

5 Digital Constitutionalism and Freedom of Expression

5.1 Expressions in the Algorithmic Society

Freedom of expression is one of the cornerstones of democratic society.¹ This non-exhaustive statement is of particular relevance in the digital age.² In the last twenty years, the Internet has become one of the primary means to exercise rights and freedoms. The possibility to access online services and content ubiquitously has played a critical role in promoting opinions and ideas on a global scale.³ Users can connect with different communities to build social and professional relationships at a global level simply by using a personal device. The global pandemic has revealed the importance of online services to overcome the limits of social distancing.

Nevertheless, this flourishing democratic framework driven by digital communication technologies firmly clashes with the troubling evolution of the algorithmic society where online platforms govern the flow of information online.⁴ By making decisions on expressions, they contribute to shaping the boundaries of freedom of expression in the digital age. More than two billion users are today governed by

¹ Cass Sunstein, *Democracy and the Problem of Free Speech* (The Free Press 1995).

² Jack M. Balkin, 'Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society' (2004) 79(1) *New York University Law Review* 1.

³ Henry Jenkins, *Convergence Culture: Where Old and New Media Collide* (New York University Press 2006).

⁴ Niva Elkin-Koren and Maayan Perel, 'Guarding the Guardians: Content Moderation by Online Intermediaries and the Rule of Law' in Giancarlo Frosio (ed.), *Oxford Handbook of Online Intermediary Liability* (Oxford University Press 2020); Kate Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2018) 131 *Harvard Law Review* 1598; Kyle Langvardt, 'Regulating Online Content Moderation' (2018) 106 *The Georgetown Law Journal* 1353.

Facebook's community guidelines,⁵ and YouTube decides how to host and distribute billions of hours of video each week.⁶

This quantitative consideration just provides a partial picture of power. An oligopoly of private entities organises transnationally online information for profit by using algorithmic technologies.⁷ The organisation of social media news feeds or the results provided by a search engine are only some examples of the role of automated decision-making systems in online content moderation, thus pushing to rethink the public sphere.⁸ The decisions of Facebook and Twitter to block the account of the former president Donald Trump in the aftermath of the violent conflict at Capitol Hill or the Facebook ban of Australian publishers as an answer to the adoption of the News Media and Digital Platforms Mandatory Bargaining Code are just two examples of their power over online information. Since algorithmic technologies are programmed according to the economic and ethical values of online platforms without any involvement of the users, the extent to which freedom of expression is protected is subject to private determinations driven by opaque business purposes.⁹ Even if political and social movements have spread in the digital environment,¹⁰ the governance of online content is increasingly privatised,¹¹ and, therefore, oriented to private purposes which would not lead to putting much hope in the safeguards of democratic values online.¹²

If content moderation plays a crucial role in shaping the boundaries of freedom of expression in the algorithmic society, it is worth wondering how to avoid freedom of expression being subject to opaque private

⁵ Ben Popper, 'A Quarter of the World's Population now Uses Facebook Every Month' *The Verge* (3 May 2017) www.theverge.com/2017/5/3/15535216/facebook-q1-first-quarter-2017-earnings accessed 21 November 2021.

⁶ Jack Nicas, 'YouTube Tops 1 Billion Hours of Video a Day, on Pace to Eclipse TV' *Wall Street Journal* (27 February 2017) www.wsj.com/articles/youtube-tops-1-billion-hours-of-video-a-day-on-pace-to-eclipse-tv-1488220851 accessed 21 November 2021.

⁷ Jack M. Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation' (2018) 51 *University of California Davis* 1151.

⁸ Andras Koltay, *New Media and Freedom of Expression. Rethinking the Constitutional Foundations of the Public Sphere* (Hart 2019).

⁹ Josè Van Dijk and Thomas Poell, 'Understanding Social Media Logic' (2013) 1(1) *Media and Communication* 2; Tarleton Gillespie, 'The Politics of Platforms' (2010) 12(3) *News Media & Society* 347.

¹⁰ Manuel Castells, *Networks of Outrage and Hope: Social Movements in the Internet Age* (Polity Press 2012).

¹¹ Andrew Tutt, 'The New Speech' (2014) 41 *Hastings Constitutional Law Quarterly* 235.

¹² Evgeny Morozov, *The Net Delusion: The Dark Side of Internet Freedom* (Public Affairs 2011).

interests rather than public values. Indeed, the primary point is to understand which remedies can mitigate the risk of exposing users just to content reflecting business logics rather than pluralism. The informational (and power) asymmetry between users and platforms leads to questioning whether the traditional liberal and negative dimension of the right to freedom of expression can ensure democratic values in the algorithmic era.

Within this clash between democratic public values and non-democratic business interests, this chapter focuses on the challenges of freedom of expression in the algorithmic society and on how European digital constitutionalism can provide remedies to deal with this troubling scenario for democracy and the rule of law. This challenge is particularly relevant for democratic societies. As underlined in Chapter 3, democratic states are open environments for pluralism and values such as liberty, equality, transparency and accountability. On the contrary, the activity of online platforms is based on business interests, opaque procedures and unaccountable decision-making. Democracy relies on individual self-determination and autonomy which are the primary drivers for developing opinions and participation in decision-making processes. The lack of pluralism as driven by online platforms could undermine the ability of users to make decisions based on a multiplicity of voices concurring to develop ideas. Therefore, freedom of expression is not only a individual fundamental right subject to the interference of powers but also a constitutional instrument to foster autonomy in a democratic society, reflecting the framework of dignity characterising European constitutionalism.

As examined in Chapter 3, the law of the platform competes with the authority exercised by public actors. While online platforms have a responsibility rather than a duty to guarantee the respect of fundamental rights and freedoms, democratic states are required to safeguard these interests to protect the entire democratic system. Such duty also encompasses a positive obligation to protect individuals against acts committed by private persons or entities.¹³ Without protecting equality, freedom of expression or assembly, it would not be possible to enjoy a democratic society.

¹³ UN Human Rights Committee (HRC), 'General comment no. 31 [80], The nature of the general legal obligation imposed on States Parties to the Covenant', 26 May 2004 www.refworld.org/docid/478b26ae2.html accessed 21 November 2021.

This chapter underlines that the vertical and negative nature of freedom of expression is no longer enough to protect democratic values in the digital environment, since the flow of information is actively organised by business interests, driven by profit-maximisation rather than democracy, transparency or accountability. This chapter demonstrates how the development of the algorithmic society has challenged the liberal paradigm of free speech requiring a complementary shift from a negative and active to a positive and passive dimension. Therefore, this chapter examines how European digital constitutionalism leads to reframing media pluralism to protect freedom of expression in the algorithmic society.

The first part of this chapter analyses the shift from a liberal economic narrative based on the metaphor of the free marketplace of ideas to the rise of platform power to moderate online content. Precisely, it focuses on the logic of content moderation, the rise of the algorithmic public sphere and the challenges to the protection of the right to freedom of expression raised by the private enforcement of fundamental rights. The second part focuses on the current status quo, underlining the first step of European digital constitutionalism towards limiting platform power and focusing on the horizontal effect doctrine as a potential way to fill the regulatory gap in the field of content moderation. The third part examines the approach of European digital constitutionalism to address the challenges of content moderation, focusing on rethinking online media pluralism through transparency and procedural safeguards.

5.2 From the Free Marketplace of Ideas . . .

The right to freedom of expression in modern and contemporary history has liberal roots. Like other civil and political liberties risen at the end of the nineteenth century,¹⁴ the right to free speech is based on the idea that liberties and freedoms can be ensured by limiting interferences coming from public actors.¹⁵ The possibility to express opinions and ideas freely is the grounding condition to develop personal identity and ensures the right to self-determination in a democratic society.

It is not by chance that one of the most suggestive legal metaphors in this field is that of the ‘free marketplace of ideas’,¹⁶ as coined for the

¹⁴ The Declaration of the Rights of Man and of the Citizen (1789).

¹⁵ Eric Barendt, *Freedom of Speech* (Oxford University Press 2017).

¹⁶ See, e.g., Daniel E. Ho and Frederik Schauer, ‘Testing the Marketplace of Ideas’ (2015) 90 *New York University Law Review* 1161; Eugene Volokh, ‘In Defense of the Market Place

first time by Justice Douglas in *United States v. Rumely*.¹⁷ This liberalist belief can be contextualised in the classical theory of market balance applied to the field of ideas.¹⁸ Since individuals act rationally, they can choose the best products and services in a free market. As in a competitive market where the best products or services prevail, the same mechanism would apply to the best information resulting from market balance.

However, the liberal grounds of freedom of expression are more in depth and older. In the seventeenth century, Milton, opposing the English Parliament's Press Ordinance, which had introduced a system of censorship to punish the promoters of ideas considered illegal, argued that freedom of expression should not be limited to allow the truth to prevail thanks to the free exchange of opinion.¹⁹ Milton compares the truth to a streaming fountain whose water constitutes the flow of information saving men from prejudice. According to this perspective, it is necessary to avoid any interference with the flow of information to lead men to the highest level of knowledge. Two centuries later, Mill shared a liberal approach to freedom of expression.²⁰ Even falsehood could contribute to reaching the truth.²¹ Otherwise, censoring falsehood would make meaningless

of Ideas / Search for Truth as a Theory of Free Speech Protection' (2011) 97(3) Virginia Law Review 591; Joseph Blocher, 'Institutions in the Marketplace of Ideas' (2008) 57(4) Duke Law Journal 820; Paul H. Brietzke, 'How and Why the Marketplace of Ideas Fails' (1997) 31(3) Valparaiso University Law Review 951; Alvin I. Goldman and James C. Cox, *Speech, Truth, and the Free Market for Ideas* (Cambridge University Press 1996).

¹⁷ *United States v. Rumely* 345 U.S. 41 (1953). 'Of necessity I come then to the constitutional questions. Respondent represents a segment of the American press. Some may like what his group publishes; others may disapprove. These tracts may be the essence of wisdom to some; to others their point of view and philosophy may be anathema. To some ears their words may be harsh and repulsive; to others they may carry the hope of the future. We have here a publisher who through books and pamphlets seeks to reach the minds and hearts of the American people. He is different in some respects from other publishers. But the differences are minor. Like the publishers of newspapers, magazines, or books, this publisher bids for the minds of men in the market place of ideas'.

