, Centre for Media, Communication and  
Information Research, University of Bremen, Germany*



United Kingdom – North America – Japan – India  
Malaysia – China

Emerald Publishing Limited  
Emerald Publishing, Floor 5, Northspring, 21-23 Wellington Street, Leeds LS1 4DL

First edition 2025

Copyright © 2025 Paloma Viejo Otero.  
Chapter 6 © 2025 Paloma Viejo Otero and Erika Irusta.  
Published under exclusive licence by Emerald Publishing Limited.

**Reprints and permissions service**

Contact: [www.copyright.com](http://www.copyright.com)

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters' suitability and application and disclaims any warranties, express or implied, to their use.

**British Library Cataloguing in Publication Data**

A catalogue record for this book is available from the British Library

ISBN: 978-1-83797-105-3 (Print)  
ISBN: 978-1-83797-104-6 (Online)  
ISBN: 978-1-83797-106-0 (Epub)



INVESTOR IN PEOPLE

*Dedicado a mi madre Margarita Otero Palacios  
a mi amor François Kervarec,  
a mi hermana Olga Viejo Otero  
y a mis abuelos David Otero Canto y Covadonga Palacios Diaz.*

This page intentionally left blank

# CONTENTS

<i>About the Author</i>	xiii
<i>Acknowledgements</i>	xv
1. Studying Platform Governance of Hate Speech	1
2. Unpacking the Debate Over Hate Speech Regulation: Neutrality, Freedom, and Equality	9
3. Approaches to Governing Hate Speech: Neutrality, Absolute Freedom of Expression and Social Justice	27
4. Mainstream Platforms and the Standard of Governance: Neutral Security Approach	41
5. Alt Tech Platforms: Freedom of Expression Absolutism	57
6. Casa Liken: A Platform With a Social Justice Approach	75
7. Platform Governance and Social Justice: Governing Hate Speech on Social Media	93
<i>References</i>	101
<i>Index</i>	113

This page intentionally left blank

## ABOUT THE AUTHOR

**Paloma Viejo Otero** is a Researcher specialising in platform governance and social justice. She received an award from Young Universities for the Future of Europe (YUFE Alliance) to collaborate with Professor Dr Christian Katzenbach's lab at the Centre for Media, Communication, and Information Research (ZeMKI), University of Bremen. At ZeMKI, Paloma is investigating socio-techno-political imaginaries in relation to hate speech, platform governance and social justice.

Paloma's academic background is in communications and social sciences. She earned her PhD in 2022 with a thesis titled 'Governing Hate: Facebook and Hate Speech', under the supervision of Prof Dr Eugenia Siapera at Dublin City University, School of Communication. Her PhD research examined the principles and values of Facebook's hate speech governing apparatus. Prior to her PhD, Paloma completed her undergraduate degree at Universidad Pontificia de Salamanca, an MA in Cultural Studies and Arts Management from the Instituto de Investigación Ortega y Gasset in Madrid and an MPhil in Race, Ethnicity and Conflict from Trinity College Dublin.

In academia, Paloma has worked as a Research Assistant at the DCU Anti-Bullying Centre and as an external collaborator for Universidad Oberta de Catalunya. As a Lecturer, she has taught Cultural Studies at Dublin City University and Platform Governance at University College Dublin. At the University of Bremen, she has taught Transcultural Communications module and has created and facilitated modules such as Hate Speech in Social Media Platforms and Media Practice, including a module called Blueprint: Designing Platforms from a Social Justice Perspective. She also has significant experience in the public and non-profit sectors, including roles with the Spanish Ministry of Foreign Affairs in Sudan and Guatemala, UNESCO, Migrants Rights Centre Dublin and the Peter McVerry Trust in Dublin.

Paloma is the author or co-author of two guides on understanding or combatting hate speech, *Guide to Understand Hate Speech* (2020) and *We can!: Taking Action against Hate Speech through Counter and Alternative*

Narratives (2017) as well as several articles, including ‘Governing Hate: Facebook and Digital Racism’ (2021), co-authored with Prof Dr Eugenia Siapera.

This is Paloma’s first book, representing eight years of research.