¹⁸ Ronald Coase, 'Markets for Goods and Market for Ideas' (1974) 64(2) American Economic Review 1974.

¹⁹ John Milton, *Aeropagitica* (1644). According to Milton: 'So Truth be in the field, we do injuriously, by licensing and prohibiting, to misdoubt her strength. Let her and Falsehood grapple; who ever knew Truth put to the worse, in a free and open encounter?'

²⁰ John S. Mill, *On Liberty* (1859).

²¹ *Ibid.* '[I]f any opinion is compelled to silence, that opinion may, for aught we can certainly know, be true. To deny this is to assume our own infallibility'.

the comparison between ideas and opinions with the risk of dogmatising the current truth.²² Both Milton and Mill agreed that the right to freedom of expression is effective when it is free from censorship and from the interferences of power.

The scope of these liberal ideas opposing public actors' interferences also emerged in the US legal framework. Justice Holmes' dissenting opinion in *Abrams v. United States* can still be considered the constitutional essence of freedom of expression in the United States as enshrined in the First Amendment.²³ The case concerned the distribution of leaflets calling for ammunition factories to strike to express a clear message of resistance against the US military intervention in Russia. According to Justice Holmes, although men try to support their positions by criticising opposing ideas, they must not be persuaded that their opinions are certain. Only the free exchange of ideas can confirm the accuracy of each position.²⁴ Freedom of speech is functional to ensure that individuals are autonomous and, therefore, responsible moral agents participating in a political society.²⁵ According to Meiklejohn, the constitutional protection of free speech aims to foster citizens' awareness about public matters.²⁶

This liberal approach has also been expressed, more recently, in the framework of the digital environment, at least in two landmark decisions of the US Supreme Court. In 1997, in *Reno v. ACLU*,²⁷ the Supreme Court ruled that the provisions of the Communication Decency Act concerning the criminalisation of obscene or indecent materials to

²² *Ibid.* '[E]ven if the received opinion be not only true, but the whole truth; unless it is suffered to be, and actually is, vigorously and earnestly contested, it will, by most of those who receive it, be held in the manner of a prejudice, with little comprehension or feeling of its rational grounds'.

²³ *Abrams v. United States* (1919) 250 U.S. 616: 'Persecution for the expression of opinions seems to me perfectly logical. If you have no doubt of your premises or your power and want a certain result with all your heart you naturally express your wishes in law and sweep away all opposition . . . But when men have realized that time has upset many fighting faiths, they may come to believe even more than they believe the very foundations of their own conduct that the ultimate good desired is better reached by free trade in ideas. . . . The best test of truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which their wishes safely can be carried out'.

²⁴ Sheldon Novick, *Honorable Justice* (Laurel 1990).

²⁵ Ronald Dworkin, *Freedom's Law: The Moral Reading of the American Constitution* (Oxford University Press 1999).

²⁶ Alexander Meiklejohn, *Free Speech and Its Relation to Self-Government* (Lawbook Exchange 2011).

²⁷ *Reno v. American Civil Liberties Union* 521 U.S. 844 (1997).

any person under eighteen was unconstitutional.²⁸ As observed by the Supreme Court, unlike traditional media outlets, ‘the risk of encountering indecent material by accident is remote because a series of affirmative steps is required to access specific material’.³¹ According to Justice Stevens, the Internet plays the role of a ‘new marketplace of ideas’ observing that ‘the interest in encouraging freedom of expression in a democratic society outweighs any theoretical but unproven benefit of censorship’.²⁹ Besides, he observed that the growth of the Internet as been phenomenal and, therefore, ‘we presume that governmental regulation of the content of speech is more likely to interfere with the free exchange of ideas than to encourage it. The interest in encouraging freedom of expression in a democratic society outweighs any theoretical but unproven benefit of censorship’.³⁰ This decision can be considered the first step towards a transformation of the public forum doctrine.³¹

Despite the passing of years and opposing positions, this liberal approach has been reiterated more recently in *Packingham v. North Carolina*.³² The case involved a statute banning registered sex offenders from accessing social networking services to avoid any contact with minors. The US Supreme Court placed the Internet and social media on the same layer of public places where the First Amendment enjoys a broad scope of protection. In the words of Justice Kennedy: ‘It is cyberspace – the “vast democratic forums of the Internet” in general, and social media in particular’.³³ The metaphor of the (digital) free marketplace of ideas is still firm in the jurisprudence of the US Supreme Court. Social media are indeed considered as an enabler of democracy rather than as a threat for public discourse. This would also contribute to explaining why social media enjoy a safe constitutional area of protection under the First Amendment, which, in the last twenty years, has constituted a fundamental ban on any attempt to regulate speech online,³⁴ thus showing the role of the First Amendment in US

²⁸ Communication Decency Act (1996).

²⁹ 521 U.S. 844 (n. 27).

³⁰ *Ibid.*, 885.

³¹ Dawn C. Nunziato, ‘The Death of The Public Forum in Cyberspace’ (2005) 20 *Berkeley Technology Law Journal* 1115.

³² *Packingham v. North Carolina* (2017) 582 U.S. ____.

³³ *Ibid.*

³⁴ See, e.g., *Ashcroft v. Free Speech Coalition* (2002) 535 U.S. 234; *Ashcroft v. American Civil Liberties Union* (2002) 535 U.S. 564.

constitutionalism,³⁵ as ‘the paramount right within the American constellation of constitutional rights’.³⁶

Nevertheless, it would be enough just to cross the Atlantic to understand how this general trust for a vertical paradigm of free speech is not shared worldwide by other democracies, especially when the right to freedom of expression is framed in the digital environment. While, in the United States, the Internet and social media still benefit from the frame coming from the traditional liberal metaphor of the free marketplace of ideas as a safeguard for democracy,³⁷ in Europe, freedom of expression online does not enjoy the same degree of protection.³⁸ In the European framework, the right to freedom of expression is subject to a multilevel balancing,³⁹ precisely with other rights enshrined in the Charter,⁴⁰ the Convention⁴¹ and national constitutions. Unlike the US Supreme Court, the Strasbourg Court has shown a more restrictive approach to the protection of the right to freedom of expression in the digital environment, perceived more like a risk rather than as an opportunity for the flourishing of democratic values.⁴²

Such a cautious approach in Europe does not only aim to balance different constitutional interests but also to avoid that granting absolute protection to one right could lead to the destruction of other

³⁵ See, e.g., Lee C. Bollinger and Geoffrey R. Stone (eds.), *The Free Speech Century* (Oxford University Press 2019); Floyd Abrams, *The Soul of the First Amendment* (Yale University Press 2017); Frederik Schauer, ‘The Exceptional First Amendment’ in Michael Ignatieff (ed.), *American Exceptionalism and Human Rights* 29 (Princeton University Press 2005); Alexander Meiklejohn, ‘The First Amendment Is an Absolute’ (1961) *The Supreme Court Review* 245.

³⁶ Michel Rosenfeld and Andras Sajó, ‘Spreading Liberal Constitutionalism: An Inquiry into the Fate of Free Speech Rights in New Democracies’ in Sujit Choudhry (ed.), *The Migration of Constitutional Ideas* 152 (Cambridge University Press 2007).

³⁷ Claudia E. Haupt, ‘Regulating Speech Online: Free Speech Values in Constitutional Frames’ Northeastern University School of Law Research Paper No. 402-2021 (22 July 2021) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3794884 accessed 19 November 2021.

³⁸ Oreste Pollicino and Marco Bassini, ‘Free Speech, Defamation and the Limits to Freedom of Expression in the EU: A Comparative Analysis’ in Andrej Savin and Jan Trzaskowski (eds.), *Research Handbook on EU Internet Law* 508 (Edward Elgar 2014); Vincenzo Zeno-Zencovich, *Freedom of Expression: A Critical and Comparative Analysis* (Routledge 2008).

³⁹ Ingolf Pernice, ‘The Treaty of Lisbon: Multilevel Constitutionalism in Action’ (2009) 15 (3) *Columbia Journal of European Law* 349.

⁴⁰ Charter of Fundamental Rights of the European Union (2012) OJ C326/12, Arts. 11, 52.

⁴¹ European Convention on Human Rights (1950), Art. 10.

⁴² Oreste Pollicino, ‘Judicial Protection of Fundamental Rights in the Transition from the World of Atoms to the World of Bits: The Case of Freedom of Speech’ (2019) 25(2) *European Law Journal* 155.

fundamental interests undermining de facto their constitutional relevance.⁴³ This is an expression of the different understanding of the role of dignity on the western side of the Atlantic as mentioned in Chapter 1. In Europe, freedom of expression is not just a liberal value whose protection needs to be safeguarded at any cost to protect democracy. Allowing such an approach would also entail that speech could be used as a constitutional excuse to hinder democracy itself. From a European constitutional perspective, freedom of expression is instead a fundamental right whose protection needs to take into account the other constitutional interests at stake. Unlike the frame of liberty in the US constitutional framework, freedom of expression in Europe does not enjoy absolute protection but is subject to the logic of balancing intimately connected to human dignity.⁴⁴ Bognetti underlined the European reluctance to read freedom of speech in ways that would sacrifice other constitutional values. He observed: ‘At times the necessity of preserving the values of liberal democracy has been felt so intensely as to lead to the prohibition of political parties and to deny legitimacy to speech that has been seen to undermine these values’.⁴⁵

This non-exhaustive framework provides clues to understand why the Union has not adopted an ommissive approach to the challenges to freedom of expression raised by the algorithmic society, thus paving the way towards a new approach, precisely focusing on regulating the process of content moderation. Despite the difference in the protection of the right to freedom of expression in the EU and the United States, this fundamental right is still the prerequisite for a democratic society. However, in the digital environment, the protection of this fundamental right is no longer a matter of quantity but a matter of quality because of the crucial role of online platforms in determining the standard of protection of freedom of expression and other fundamental rights on a global scale. The case of disinformation is a paradigmatic example of the challenges to the right to freedom of expression in the algorithmic society.⁴⁶ In other words, the primary challenge for democracies is no longer that of

⁴³ Charter (n. 40), Art. 54; Convention (n. 41), Art. 17.

⁴⁴ Mattias Kumm and Alec D. Walen, ‘Human Dignity and Proportionality: Deontic Pluralism in Balancing’ Grant Huscroft and others (eds.), *Proportionality and the Rule of Law: Rights, Justification, Reasoning* (Cambridge University Press 2014).

⁴⁵ Giovanni Bognetti, ‘The Concept of Human Dignity in U.S. and European Constitutionalism’ in Georg Nolte (ed.), *European and US Constitutionalism* 77 (Cambridge University Press 2005).

⁴⁶ Giovanni Pitruzella and Oreste Pollicino, *Disinformation and Hate Speech: An European Constitutional Perspective* (Bocconi University Press 2020).

protecting freedom of expression extensively by granting access to new digital channels and avoiding interferences from public actors, but, rather, that of ensuring exposure and pluralism in a democratic digital environment.

5.3 ... To the Algorithmic Marketplace of Ideas

At the World Summit on the Information Society in 2004, Lessig underlined the significant potentialities afforded by the digital environment: '[f]or the first time in a millennium, we have a technology to equalize the opportunity that people have to access and participate in the construction of knowledge and culture, regardless of their geographic placing'.⁴⁷ Likewise, Shapiro stated: 'Hierarchies are coming undone. Gatekeepers are being bypassed. Power is devolving down to "end users" ... No one is in control except you'.⁴⁸ These were positive news for the free marketplace of ideas. Information sources have spread online. The new online communication channels have enabled users to potentially reach a global audience without relying any longer on the traditional channels of communications in the hand of publishers like newspapers and televisions.⁴⁹ Put another way, the Internet as a new channel of communication promised to overcome the problem of concentration of power in traditional media warned of by Habermas.⁵⁰

Although it is true that the possibility for users to express opinions and ideas without traditional filters cannot be contested, nonetheless, the lack of control over information online has been revealed to be just a libertarian dream. It is true that users can still run their blogs and websites to share their ideas or opinions, sell products or keep social relationships. However, it would be naïve to believe that this is how most information flows online. As underlined in Chapter 3, to exercise online rights and freedoms, it is almost necessary to rely on online platforms, primarily social media. These entities aim to maximise

⁴⁷ Lawrence Lessig, 'An Information Society: Free or Feudal' (2004) World Summit on the Information Society (WSIS) www.itu.int/wsis/docs/pc2/visionaries/lessig.pdf accessed 22 November 2021.

⁴⁸ Andrew L. Shapiro, *The Control Revolution: How the Internet is Putting Individuals in Charge and Changing the World we Know* 11, 30 (Public Affairs 1999).

⁴⁹ Jack M. Balkin, 'Old-School/New-School Speech Regulation' (2014) 127 *Harvard Law Review* 2296.

⁵⁰ Marianne Franklin, *Digital Dilemmas: Power, Resistance, and the Internet* (Oxford University Press 2013).

their profit, and expressions – to say nothing of data – are the perfect means to achieve this purpose. By processing content, platforms can extract information, collect data and even map emotions to provide the most granular advertising services on the market and finding new ways to attract customers.⁵¹ It would be enough to observe the business models of Facebook and Google based on more than 80 per cent on advertising revenues coming from advertising services.⁵² Just these two platforms absorb 75 per cent of the \$73 billion digital advertising market in the United States.⁵³ In other words, users are subject to the private governance of the space where information flows based on business logic of online platforms.

The moderation of expressions for profit reflects the logic of digital capitalism, or better information capitalism, which leads platforms to express surveillance and governance as expressions of powers.⁵⁴ At first glance, there would be not so many differences with traditional media outlets governing and filtering information. Nonetheless, in the digital environment, the source of platform power comes primarily from algorithmic technologies processing a vast amount of data and information that platforms can accumulate, revealing users' intimate information which is enormously valuable for commercial interests, governments' public tasks and political campaigns. If these considerations are mixed with the immunity of online intermediaries from liability for hosting third-party content, it should not come as a surprise how profitable it is for platforms to run their business with a very low degree of risk. In other words, by relying on their immunity, platforms have developed business models profiting from online speech without accountability.

However, although the private governance of content frames online speech in a mercantilist environment where the space for democratic values is only a matter of business incentives, the role of algorithms in organising content as well has positive effects to help users interact and

⁵¹ Vindu Goel, 'Facebook Tinkers with Users' Emotions in News Feed Experiment, Stirring Outcry' *The New York Times* (29 June 2014) www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html accessed 21 November 2021.

⁵² Mathew Ingram, 'How Google and Facebook Have Taken Over the Digital Ad Industry' *Fortune* (4 January 2017) <https://fortune.com/2017/01/04/google-facebook-ad-industry/> accessed 21 November 2021.

⁵³ Shannon Bond, 'Google and Facebook Build Digital Duopoly' *Financial Times* (14 March 2017) ft.com/content/30c81d12-08c8-11e7-97d1-5e720a26771b accessed 21 November 2021.

⁵⁴ Julie Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism* (Oxford University Press 2019).

access the vast amount of information in a framework of scarcity of time and attention.⁵⁵ Information has spread online with the result that what is now scarce is not the source of information but the attention of the listeners.⁵⁶ This change has led to the emergence of the 'attention economy' pushing towards new strategies to attract consumers.⁵⁷ If social media programme their algorithms to achieve business purposes through content moderation, it should not come as a surprise that content moderation does not reflect necessarily democratic values like diversity or truthfulness. The primary goal is just increasing the probability of an interaction between users and the time and quantity of content they share on social media spaces. Even more importantly, as examined in Chapter 3, such discretion in the moderation of expressions contributes to shaping online speech and the principle of the rule of law. The price to pay for such an intermediation consists of accepting the private values translated by algorithmic determinations.

These considerations show why considering public actors as the only source of interference for freedom of expression online could today seem anachronistic. A further challenge raised by the algorithmic society concerns how to address the discretion of private actors freely influencing the limits of freedom of expression on a global scale without any public guarantee. The metaphor of the marketplace of ideas is critical now more than ever to represent the current situation, but with a small makeup. The difference consists of the change of the expression 'free' with 'algorithmic', that moves the perspective from democratic and collective values to business and individualist purposes. Ideas do not reach a market balance through the invisible hand, but are driven by oligopolist logics where decisions are centralised. In the algorithmic marketplace of ideas, speech is still central but not quite as much from the perspective of users' freedoms as from that of platforms' profits. Within this framework, the following subsections focus on the characteristics of the algorithmic public sphere, the logic of moderation and the private enforcement of freedom of expression online.

⁵⁵ Natali Helberger, 'On the Democratic Role of News Recommenders' (2019) 7(8) *Digital Journalism* 993.

⁵⁶ Herbert A. Simon, 'Designing Organizations for an Information-Rich World' in Martin Greenberger (ed.), *Computers, Communications, and the Public Interest* 37 (Johns Hopkins Press 1971).

⁵⁷ Tim Wu, *The Attention Merchants: The Epic Scramble to Get Inside our Heads* (Knopf 2016).

5.3.1 The Public Sphere in the Age of Algorithms

'Imagine a future in which your interface agent can read every news-wire and newspaper and catch every TV and radio broadcast on the planet, and then construct a personalised summary. This kind of newspaper is printed in an edition of one'. These were the words of Negroponte in 1995 in the aftermath of the Internet.⁵⁸ The situation of centralisation and personalisation of expression which users are experiencing was already in these sentences.

In the algorithmic society, online platforms mediate the ability of users to share their opinions and ideas online. The use of Google or Facebook is almost a mandatory step for entering the public debate and building social interactions online.⁵⁹ Already in 1962, Habermas observed that 'the process in which societal power is transformed into political power is as much in need of criticism and control as the legitimate exercise of political domination over society'.⁶⁰ The lack of control in the shift from social to political is what already happened in the field of traditional media outlets. Once again, Habermas already underlined the debasement of the public sphere consisting of the high societal barriers to access channels of communication (e.g. print media) and the intertwined relationship with politics.⁶¹ In this bottleneck, a bunch of national mass media institutions governed public discourse.

These considerations would not sound brand new in the digital environment. Like any other libertarian dream, the idea of an alternative space overcoming traditional forms of control failed. Together with states, other entities contribute to producing norms regulating spaces. As Fraser explained, it is not possible to think a public sphere free from manipulation in a capitalist economy where different forces tend to influence the formation of the public opinion and societal beliefs.⁶² Benkler already underlined how the digital environment projects users in a 'networked public sphere'.⁶³ The difference is the mediating subject which has changed from a bunch of traditional media outlets to

⁵⁸ Nicholas Negroponte, *Being Digital* 153 (Alfred A. Knopf 1995).

⁵⁹ Taina Bucher, 'Want to Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook' (2012) 14(7) *New Media & Society* 1164.

⁶⁰ Jürgen Habermas, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* 210 (MIT Press 1991).

⁶¹ Jürgen Habermas, *Between Facts and Norms* (MIT Press 1998).

⁶² Nancy Fraser, 'Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy' (1990) 25/26 *Social Text* 56.

⁶³ Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (Yale University Press 2006).

an oligopoly of online providers. While, at first glance, the digital environment could be a solution to overcome centralised powers in the media sector, realising Habermas' dream of a bourgeois public sphere, a closer look shows how similar dynamics of centralisation and control over information have been reproduced in the digital environment creating a quasi-public sphere.⁶⁴ Platforms' ability to massively organise or amplify certain voices (and decide how to do that) leads to thinking about the future of the public sphere online.

This framework of power does not mean that the digital environment has not provided opportunities to express ideas and opinions. Although the rise of information pluralism should generally be welcomed for the development and maintenance of a democratic environment, the characteristics of the information flow online and its moderation raise serious concerns in terms of 'quantity' and 'quality' of the information sources.