# ACKNOWLEDGEMENTS

Firstly, I would like to thank Prof Dr Eugenia Siapera, who has accompanied me and helped this book to mature since 2015. Thank you for teaching me how collaborative academia should be. My most sincere thanks to the editor of this book Prof Dr Athina Karatzogianni for all her insights. Thanks to Prof Dr Christian Katzenbach, whose support and trust in me helped me to secure the YUFE Fellowship at the ZeMKI, University of Bremen, which enabled me to work on this book. I particularly appreciate your support and respect for your lab's individual intellectual freedom, as well as your leadership model, which is rare. Thanks to the ZeMKI, and in particular to Prof Dr Andreas Hepp, for his direction of the institute which hosts originality and academic freedom. Thanks to Prof Dr Gavan Titley, who inspired thoughts for this book and, of course, for your friendship (Church). Thanks to Prof Dr Andrew Finlay for teaching me to think things through thoroughly. And thank you, Prof Dr Ronit Lentin, for being a catalyst in my long-desired academic career. I want to thank all my colleagues at the ZeMKI, with special thanks to Anke Offerhaus and Janina Fadil-Kerstein.

Special thanks to Dr Mattias Ekman, Dr Elizabeth Farries, Prof Dr Roderick Flynn, Prof Dr Debbie Ging, Dr Robert Gorwa, Dr Miira Hill, Dr Tanya Lokot, Dr Ariadna Matamoros-Fernandez, Dr Tijana Milosevic, Prof Dr Kaarina Nikunen, Dr James O'Higgins Norman, Dr Leticia Ortega, Dr Matti Pohjonen, Dr Rebecca Scharlach and Dr Marielle Wijermars.

This page intentionally left blank

# STUDYING PLATFORM GOVERNANCE OF HATE SPEECH

*The study of governance involves the understanding of all possible techniques that intervene in the life of individuals, and the knowledge that shapes the design of those techniques. (Miller & Ross, 2008, p. 30)*

The study of platform governance involves the analysis of a multitude of ‘infinitesimal techniques’ that platforms employ to influence our lives and entails comprehending the knowledge and intentions underlying these techniques. Drawing from Foucauldian studies, ‘infinitesimal techniques’ (Finlay, 2015) encompass every element identifiable within and by platforms, which, arranged in a particular way, lead to a suitable end. From a button on the left side of the screen, passing through the fabric of relationships with stakeholders, to a large set of public statements held by platform leaders, all the elements we can identify are governing techniques that wield power and unveil, for example, the foundational philosophies internally held by platforms, the significance they attribute to user engagement or how they adapt to external regulations. Fundamentally, following Foucault (2007) and Miller and Ross (2008), I argue that identifying all possible techniques created by platforms and understanding the knowledge shaping those techniques reveal the underlying intentions of platforms and the manner in which they govern our lives. It is acknowledged that not all techniques have governing intentions; nevertheless, they govern.

In addition to understanding how platforms govern our lives, we must consider an added layer of complexity, which comprises how platforms have

transformed historical and social public concepts. The digital realm, in general, is expert at converting historical debates or external mandates into operational frameworks, thereby endowing traditional concepts with new meanings. For example, concerning racism, we may refer to ‘Platform Racism’ as identified by Matamoros Fernández (2017), and in the context of the Digital Service Act, we may now discuss ‘Solutions to the Digital Service Act Requirements’ (Clegg, 2023), which suggests that platforms typically incorporate and, more critically, tailor any external regulation to their dynamics. This becomes particularly clear in the context of hate speech regulation, a notion that existed before the advent of platforms but which they quickly adopted and, as I will argue in this book, have significantly altered since 2016. Therefore, comprehending how platforms govern hate speech and its impact on our lives involves identifying all the techniques that platforms use or display to manage hateful content and analysing the underlying principles, values and knowledge that shape these techniques. It also involves understanding the historical evolution of hate speech, as this will help us to grasp the extent to which platforms have transformed and adapted the concept.

Hate speech is historically a contested term, shaped by political and social actors ranging from the Soviet Union to social media companies (Viejo Otero, 2022). The tension which characterises the lack of a universal definition of hate speech stems from the struggle between actors over the meaning and significance of the fundamental values of equality and freedom. It has been the author’s task in this book to uncover how different actors, including media platforms, define these principles.