From a quantitative perspective, in the last twenty years, a high degree of concentration of the online platforms' market has characterised the digital environment. As foreseen by Zittrain,⁶⁵ the characteristics of the information society have led to the creation of monopolies,⁶⁶ linked to the platformisation of the Internet,⁶⁷ which Srnicek would call the era of 'platform capitalism'.⁶⁸ This market concentration empowers a limited number of platforms to set the conditions on which vast amounts of content and data flow online. The effect of this process is to create barriers for entering the market of information and increase the dependency of traditional media outlets from the new opportunities of visibility offered by social media. Although, at first glance, the digital environment has empowered users to access new channels to share ideas and access sources of information, however, the aforementioned digital convergence dangerously affects media pluralism from a quantitative perspective.

⁶⁴ Jillian C. York, 'Policing Content in the Quasi-Public Sphere' Open Net Initiative' Bulletin (September 2010) <https://opennet.net/policing-content-quasi-public-sphere> accessed 21 November 2021.

⁶⁵ Jonathan Zittrain, *The Future of the Internet and How to Stop It* (Yale University Press 2008).

⁶⁶ Robin Mansell and Michele Javary, 'Emerging Internet Oligopolies: A Political Economy Analysis' in Arthur S. Miller and Warren J. Samuels (eds.), *An Institutionalist Approach to Public Utilities Regulation* (Michigan State University Press 2002).

⁶⁷ Anne Helmond, 'The Platformization of the Web: Making Web Data Platform Ready' (2015) 1(2) *Social Media + Society* 1.

⁶⁸ Nick Srnicek, *Platform Capitalism* (Polity Press 2016).

From a qualitative standpoint, pluralism is based on different manifestations of thinking and promotes heterogeneous ideas. In the digital environment, the use of artificial intelligence for online content moderation mitigates this positive effect. The European High-Level Expert Group on Media diversity underlined this point explaining the negative impact on democracy by noting that, while ‘increasing filtering mechanisms make it more likely for people to only get news on subjects they are interested in, and with the perspective, they identify with’, ‘[this reality] will also tend to create more insulated communities as isolated subsets within the overall public sphere’.⁶⁹ Democracy indeed needs a public sphere where the meeting of ideas and opinions can be a ‘societal glue’.⁷⁰ Otherwise, individuals are likely to be attracted by extreme and dogmatic poles, forgetting the alternative ideas which are the basis for consensus in a democratic society. The Habermasian idea of the public sphere is hard to realise in the digital environment where ideas are formulated, negotiated and distributed by machines. In other words, the public sphere in the age of algorithms is not under the control and guidance of public opinion but instead is governed by opaque business purposes.

In a footnote within a larger article of 2006, Habermas underlined the critical role of digital technologies for democracy, looking particularly at authoritarian regimes. However, ‘[i]n the context of liberal regimes, the rise of millions of fragmented chat rooms across the world tend instead to lead to the fragmentation of large but politically focused mass audiences into a huge number of isolated issue publics’.⁷¹ Despite the criticisms and disappointment sparked by this non-exhaustive comment,⁷² these sentences underline the double face of the online public sphere: a great opportunity for democracy as a liberation

⁶⁹ High-Level Group on Media Freedom and Pluralism, ‘A free and pluralistic media to sustain European democracy’ (2013), 27 <https://ec.europa.eu/digital-single-market/sites/digital-agenda/files/HLG%20Final%20Report.pdf> accessed 22 November 2021.

⁷⁰ Cass R. Sunstein, *Republic.com* 9 (Princeton University Press 2002).

⁷¹ Jurgen Habermas, ‘Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research’ (2006) 16(4) *Communication Theory* 411, 423.

⁷² Howard Rheingold, ‘Habermas Blows Off Question about the Internet and the Public Sphere’, *SmartMobs* (5 November 2007) www.smartmobs.com/2007/11/05/habermas-blows-off-question-about-the-internet-and-the-public-sphere/ accessed 19 November 2021; Stuart Geiger, ‘Does Habermas Understand the Internet? The Algorithmic Construction of the Blog/Public Sphere’ (2009) 10(1) *Gnovis: A Journal of Communication, Culture, and Technology* www.gnovisjournal.org/2009/12/22/does-habermas-understand-internet-algorithmic-construction-blog-public-sphere/ accessed 19 November 2021.

technology, but also as a risk for the fragmentation of the public sphere driven by business purposes. According to Habermas, a solid democracy is highly dependent on the public opinion. The shift from ‘public’ to ‘artificial’ opinion due to the lack of ability of individuals to act as rational agents is one of the reasons why democracy could be threatened in the algorithmic society.

Such a liberal root of the public sphere, naturally and deeply connected with that of freedom of expression, is not just put under pressure, but it is basically frustrated. It is worth wondering how individuals can be rational users in the algorithmic public sphere if they are subject to a top-down power exercised by online platforms driving the public sphere through artificial intelligence systems whose decision-making processes cannot be always explained. In other words, the same failure of freedom of expression as a negative right to protect democratic values also extends to the liberal vision of the digital public sphere.

A digital liberal approach to the public sphere based on the autonomy and rationality of users seems not to be enough to ensure democratic values any longer. The shift from the ‘free’ to the ‘algorithmic’ marketplace of ideas has shown the fallacies of the traditional instruments of pluralism when implemented in the digital environment. Accessing more information does not necessarily imply accessing better information. The organisation of content aims to engage users based on their data and preferences, leading to the polarisation of the debate due to the creation of ‘filter bubbles’ or ‘information cocoons’,⁷³ which Sunstein defines as ‘communication universes in which we hear only what we choose and only what comforts us and pleases us’.⁷⁴ The personalisation of online content leads to the creation of echo chambers,⁷⁵ where each user is isolated and marginalised from opposing positions as resulting

⁷³ Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (Viking 2011); Cass R. Sunstein, *Republic.com 2.0* (Princeton University Press 2007).

⁷⁴ Cass R. Sunstein, *Infotopia: How Many Minds Produce Knowledge* 9 (Oxford University Press 2006).

⁷⁵ Empirical evidence of filter bubbles is scarce. See, e.g., see Judith Moeller and Natali Helberger, ‘Beyond the Filter Bubble: Concepts, Myths, Evidence and Issues for Future Debates. A Report Drafted for the Dutch Media Regulator’ (2018) <https://dare.uva.nl/search?identifier=478edb9e-8296-4a84-9631-c7360d593610> accessed 19 November 2021; Richard Fletcher and Rasmus K. Nielsen, ‘Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication’ (2017) 67(4) *Journal of Communication* 476; Ivan Dylko and others, ‘The Dark Side of Technology: An Experimental Investigation of the Influence of Customizability Technology on Online Political Selective Exposure’ (2017) 73 *Computers in Human Behavior* 181.

from a mere algorithmic calculation. There are already studies showing the role of algorithmic bias in reflecting and amplifying existing human beliefs.⁷⁶ In other words, users are encouraged to interact only with information inside the area of their preferences.

This effect primarily results from the logic of moderation. Personalisation, more than removal or organisation, allows indeed platforms to maximise online attention, thus meeting the interests of companies interested in advertising their products and services online. Social media exploit the characteristics of human communication based on the tendency to avoid dissensus.⁷⁷ Since advertising revenues are highly dependent on attracting scarce attention, discovering new ways to manipulate users' behaviours is the mission of online platforms. Automation is implemented not only to remove but also organise and recommend content, thus influencing users' interactions. It would be enough to think about how the search results of Google or the Facebook newsfeed are not the same for each individual,⁷⁸ but they create what, at the beginning of this century, has already been defined as distinguished public spheres.⁷⁹

The fragmentation of the public sphere is also driven by micro-targeting strategies which aim to limit the audience to certain content to increase the likelihood of capturing attention. While, like price discrimination, this is not an issue in the market field, it is instead when this practice is applied to the democratic debate that it shows how believing in a uniform public sphere in the information society could not be possible. Micro-targeting strategies intentionally focus just on certain groups giving the possibility to reach only those who are potentially interested in that content, no matter if the information is of commercial or political nature.⁸⁰ In this case, platforms become the arbiter of content online, including political speech.⁸¹

⁷⁶ Safiya U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York University Press 2018).

⁷⁷ Leon Festinger, *A Theory of Cognitive Dissonance* (Stanford University Press 1957).

⁷⁸ Micheal A. DeVito, 'From Editors to Algorithms' (2017) 5(6) *Digital Journalism* 753.

⁷⁹ There is not a unitary notion of public sphere. See, e.g., Todd Gitlan, 'Public Sphere or Public Sphericules?' in Tamar Liebes and James Curran (eds.), *Media, Ritual and Identity* 168 (Routledge 2002); Micheal Warner, *Publics and Counterpublics* (MIT University Press 2002); Catherine R. Squires, 'Rethinking the Black Public Sphere: An Alternative Vocabulary for Multiple Public Spheres' (2002) 12(4) *Communication Theory* 446.

⁸⁰ Frederik J. Zuiderveen Borgesius and others, 'Online Political Microtargeting: Promises and Threats for Democracy' (2018) 14(1) *Utrecht Law Review* 82.

⁸¹ Daniel Kreiss and Shannon C. McGregor, 'The "Arbiters of What Our Voters See": Facebook and Google's Struggle with Policy, Process, and Enforcement around Political

Although traditional media outlets could be accused of filtering relevant news or even manipulating information, they just provide unique platforms to discuss. On the opposite, online platforms create different places driven by business purposes for each user. Algorithms can indeed decide what deserves to be on top and what instead is best to hide. They choose who is a best friend rather than recommending that journal article or blog post to read. By processing a vast amount of information and data, artificial intelligence systems can select the relevant item to put in front of the user's eyes. The problem is that information that is relevant for the public debate is not defined by the exchange of views and opinions but machines. These systems are far from being perfect, leading to potential discriminatory bias or to exposure to objectionable content.⁸²

Within this framework, content moderation contribute to generating intertwined public spheres whose sum then makes the single (and invisible) public sphere. This is also why, according to Schudson, the public sphere was never entirely based on agents' rational independency.⁸³ It has been always shaped by a form of intimate tribality governing the transmission of knowledge and ideas across society. What makes the public sphere is the sense of community or namely the function of communication towards building a global village,⁸⁴ where people consume information to underline their connection and define their place in the world.

Within this framework, users cannot access transparent information about what happens behind the screen. Between self-selected and pre-selected personalisation, also known as explicit or implicit personalisation,⁸⁵ the latter mostly prevail over the former.⁸⁶ In the

Advertising' (2019) 36(4) Political Communication 499; Shannon C. McGregor, 'Personalization, Social Media, and Voting: Effects of Candidate Self-Personalization on Vote Intention' (2017) 20(3) News Media & Society 1139.

⁸² Muhammad Ali and others, 'Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes' in *Proceedings of the ACM on Human-Computer Interaction* (ACM 2019); Reuben Binns and others, 'Like Trainer, Like Bot? Inheritance of Bias in Algorithmic Content Moderation' in Giovanni L. Ciampaglia, Afra Mashhadi and Taha Yasseri (eds.), *Social Informatics* 405 (Springer 2017).

⁸³ Micheal Schudson, 'Was There Ever a Public Sphere? If So, When? Reflections on the American Case' in John Calhoun (ed.), *Habermas and the Public Sphere* 143 (MIT Press 1992).

⁸⁴ Marshall McLuhan, *Understanding Media. The Extensions of Man* (MIT Press 1994).

⁸⁵ Neil Thurman and Steve Schifferes, 'The Future of Personalization at News Websites: Lessons from a Longitudinal Study' (2012) 13(5-6) *Journalism Studies* 775.

⁸⁶ Frederik J. Zuiderveen Borgesius and others, 'Should We Worry about Filter Bubbles?' (2016) 5(1) *Internet Policy Review* <https://policyreview.info/node/401/pdf> accessed 21 November 2021.

first case, users have more discretion in defining the criteria according to which online platforms organise their content through automated systems (i.e. selective exposure).⁸⁷ These options can include filters for certain types of content or topics rather than specific users or groups. This case is also relevant in the atomic world where individuals chose which kind of media outlets they wanted to rely on when buying a newspaper or watching television. This type of personalisation can also be beneficial for users since it leaves in the hands of individuals the possibility to choose their degree of exposure.⁸⁸ On the contrary, pre-selected personalisation is driven not only by online platforms but also exogenous factors as the goal to reach a new advertising strategy required by the market. Therefore, algorithmic accountability and transparency play a critical role in increasing users' autonomy and reduce the fragmentation of the public sphere.⁸⁹

The challenges of content moderation could lead to the debasement of information pluralism in the digital environment. Instead of a democratic and decentralised society as defined at the end of the last century, an oligopoly of private entities has emerged, controlling information and determining how people exchange it.⁹⁰ Arendt described the public domain as a place 'where men exist not merely like other living or inanimate things, but to make their appearance explicitly' (i.e. the 'space of appearance').⁹¹ Nonetheless, this space is not stable but highly dependent on the performance of deeds or the utterance of words. Indeed, 'unlike the spaces which are the work of our hands, it does not survive the actuality of the movement which brought it into being, but disappears not only with the dispersal of men – as in the case of great catastrophes when the body politic of a people is destroyed – but with the disappearance or arrest of the activities themselves'.⁹²

The primary question is whether platform determinations shaping the public debate would lead to a qualitative arrest of human activities. Public actors are no longer the only source of concern in the (algorithmic) marketplace of ideas. The lack of transparency and accountability

⁸⁷ Natalie J. Stroud, 'Polarization and Partisan Selective Exposure' (2010) 60(3) *Journal of Communication* 556.

⁸⁸ Natalie Helberger, 'Diversity by Design' (2011) 1 *Journal of Information Policy* 441.

⁸⁹ Nikolas Diakopoulos, 'Algorithmic Accountability. Journalistic Investigation of Computational Power Structures' (2014) 3 *Digital Journalism* 398.

⁹⁰ Martin Moore and Damian Tambini (eds.), *Digital Dominance. The Power of Google, Amazon, Facebook, and Apple* (Oxford University Press 2018).

⁹¹ Hannah Arendt, *The Human Condition* (University of Chicago Press 1998).

⁹² *Ibid.*, 199.

in online content moderation frustrates the exercise of freedoms in the public sphere encouraging to rethink the role of freedom of expression as a negative liberty in the algorithmic society. Platforms govern the flow of information online by defining, enforcing and balancing the right to freedom of expression online according to their business logics as the next subsection explains.

5.3.2 The Logic of Moderation

The shift from the free to the algorithmic marketplace of ideas can also be understood by focusing on the logic of moderation. Moderation can be defined as ‘the screening, evaluation, categorization, approval or removal/hiding of online content according to relevant communications and publishing policies. It seeks to support and enforce positive communications behaviour online, and to minimize aggression and anti-social behaviour’.⁹³ By focusing on the virtues of moderation, Grimmelmann has defined this process as ‘the governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse’.⁹⁴ Content moderation decisions can be entirely automated, made by humans or a mix of them. While the activities of pre-moderation like prioritisation, delisting and geo-blocking are usually automated, post-moderation is usually the result of a mix between automated and human resources.⁹⁵ This activity usually implies the use of different kinds of automated systems to manage vast amounts of information in different phases.⁹⁶

Moderation occurs before content is published (i.e. pre-moderation) or after publication (i.e. post-moderation). Precisely, post-moderation consists of the organisation of content, and it is implemented as a reactive measure to assess noticed content and as a proactive tool to actively monitor published content. Besides, removal is not the only way. For example, YouTube demonetises content by terminating any

⁹³ Terry Flew and others, ‘Internet Regulation as Media Policy: Rethinking the Question of Digital Communication Platform Governance’ (2019) 10(1) *Journal of Digital Media & Policy* 33, 40.

⁹⁴ James Grimmelmann, ‘The Virtues of Moderation’ (2015) 17 *Yale Journal of Law and Technology* 42, 47.

⁹⁵ Sarah T. Roberts, ‘Content Moderation’ in Laurie A. Schintler and Connie L. McNeely (eds.), *Encyclopedia of Big Data* (Springer 2017).

⁹⁶ Robert Gorwa, Reuben Binns and Christian Katzenbach, ‘Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance’ (2020) 7(1) *Big Data & Society* <https://journals.sagepub.com/doi/pdf/10.1177/2053951719897945> accessed 19 November 2021.

revenue sharing agreement with the content provider. This process can be a powerful tool to silence certain speakers who rely on YouTube as a source of income. Another alternative to content removal is down-ranking or shadow banning. In this case, content is deprioritised in news feeds and other recommendation systems. This constitutes an editorial decision on the organisation of content affecting how public discourse is shaped online. Platforms can decide whether certain content is visible and, therefore, affect its potential reach and dissemination.

These considerations only partially explain why moderation is a need for social media. As observed by Gillespie, ‘moderation is not an ancillary aspect of what platforms do. It is essential, constitutional, definitional. Not only can platforms not survive without moderation, they are not platforms without it’.⁹⁷ Moderation of online content is an almost mandatory step for social media not only to manage removal requests coming from governments or users but also to prevent that their digital spaces turn into hostile environments due to the spread, for example, of incitement to hatred. The implementation of these systems has become necessary as a filter to protect good expression from the massive presence of objectionable content.

However, the interest of platforms is not just focused on facilitating the spread of opinions and ideas across the globe to foster freedom of expression. They aim to create a digital environment where users feel free to share information and data that can feed commercial networks and channels and, especially, attract profits coming from advertising revenues.⁹⁸ Facebook, for instance, aims to maximise the amount of time users spend in their digital spaces to collect data and information.⁹⁹ Therefore, this approach leads to developing addictive technologies and capture users’ attention, for instance, with inflammatory content and a low degree of privacy.¹⁰⁰ In other words, the activity of content moderation is performed to attract revenues by ensuring a healthy online community, protecting

⁹⁷ Tarleton Gillespie, *Custodians of the Internet. Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* 21 (Yale University Press 2018).

⁹⁸ Tarleton Gillespie, ‘Regulation of and by Platforms’ in Jean Burgess, Alice E. Marwick and Thomas Poell (eds.), *The SAGE Handbook of Social Media* 254 (Sage 2018).

⁹⁹ Adam Alter, *Irresistible: The Rise of Addictive Technology and the Business of Keeping us Hooked* (Penguin Press 2017).

¹⁰⁰ Emily Bell and Taylor Owen, ‘The Platform Press: How Silicon Valley Reengineered Journalism’ Tow Centre for Digital Journalism (29 March 2017) www.cjr.org/tow_center_reports/platform-press-how-silicon-valley-reengineered-journalism.php accessed 21 November 2021.

the corporate image and showing commitments to ethical values. Within this business framework, users' data are the central product of online platforms under a logic of accumulation.¹⁰¹

The story of moderation legally began in the aftermath of the Internet. The Big Bang of moderation can indeed be connected to the system of online intermediaries' liability based on a liberal regulatory approach adopted by the United States and EU as described in Chapter 2. As for the evolution of the universe, it took some phases to make the digital environment profitable. It has been only with the first experiments of processing users' information for advertising that digital capitalism understood the potentialities of the digital environment.¹⁰²

At the end of the last century, there were no large corporations exercising powers in the digital environment. The political choice to follow a digital liberal path has led platforms to exploiting the legal framework to their advantage. According to Pasquale, online platforms try to avoid regulatory burdens by relying on the protection recognised by the First Amendment, while, at the same time, they claim immunities as passive conduits for third-party content.¹⁰³ Likewise, Citron and Norton observe how social media 'not only are free from First Amendment concerns as private actors, they are also statutorily immunized from liability for publishing content created by others as well as for removing that content'.¹⁰⁴ As Tushnet underlined, Section 230 'allows Internet intermediaries to have their free speech and everyone else's too'.¹⁰⁵

This framework leads to the content moderation paradox. Notwithstanding several social media exploit rhetoric statements advocating to represent a global community and enhance free speech transnationally,¹⁰⁶ however, online platforms need to moderate

¹⁰¹ Shoshana Zuboff, 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization' (2015) 30(1) *Journal of Information Technology* 75.

¹⁰² Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Public Affairs 2019).

¹⁰³ Frank Pasquale, 'Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power' (2016) 17 *Theoretical Inquiries in Law* 487.

¹⁰⁴ Danielle Keats Citron and Helen L. Norton, 'Intermediaries and Hate Speech: Fostering Digital Citizenship for our Information Age' (2011) 91 *Boston University Law Review* 1436, 1439.