Regarding the history of the concept of hate speech, there is a general tendency in specialist literature to place the ‘birth’ of the debate between the 1980s and 1990s, on university campuses in the United States. At that time, universities drafted codes of conduct to regulate racial attacks occurring not only in classrooms but also in dormitories, corridors, and cafeterias (Shiell, 2008). These regulations, along with responses from their opponents, sparked extensive discussions and multiple publications that have led today’s authors to identify the 1980s as the origin of the debate. However, this book will argue – or rather, reinforce the argument made by Morsink (1999) – that the discussion on hate speech regulation did not start on US campuses in the 1980s but actually began to gain traction in 1946, during the debates preceding the Declaration of Human Rights. In particular, the debate was initiated by the Soviet Bloc, whose representatives argued for the need to regulate historical privileges. Indeed, as Morsink points out, for the representatives of the Soviet Bloc, the Declaration of Human Rights was seen not so much as a legal document but as a moral compass that would contribute to emancipating

historically oppressed groups, allowing them to achieve the universal levels of equality and freedom the Charter promised. However, the history of capitalism and the profound effects of primitive accumulation, traceable to slavery, the history of the colonies and the domestication of women, set the pace of the conversation. The Western bloc, particularly that of the 1940s, was largely benefiting from these forms of exploitation, and the Soviet Bloc's effort, as we will recount in the following chapter, was fruitless. Instead, in the Article 7 of the Chart, we obtained a diluted regulation against the act of discrimination which avoided discussing hate or history and did not significantly change the situation for those historically oppressed. We therefore learn from this historical moment that the regulation of hate speech was born and developed to the rhythm imposed by the Cold War and continued to follow its beat through various historical periods – the Civil Rights movements of the 1960s, the US campuses of the 1980s, the fall of the Berlin Wall and the Soviet Union in the 1990s and the solidification of the European Union – until the debate resurfaced in 2016. This was the year when social media platforms began to take an active role in the regulation of 'harmful or objectionable content'. It is at this moment that a substantial change occurs: the development of mainstream platforms, and especially the emergence of alternative tech and progressive platforms, has shifted the regulation of hate speech from being influenced by the Cold War to moving to the rhythm of the so-called 'culture wars'.

Platforms have transformed the way we communicate, notably by increasing the frequency of our interactions through individualised systems. This shift has disrupted the traditional timing patterns established by newspapers, radio and television. Thus, the way in which platforms have evolved and the extent of their influence in the social fabric has been the subject of various investigations. Among these, I would highlight the contributions of those associated with Platform Governance Studies. This group of researchers has largely focused on understanding social media platforms as active social entities whose intentions and designs have significantly shaped the characteristics of this new communication pattern, thereby impacting the social fabric.

Researchers in the field of Platform Governance have traced the development of platforms through what can be broadly categorised into three stages. The first, a tentative phase, refers to the emergence of platforms as free and public spaces, embodying ideas that resonate with McLuhan's concept of a global village – a vision of a happy, free and connected world that aimed to transcend the strict confines of the Cold War. The second stage is marked by optimism and capitalism, during which major platforms like Facebook and YouTube expanded and evolved, serving social movements and 'democracy'. The third stage is characterised by decline and criticism and is possibly the

genesis of Platform Governance Studies themselves. This phase looks at platforms as active participants, as actors with agency to address and transform what traditionally were considered public themes, such as hate speech. Therefore, this book can be situated within the ‘third stage’ of Platform Governance Studies. Additionally, as Platform Studies have evolved, scholars have distinguished between governance *on* platforms (referring to user behaviour), governance *of* platforms (what public institutions and legal bodies impose on platforms) (Gorwa, 2019) and governance *by* platforms (how platforms govern our lives). This book fits into the third line of study, with one crucial observation: hate speech regulation introduces its own tradition, language, values and tools to platforms, which, as I will argue, platforms have interpreted and adapted according to their own needs.

In this context and through a series of life events, I have devoted several years to studying the example of Facebook. Facebook, and therefore Meta, has been operating for a long time and Facebook remains a leading platform to this day, accumulating knowledge, establishing relationships with NGOs and academics, accumulating cultural and financial capital, developing know-how and expertise sufficient to enable the company to adapt to external pressures. As a result, Facebook and the cohort of platforms that follow their business model have set a moderation standard for hate speech based on controlling the circulation of content through a supposed balance between security of the platform and freedom of expression, alongside an ethos of universal arithmetic equality (Siapera & Viejo Otero, 2021). It is the reaction and response to this ‘moderation standard’ by Alt Tech platforms that has shifted the debate on hate speech from the Cold War to the ‘culture wars’.