¹⁰⁵ Rebecca Tushnet, 'Power Without Responsibility: Intermediaries and the First Amendment' (2008) 76 *The George Washington Law Review* 986, 1002.

¹⁰⁶ Mark Zuckerberg, 'Building Global Community' Facebook (16 February 2017) www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634/ accessed 21 November 2021.

content to protect their business interests. As observed by Roberts, ‘videos and other material have only one type of value to the platform, measured by their ability to either attract users and direct them to advertisers or to repel them and deny advertisers their connection to the user’.¹⁰⁷ An eventual escape of users due to the dissemination of content like terrorism and hate could severely harm advertising revenues. Other incentives are still linked to profit but come from concerns relating to corporate identity and reputation. For instance, online platforms aim to maintain control over the enforcement of their community guidelines and agreements to demonstrate that they act responsibly by complying with government requests relating to specific content like terrorist expressions.

At the same time, the grounding principle of content moderation is attracting profits by governing users’ attention.¹⁰⁸ The frequency of interaction, emotional reactions or comments are just some examples of the information which platforms can extract from users’ behaviours. This amount of information is then analysed to influence visibility and engagement which are usually fostered by matching similar content or standpoints according to micro-targeting strategies.¹⁰⁹ The numbers of likes or shares together with the analysis of users’ similarities are then used for moderating information online and profiting from advertising revenues.¹¹⁰ The revelations of platform’s whistle-blowers have contributed to confirming how the system of moderation tends to be driven by the logic of virality through engagement among users,¹¹¹ and the Facebook Files have confirmed the failure of online platforms to behave responsibly when moderating online content.¹¹² The spread of hate in

¹⁰⁷ Sarah T. Roberts, ‘Digital Detritus: “Error” and the Logic of Opacity in Social Media Content Moderation’ (2018) 23(3) *First Monday* <https://firstmonday.org/ojs/index.php/fm/rt/prINTERfriendly/8283/6649> accessed 21 November 2021.

¹⁰⁸ James G. Webster, ‘User Information Regimes: How Social Media Shape Patterns of Consumption’ (2010) 104 *Northwestern University Law Review* 593.

¹⁰⁹ Philipp M. Napoli, *Social Media and the Public Interest: Media Regulation in the Disinformation Age* (Columbia University Press 2019).

¹¹⁰ Engin Bozdag, ‘Bias in Algorithmic Filtering and Personalization’ 15(3) *Ethics and Information Technology* 209.

¹¹¹ Kari Paul and Dan Milmo, ‘Facebook Putting Profit Before Public Good, Says Whistleblower Frances Haugen’ *The Guardian* (4 October 2021) www.theguardian.com/technology/2021/oct/03/former-facebook-employee-frances-haugen-identifies-herself-as-whistleblower accessed 24 November 2021.

¹¹² See, e.g., Jeff Horwitz, ‘Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That’s Exempt’ *The Wall Street Journal* (13 September 2021) www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353?moD=article_inline accessed 24 November 2021.

Myanmar, or the attack at the Capitol Hill in the United States, are examples of the pitfalls of content moderation and how platforms could contribute to producing harms beyond digital boundaries, without mentioning the possibility that social media become instruments to further harm through surveillance and computational propaganda.¹¹³

Therefore, content as data is 'food' for feeding the business model of social media using algorithms which tend to show users content which is related to their algorithmic profile. This is not entirely new but based on the tendency of humans to create relationships with people who share their ideas and values, what has been called the 'homophily of networks'.¹¹⁴ This system also affects political speech by politicians or news media organisations.¹¹⁵ According to Sajó, 'instead of creating a common space for democratic deliberation, the Internet and social media enabled fragmentation and segmentation. Discourse is limited to occur within self-selecting groups and there are tendencies of isolation. Views are more extreme and less responsive to external arguments and facts, resulting in polarization around alternative facts'.¹¹⁶ The activity of content moderation indeed contributes to locking each user within personalised public spheres shaped by opaque business logics. Such a process turns online platforms into a manipulation machine.¹¹⁷ Put another way, no matter what kind of speech, this is in the filtering hands of online platforms.

This content moderation paradox explains why, on the one hand, social media commit to protecting free speech, while, on the other hand, they moderate content regulating their communities for business purposes. Therefore, one of the primary issues concerns the compatibility between their private interests and public values.¹¹⁸

¹¹³ Zeynep Tufekci, 'Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency' (2015) 13 *Colorado Technology Law Journal* 203.

¹¹⁴ Miller McPherson, Lynn Smith-Lovin and James M. Cook, 'Birds of a Feather: Homophily in Social Networks' (2001) 27 *Annual Review of Sociology* 415.

¹¹⁵ David Tewksbury and Jason Rittenberg, 'Online News Creation and Consumption: Implications for Modern Democracies' in Andrew Chadwick and Philipp N. Howard (eds.), *The Handbook of Internet Politics* 186 (Routledge 2008).

¹¹⁶ European Centre for Press and Media Freedom, 'Promoting Dialogue Between the European Court of Human Rights and the Media Freedom Community. Freedom of Expression and the Role and Case Law of the European Court of Human Rights: Developments and Challenges' (2017) www.ecpmf.eu/archive/files/ecpmf-ecthr_conference_e-book.pdf accessed 21 November 2021.

¹¹⁷ Siva Vaidyanathan, *Anti-Social Media* (Oxford University Press 2018).

¹¹⁸ José van Dijck, Thomas Poell, and Martijn de Waal, *The Platform Society: Public Values in a Connective World* (Oxford University Press 2018).

This situation is not only the result of the complexity of content moderation systems but also of the logic of opacity. Platforms are interested in pursuing their depoliticisation to escape from their social responsibilities coming from their key social functions. As argued by Roberts, platform tries to make the process obscure trying to denying 'the inherent gatekeeping baked in at the platform level by both its function as an advertising marketplace and the systems of review and deletion that have, until recently, been invisible to or otherwise largely unnoticed by most users'.¹¹⁹

To achieve this purpose, a critical piece of the moderation logic consists of the use of artificial intelligence systems. Platforms rely on automated technologies to cope with the amount of content uploaded by users whose non-automated management would require enormous costs in terms of human, technological and financial resources. Klonick has underlined the creation of a content moderation bureaucracy made of the work of humans and machines according to internal guidelines.¹²⁰ If, on the one hand, content moderation constitutes a valuable resource (and burden) for social media, on the other hand, the use of automated technologies for moderating content on a global scale challenges the protection of freedom of expression in the digital environment with effects extending far beyond domestic boundaries. The information uploaded by users is processed by automated systems that define (or at least suggest to human moderators) content to remove in a bunch of seconds according to non-transparent standards and providing the user access to limited remedies against a specific decision. It would not be possible to talk about content moderation online without considering to what extent algorithms are widely used for organising, filtering and removal procedures.¹²¹

The process (and the logic) of moderation is based on automated or semi-automated systems.¹²² Decisions about users' expressions are left to the discretion of machines (and unaccountable moderators) operating on behalf of online platforms.¹²³ These procedures govern all the

¹¹⁹ Roberts (n. 95).

¹²⁰ Klonick (n. 4).

¹²¹ Jennifer M. Urban and others, *Notice and Takedown in Everyday Practice* (American Assembly 2016).

¹²² Ben Wagner, *Global Free Expression: Governing the Boundaries of Internet Content* (Springer 2016).

¹²³ Paul M. Barrett, 'Who Moderates the Social Media Giants? A Call to End Outsourcing' NYU Stern (June 2020) https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/5ed9854bf618c710cb55be98/1591313740497/NYU+Content+Moderation+Report_June+8+2020.pdf accessed 22 November 2021.

phases of content moderation in the platform environment from indexing, organisation, filtering, recommendation and, eventually, removal of expressions and accounts. The role of human intervention is also critical,¹²⁴ even if this could not be the solution for digital firms like Facebook due to the high amount of content to moderate.¹²⁵

The pandemic has amplified these concerns and showed how the implementation of artificial intelligence to moderate content has contributed to spreading disinformation and to the blocking of accounts.¹²⁶ The decision of Google and Facebook to limit the employment of human moderation has affected the entire process with the result that different accounts and contents have been automatically suspended unnecessarily.¹²⁷ Notwithstanding the cooperative efforts of platforms to fight this situation,¹²⁸ the pandemic has underlined the limits of artificial intelligence in content moderation, particularly to tackle the spread of disinformation in a time where reliance over good health information has been critical.¹²⁹ This global health emergency has provided further clues concerning the role of online platforms as essential facilities or public utilities in the algorithmic society.¹³⁰

¹²⁴ Sarah T. Roberts, *Behind the Screen. Content Moderation in the Shadows of Social Media* (Yale University Press 2019); Paško Bilić, 'Search Algorithms, Hidden Labour and Information Control' (2016) 3(1) *Big Data & Society* 1.

¹²⁵ Jessica Lessin, 'Facebook Shouldn't Fact Check' *The New York Times* (29 November 2016) www.nytimes.com/2016/11/29/opinion/facebook-shouldnt-fact-check.html accessed 21 November 2021.

¹²⁶ Common position of European Commission and Consumer Protection Cooperation Network 20 March 2020 on stopping scams and tackling unfair business practices on online platforms in the context of the Coronavirus outbreak in the EU (20 March 2020) https://ec.europa.eu/info/sites/info/files/live_work_travel_in_the_eu_consumers/documents/cpc_common_position_covid19.pdf accessed 21 November 2021.

¹²⁷ Elizabeth Dwoskin and Nitasha Tiku, 'Facebook Sent Home Thousands of Human Moderators due to the Coronavirus. Now the Algorithms are in Charge' *The Washington Post* (24 March 2020) www.washingtonpost.com/technology/2020/03/23/facebook-moderators-coronavirus/ accessed 21 November 2021.

¹²⁸ See, e.g., joint industry statement of 17 March 2020 of Facebook, Google, LinkedIn, Microsoft, Reddit, Twitter and YouTube on working together to combat misinformation (16 March 2020) <https://about.fb.com/news/2020/06/coronavirus/> accessed 21 November 2021.

¹²⁹ Tobias R. Keller and Rosalie Gillett, 'Why Is It So Hard to Stop COVID-19 Misinformation Spreading on Social Media?' *The Conversation* (13 April 2020) <https://theconversation.com/why-is-it-so-hard-to-stop-covid-19-misinformation-spreading-on-social-media-134396> accessed 21 November 2021.

¹³⁰ K. Sabeel Rahman, 'The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept' (2018) 39 *Cardozo Law Review* 1621.

Within this framework, it is worth stressing that content moderation is not only a necessity for online platforms but also a way for governments to enforce public policies online, and even for surveillance.¹³¹ The case of India requiring Twitter to block more than 250 accounts of farmers protesting against a new farm law is just one example of how public authorities rely on online platforms to cope with dissent.¹³² Governments could potentially enforce their policies online. Nonetheless, it is a matter of technical capabilities and resources. It is indeed easier to regulate or even rely on gatekeepers (e.g. telco or online platforms) to address illicit content across multiple jurisdictions, without considering that some of the alleged wrongdoers could also be artificial like bots. As examined in Chapter 3, governments and online platforms can profit much more from the benefits of an indivisible handshake rather than from regulation.¹³³ On the one hand, regulating content moderation would decrease the flexibility to use online platforms as instruments of public surveillance or collection of data, transforming digital spaces from areas fostering free expression in a cage for liberties. On the other hand, online platforms aim to maintain a cooperative approach to protect their freedoms to run their business and limit attempts to increase regulatory pressures, unless regulation can create legal barrier to enter the market, thus increasing their power by liming competition.

Therefore, the cooperation between public and private actors is inside the logic of moderation, even if it could seem irrelevant or even invisible at first glance. This relationship is also the reason why the regulation of online platforms has not changed until recently and just in Europe. Balkin has underlined that ‘public/private cooperation – or cooptation – is a natural consequence of new-school speech regulation’.¹³⁴ Likewise, Reidenberg clarified that one of the systems to enforce public policies

¹³¹ Hannah Bloch Wehba, ‘Content Moderation as Surveillance’ (2021) 36 Berkeley Technology Law Journal 102.

¹³² Sangeeta Mahapatra, Martin Fertmann and Matthias C. Kettemann, ‘Twitter’s Modi Operandi: Lessons from India on Social Media’s Challenges in Reconciling Terms of Service, National Law and Human Rights Law’ *Verfassungsblog* (24 February 2021) <https://verfassungsblog.de/twitters-modi-operandi/> accessed 23 November 2021.

¹³³ Michael D. Birnhack and Niva Elkin-Koren, ‘The Invisible Handshake: The Reemergence of the State in the Digital Environment’ (2003) 8 *Virginia Journal of Law & Technology* 6.

¹³⁴ Jack M. Balkin, ‘Old-School/New-School Speech Regulation’ (2014) 127 *Harvard Law Review* 2296, 2305.

online consists of not only regulating the architecture of the digital environment but also of relying on online intermediaries.¹³⁵ Within this framework, governing by proxy online could be almost a mandatory step for public actors to address unlawful content online even if it raises high risks for fundamental rights and liberties as the next subsections underline in the case of freedom of expression.

5.3.3 Private Enforcement of Freedom of Expression

The mix of digital liberalism and algorithmic technologies is one of the reasons for the troubling scenario of online speech in the digital environment. The legal immunity, mixed together with profiling technologies to moderate content, has constituted a green light for online platforms to freely choose the values they want to protect and promote, no matter if democratic or anti-democratic and authoritarian. This is a perfect environment to profit while escaping responsibility. Since online platforms are private businesses, given the lack of incentives, they would likely focus on minimising economic risks rather than ensuring a fair balance between fundamental rights in the digital environment. In other words, the system of immunity has indirectly entrusted online platforms with the role of moderating content and encouraged them to develop new profitable automated systems to organise, select and remove content based on a standard of protection of free speech influenced by business purposes.

The scope of platform power can be better understood by focusing on how these actors set and enforce their internal rules of moderation after balancing conflicting interests. When organising, recommending or removing, platforms make decisions on which kind of speech should be protected or fostered.¹³⁶ This is evident in the process of removal reflecting some characteristics of the powers traditionally vested in public authorities as underlined in Chapter 3. Human moderators refer to community guidelines or internal documents as a 'private legal basis' to remove content. Social media usually provide ToS and community guidelines where they explain to users the acceptable conducts and content, creating 'a complex interplay between users and platforms, humans and algorithms, and the social norms and regulatory structures of social media'.¹³⁷

¹³⁵ Joel R. Reidenberg, 'States and Internet Enforcement' (2004) 1 *University of Ottawa Law & Technology Journal* 213.

¹³⁶ Hannah Bloch-Webba, 'Global Platform Governance: Private Power in the Shadow of the State' (2019) 72 *SMU Law Review* 27.

¹³⁷ Kate Crawford and Tarleton Gillespie, 'What Is a Flag for? Social Media Reporting Tools and the Vocabulary of Complaint' (2016) 18 *New Media & Society* 410, 411.

However, these community rules do not necessarily represent the reality of content moderation. Facebook, for example, relies on internal guidelines which users cannot access and whose drafting process is unknown.¹³⁸ According to Klonick, Facebook's content moderation is 'largely developed by American lawyers trained and acculturated in American free-speech norms, and it seems that this cultural background has affected their thinking'.¹³⁹ Whatever American or European values are at stake, this process is far from being close to any democratic value. Besides, the use of internal guidelines which are not publicly disclosed, leads to looking at this process more as an authoritarian determination than a democratic expression.

The situation is even more complicated when internal standards are solely implemented by machines which translate top-down rules in an enforceable series of code, defining another layer of complexity in the moderation of expressions. From a technical perspective, the opacity of content moderation also derives from the implementation of machine learning techniques subject to the 'black box' effect.¹⁴⁰ On the one hand, algorithms can be considered as technical instruments facilitating the organisation of online content. Nevertheless, on the other hand, such technologies can constitute opaque self-executing rules, obviating any human control with troubling consequences for democratic values such as transparency and accountability.

This mix of human and machine definition of freedom of expression constitutes the basis for enforcing decisions which are the results of a balance between conflicting interests. Taking as an example the case of hate speech, this concept is then mediated by the private determinations of human moderators or machines. This process then leads to the hybridisation of freedom of expression where traditional dichotomies like public/private or human/machine merge into a unique soul.

¹³⁸ Max Fisher, 'Inside Facebook's Secret Rulebook for Global Political Speech' *The New York Times* (27 December 2018) www.nytimes.com/2018/12/27/world/facebook-moderators.html accessed 21 November 2021.

¹³⁹ Klonick (n. 4), 1622.

¹⁴⁰ Frank Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information* (Harvard University Press 2015).

Within this framework, the lack of horizontal remedies leads online platforms to exercise the same discretion of an absolute power over its community. Despite the fundamental role of online platforms in establishing the standard of free speech and shaping democratic culture on a global scale,¹⁴¹ the information provided by these companies about content moderation is opaque or lawless.¹⁴² Online platforms are free to decide how to show and organise online content according to predictive analyses based on the processing of users' data. In other words, although at first glance social media foster freedom of expression by empowering users to share their opinions and ideas cross-border, however, the high degree of opacity and inconsistency of content moderation frustrates democratic values.

Content moderation does not only constitute an autonomous set of technical rules to control digital spaces but also contributes to defining the standard of protection of fundamental rights online, thus shaping the notion of public sphere and democracy. This situation leads towards computing legality as defined by a mere algorithmic calculation. The power of online platforms to shape the scope of protection of rights lies mostly in their ability to mathematically materialise abstract notions through digital means. Since artificial intelligence technologies are becoming more pervasive in online content moderation, the opacity of these technologies raises legal (and ethical) concerns for democracy.¹⁴³ Individuals are increasingly surrounded by technical systems influencing their decisions without the possibility to understand or control this phenomenon.¹⁴⁴ In other words, although the Internet has provided opportunities for users to access different types of information, the mediation of automated technologies leads to a process of hybridisation of freedom of expression becoming a mix of legal rules, platform guidelines and algorithmic determination. This trend towards computing abstract notions of law is a call for European digital constitutionalism to protect freedom of expression, and, more generally, constitutional values, in the algorithmic society.

¹⁴¹ Marvin Ammori, 'The "New" New York Times: Free Speech Lawyering in the Age of Google and Twitter' (2014) 127 *Harvard Law Review* 2259.

¹⁴² Nicolas Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* (Cambridge University Press 2019).

¹⁴³ Brent D. Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3(2) *Big Data & Society* <https://journals.sagepub.com/doi/pdf/10.1177/2053951716679679> accessed 21 November 2021.

¹⁴⁴ Paul Nemitz, 'Constitutional Democracy and Technology in the Age of Artificial Intelligence' (2018) 376 *Royal Society Philosophical Transactions A* 89.

5.4 The First Reaction of European Digital Constitutionalism

In the process of content moderation, users are not only subject to the private determinations of online platforms on freedom of expression but, more importantly, they cannot generally rely on procedural safeguards in this process. In other words, as observed by Myers West, ‘they are exactly the kinds of users who make up the kind of “town square”, “global village”, or “community” that these platforms themselves say they seek to cultivate – but current content moderation systems do not give them much opportunity to participate or grow as citizens of these spaces’.¹⁴⁵

From an international perspective, both the Manila principles on intermediary liability and the best practises proposed by the IGF Dynamic Coalition on Platform Responsibility are just two examples of proposals towards the proceduralisation of content moderation.¹⁴⁶ Similarly, the Santa Clara principles on Transparency and Accountability in Content Moderation suggest the adoption of due process safeguards regarding how content moderation should be performed and what rights users can rely on in the context of this process.¹⁴⁷ Article 19 has proposed the creation of social media councils based on a self-regulatory and multi-stakeholder system of accountability for content moderation complying with international human rights’ standards.¹⁴⁸ Likewise, in 2019, Facebook launched its oversight board.¹⁴⁹ At the same time, Twitter set an independent research group whose task is to develop standards for content moderation.¹⁵⁰

¹⁴⁵ Sarah Myers West, ‘Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms’ (2018) 20(11) *New Media & Society* 4380. See, also, Trevor Puetz, ‘Facebook: The New Town Square’ (2014) 44 *Southwestern Law Review* 385.