As can be observed, the term ‘culture wars’ appears in quotation marks throughout this book. The term is often misused as a derogatory label, unjustly relegating complex social issues – those that fall outside the traditional domains of politics or labour movements – to a realm of perceived triviality. Indeed, such a label indiscriminately bundles together discussions on gender, sexuality, cultural production and labour struggles, among others, effectively diminishing them to components of a superficial mixture. This serves to obscure the underlying struggles with robust theoretical and political underpinnings in post-structuralism, constructivist theories, queer theory and anti-racism efforts. Ultimately, these movements aim to challenge and transform cultural norms, often clashing with liberal or conservative forces intent on preserving the status quo. Therefore, the ‘culture wars’ need to be referenced with care, as this ultimately acts as a hollow term employed to highlight a superficial engagement with these critical debates by certain conservative sectors, particularly in the digital realm, who treat these discussions merely as

content fodder without a genuine commitment to addressing their implications or driving substantial societal change.

This perspective particularly resonates in the discussion of the final chapter which focuses on Casa Liken, a platform known for its critical stance and leftist techno-political efforts. While at first sight, it might be tempting to categorise Casa Liken within the broader narrative of the ‘culture wars’, I, having authored this book and interviewed its founder Erika Irusta, question the fairness of such an association. Instead, the analysis reveals the nuances of Casa Liken as a platform that critically engages with the digital landscape and, more importantly, with its effects on the social landscape, advocating for respectful dialogue, deep thinking and continued resistance to what constitutes freedom, equality or safety. This additionally leads me to think that perhaps, evidence that the ‘culture wars’ are not as meaningful or profound in the digital space as liberals and conservatives want to claim, is the ironic fact that – from the Cold War to the ‘culture wars’ – there has been no variation in the terms of the debate about hate speech. That is, all actors from the 1940s to the present day appear to be grappling with the same questions, namely: What does hate speech regulation entail, and what should it achieve? Is regulating hate speech a means to facilitate universal equality? Does it pose a threat to freedom of expression? Or is it simply a tool for creating environments free from hostility? These questions remain the same, with one notable difference: whereas discussions of hate speech during the Declaration of Human Rights era, or on US campuses, were based on ethical, philosophical and legal grounds, platforms have shifted the debate on the adoption (or lack thereof) of hate speech regulation into a conversation that focuses on the management and operability of hateful, harmful, illegal content.

The material used for this book comes from various sources, and the studies presented in these pages have been conducted in different stages over the last 8 or 9 years. A significant part of the study on what constitutes hate, or the genealogy I have traced on hate speech, was carried out from 2015 to 2018. Three years of intensive reading resulted in substantive knowledge about the actors in, and terms of, the debate around hate speech regulation, which is addressed in Chapter 2. Additionally, these 3 years, and my study of these issues, have enabled the construction of a typology on approaches to regulating hate speech. This is presented in Chapter 3, the content of which I developed in order to gain greater mental clarity on the concept of hate speech and hate speech regulation. Essentially, I deduced this typology for myself, for my own clarity of mind, but ultimately, it has served to guide this book, and I believe it may be useful for other researchers interested in understanding hate speech regulation.

This book also makes use of the exhaustive analysis of the Facebook platform that I conducted for my doctoral thesis, and which in these pages serves, here and there, as an example of a platform whose evolution needs to be understood in order to comprehend the fragmented landscape of platforms to which we have access today. Additionally, new material created ad hoc for this book includes analyses of Alt Tech platforms, for which I used my own personal research and novel studies produced by colleagues. Most importantly, there is analysis of a progressive platform, Casa Liken. Casa Liken is a Spanish platform about which I have been in conversation with its creator for over a year, through meetings and interviews specifically designed for this book. All of this can be found in the following chapters, which I have structured as follows.

Following this introduction, Chapter 2 unveils the history of hate speech regulation. My analysis focuses on the key actors in the debate and the tensions that were prevalent at the time. As the reader will note, the discussion was not limited to hate speech or regulation per se but conjured up other fundamental terms associated with hate regulation, such as privilege, fascism, discrimination, oppression, superiority, self-identification, racism, sexism and of course freedom, equality and safety. This chapter therefore aims to showcase the term's complexity by unfolding its evolution across five significant historical periods, namely the drafting of the Declaration of Human Rights in the 1940s, the Convention on Minorities in the 1969s, the US campus debates of the 1980s, the consolidation of the European Union in the 1990s, and the adoption of hate speech regulation by mainstream social media platforms circa 2016.

Chapter 3 tackles the complexities surrounding the concept of hate speech and its regulation, aiming to simplify and facilitate the study of hate speech on social media platforms for those interested in delving into it. Therefore, while Chapter 2 exposes its complexity by showing 'all the piano keys', Chapter 3 compresses the concept into what I refer to as 'approaches to hate speech regulation'. Specifically, this exercise consists of grouping the actors who, in different periods, understood hate speech regulation through a common lens. This has resulted in at least three major approaches to hate speech: the Freedom of Expression Absolutist Approach, the Neutral Approach and the Social Justice Approach, along with two derivatives: the European Approach and the Neutral-Security Approach, the latter being characteristic of social networks that have emerged in Silicon Valley. An important point is that this typology, especially as it relates to the Neutral Security Approach, the Freedom of Expression Absolutist Approach and the Social Justice Approach, serves to outline the following chapters.

Chapter 4 examines the creation of a Neutral-Security approach by mainstream platforms. I examine how platforms have adopted neutral stances towards hate speech and assigned it a role in security; I then discuss how these platforms define hate speech in terms of security and how they have established a moderation standard based on the tension between freedom of expression and platform safety. Furthermore, I theorise the elements that lead or mobilise platforms to lean, one comment or user's profile at a time, towards freedom of expression, or the free circulation of content, or towards security, also known as content restriction. These elements are: scarcity, accumulation, diversity and content production. Inspired by Foucault's governance studies, these concepts aim to provide a theoretical foundation for how mainstream platforms govern the control and circulation of content. I note that this theoretical lens is still absent in the field of Platform Governance and the phenomenon of moderation, and thus, Chapter 4 aims to address this gap.

In Chapter 5, I delve into platforms that champion the absolutism of freedom of expression, known as Alt Tech. These platforms position themselves as a 'counter-force' against the norms established by mainstream platforms. However, researchers often regard Alt Tech platforms merely as alternatives. This chapter will not portray Alt Tech as alternatives per se, but rather as manifestations of the same communicative phenomenon that advocates for a distinct interpretation of freedom values, closer to a neo-Christian Freedom than notions of freedom used during US campus debates about hate speech. Indeed, Alt Tech does not alter our communication patterns nor does it deviate from the capitalist framework. In essence, it represents variations of mainstream platforms which prioritise different values, while mutating them into the technological realm. For instance, instead of content moderation, they advocate for unrestricted freedom of expression; in opposition to security measures, they denounce them as acts of censorship; and rather than fostering equality, they adopt universal arithmetical principles, leading to ineffective hate speech management or even, in some instances, its outright neglect. This stands in stark contrast to the concept of freedom of speech, which is extended to extremes that distort its original meaning, associating it with divine rights and the complete politicisation of life – a notion that Hannah Arendt discussed and which serves as a theoretical foundation for this chapter.

We then reach Chapter 6. This chapter is dedicated to the only platform I have managed to identify that follows the Social Justice Approach: Casa Liken, a Spanish feminist and trans-priority inclusive platform which has been operational since 2014. It has successfully integrated, in both essence and structure, an understanding of hate speech focused on equality and a safe space for historically discriminated bodies. The platform emphasises the

importance of these groups being able to express themselves freely in a safe environment, where safety is not simply the absence of hostility but active and conscious care for the members of its community. As the chapter will show, Casa Liken is a modest platform which could have grown significantly, but its developers chose to limit its growth for the benefit of its users. It is a space where the moderation techniques employed accompany users, and the methods utilised across the entire platform are designed to align with an ethos of equality in every governing approach. This commitment ensures that sexual or racial discrimination, as manifested through content, finds no foothold or conditions of possibility within Casa Liken.

Finally, a conclusion chapter. In these last pages, I summarise the ideas on social justice that this book portrays, and with it, I present the results of 8, perhaps 9 years of research.