¹⁴⁶ Manila Principles on Intermediary Liability (2017) <https://manilaprinciples.org/index.html> accessed 21 November 2021; the DCPR Best Practices on Platforms’ Implementation on the Right to Effective Remedy www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/1550 accessed 20 November 2021.

¹⁴⁷ Santa Clara Principles on Transparency and Accountability in Content Moderation (2018) <https://santaclaraprinciples.org/> 20 November 2021.

¹⁴⁸ Article 19, ‘The Social Media Councils: Consultation Paper’ (2019) www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf 20 November 2021.

¹⁴⁹ Kate Klonick, ‘The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression’ (2020) 129 *Yale Law Journal* 2418; Evelyn Douek, ‘Facebook’s “Oversight Board:” Move Fast with Stable Infrastructure and Humility’ (2019) 21(1) *North Carolina Journal of Law & Technology* 1.

¹⁵⁰ Katie Paul and Munsif Vengattil, ‘Twitter Plans to Build “Decentralized Standard” for Social Networks’ *Reuters* (11 December 2019) www.reuters.com/article/us-twitter

However, despite the relevance of these steps, users still have to deal with discretionary and voluntary mechanisms. The lack of any binding force of this system leaves online platforms free to decide whether to participate in this mechanism or formally comply with these standards while maintaining their internal rules of procedures. At the same time, the former UN Special Rapporteur for Freedom of Expression, David Kaye, underlined the increasing pressure on private actors to comply with international human rights law when moderating online content.¹⁵¹ According to Kaye, since social media exercise regulatory functions in the digital environment, these private actors should refer to the existing international human rights law regime when setting their standard for content moderation.¹⁵² International human rights law could help platforms apply a universal reference in their activities of content moderation, but there are still challenges concerning the promises of human rights law in content moderation.¹⁵³

However, as already underlined, since online platforms are private actors, they are not obliged to respect human rights since international human rights law vertically binds only state actors with the result that the governance of online platforms is based on fragmented national and regional laws as well as soft-regulatory efforts.¹⁵⁴ The same consideration extends to fundamental rights since constitutional provisions bind only public actors to respect them even if there could be some cases where fundamental rights horizontally apply in the relationship between private actors.¹⁵⁵ Despite the role of self-regulation and

[-content/twitter-plans-to-build-decentralized-standard-for-social-networks-idUSKBN1YF2EN](#) accessed 21 November 2021.

¹⁵¹ David Kaye, *Speech Police: The Global Struggle to Govern the Internet* (Columbia Global Reports 2019).

¹⁵² Report of the Special Rapporteur to the Human Rights Council on online content regulation, A/HRC/38/35 (2018); See, also, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/73/348 (2018); Guiding Principles on Business and Human Rights (2011).

¹⁵³ 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(4) *Fordham Journal of International Law* 939.

¹⁵⁴ Jennifer Grygiel and Nina Brown, 'Are Social Media Companies Motivated to Be Good Corporate Citizens? Examination of the Connection Between Corporate Social Responsibility and Social Media Safety' (2019) 43 *Telecommunications Policy* 445.

¹⁵⁵ Some constitutions around the world (e.g. South Africa) horizontally extends the application of fundamental rights in the relationship between private actors. In other case, horizontal application is not the result of a direct constitutional provision but the result of judicial interpretation.

corporate social responsibility in building a shared global framework which could overcome any regulatory vacuum,¹⁵⁶ the remedies voluntarily provided by online platforms are highly fragmented and left to their discretion.¹⁵⁷ Moreover, the differences between (publicly available) community guidelines and (privately hidden) internal policy as well as the opacity about the use of automated systems in content moderation create a grey area of cases where organisation, recommendation and removal of content are set outside any democratic control.

While, in the US, the legal framework has not changed in the last twenty years, apart from some exception,¹⁵⁸ and the executive order on preventing online censorship adopted in 2020 which was then withdrawn by President Biden,¹⁵⁹ the Union has started to pave the way towards a new regulatory phase of online content moderation modernising the framework of the e-Commerce Directive.¹⁶⁰ The European objectives to protect constitutional values could be considered the political manifesto of the new European approach.¹⁶¹ Such a shift towards wider responsibilities is not a mere political decision but the expression of the first steps of European digital constitutionalism.¹⁶² As underlined in Chapter 2, the Directive on Copyright in the Digital Single Market,¹⁶³ the amendments to the Audiovisual Media Service

¹⁵⁶ Rolf H. Weber, 'Corporate Social Responsibility As a Gap-Filling Instrument' in Andrew P. Newell (ed.), *Corporate Social Responsibility: Challenges, Benefits and Impact on Business* 87 (Nova 2014).

¹⁵⁷ IGF Dynamic Coalition, 'Best Practices on Platforms' Implementation of the Right to an Effective Remedy' (2018) www.intgovforum.org/multilingual/content/dcpr-best-practices-on-due-process-safeguards-regarding-online-platforms-implementation-of accessed 20 November 2021.

¹⁵⁸ See the Stop Enabling Sex Traffickers Act (SESTA) and the Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA) adopted in 2018.

¹⁵⁹ Executive Order on Preventing Online Censorship (28 May 2020) www.federalregister.gov/documents/2020/06/02/2020-12030/preventing-online-censorship accessed 22 November 2021.

¹⁶⁰ Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce') (2000) OJ L 178/1.

¹⁶¹ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Online Platforms and the Digital Single Market Opportunities and Challenges for Europe COM (2016) 288 final.

¹⁶² Ibid.

¹⁶³ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (2019) OJ L 130/92.

Directive,¹⁶⁴ as well as the Regulation on terrorist content,¹⁶⁵ have constituted a first turning point in online content moderation, requiring online platforms to establish transparent and accountable mechanisms.

These measures are part of a broader strategy of the Union to foster accountability and transparency in online content moderation. Just to mention two examples, it would be enough to refer to the Code of Conduct on Countering Illegal Hate Speech Online and the Code of Practice on Online Disinformation,¹⁶⁶ resulting from the Communication on Tackling Online Disinformation and, especially, the Communication on tackling illegal content online,¹⁶⁷ then implemented in the Recommendation on measures to effectively tackle illegal content online.¹⁶⁸

The approach of the Union in this field shows a shift from a liberal approach in online content moderation to transparency and accountability obligations and recommendations. Rather than just focusing on content regulation, the European approach focuses on introducing procedural safeguards to dismantle the logic of opacity.

In the meantime, in *Eva Glawischnig-Piesczek v. Facebook Ireland Limited*,¹⁶⁹ the ECJ has contributed to providing guidance in the process of content moderation in a case involving the removal of identical and equivalent content. The ECJ underlined the role of social media in promoting the dissemination of information online, including illegal content. In this case, a national judge's order of removing or blocking identical content does not conflict with the monitoring ban established

¹⁶⁴ Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities (2018) OJ L 303/69.

¹⁶⁵ Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online OJ L 172/79.

¹⁶⁶ Code of Conduct on Countering Illegal Hate Speech Online (2016) http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=54300 accessed 21 November 2021; Code of Practice on Disinformation (2018) <https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation> accessed 21 November 2021.

¹⁶⁷ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Tackling Illegal Content Online Towards an enhanced responsibility of online platforms COM(2017) 555 final.

¹⁶⁸ Recommendation of 1 March 2018 on measures to effectively tackle illegal content online (C(2018) 1177 final).

¹⁶⁹ Case C-18/18 *Eva Glawischnig-Piesczek v. Facebook Ireland Limited* (2019).

by the e-Commerce Directive.¹⁷⁰ As the Advocate General Szpunar underlines, an order to remove all identical information does not require ‘active non-automatic filtering’.¹⁷¹ The ECJ addressed the question concerning the removal of ‘equivalent’ content. According to the court, in order to effectively cease an illegal act and prevent its repetition, the order of the national judge has to be able to also extend to ‘equivalent’ content defined as ‘information conveying a message the content of which remains essentially unchanged and therefore diverges very little from the content which gave rise to the finding of illegality’.¹⁷² Otherwise, users would only access a partial remedy that could lead to resorting to an indefinite number of appeals to limit the dissemination of equivalent content.¹⁷³

However, such an extension is not unlimited. The ECJ reiterated that the ban on imposing a general surveillance obligation established by the e-Commerce Directive is still the relevant threshold for Member States’ judicial and administrative orders. If, on the one hand, the possibility of extending the orders of the national authorities to equivalent content aims to protect the victim’s honour and reputation, on the other hand, such orders cannot entail an obligation for the hosting provider to generally monitor information to remove equivalent content. In other words, the ECJ defined a balance between, on the one hand, the freedom of economic initiative of the platform, and, on the other, the honour and reputation of the victim. The result of such a balance, therefore, leads to reiterate that the national orders of the judicial and administrative authorities have to be specific without being able to extend to the generality of content.

In order to balance these conflicting interests, the ECJ provided other conditions applying to equivalent content. Precisely, expressions have to contain specific elements duly identified by the injunction such as ‘the name of the person concerned by the infringement determined previously, the circumstances in which that infringement was determined and equivalent content to that which was declared to be illegal’.¹⁷⁴ Under these conditions, the protection granted to the victim would not constitute an excessive obligation on the hosting provider

¹⁷⁰ *Ibid.*, 37.

¹⁷¹ Opinion of Advocate General in C-18/18 *Eva Glawischnig-Piesczek v. Facebook Ireland Limited*, 61.

¹⁷² C-18/18 (n. 169), 39.

¹⁷³ *Ibid.*, 41.

¹⁷⁴ *Ibid.*, 45.

since its discretion is limited to certain information without leading to general monitoring obligation that could derive from an autonomous assessment of the equivalent nature of the content. If, on the one hand, the ECJ clarified how platforms should deal with users' requests for removal of identical and equivalent content, nonetheless, even in this case, the court did not define transparency and accountability safeguards in the process of content moderation.

These first steps of European digital constitutionalism have not solved the asymmetry of power in the field of content. Users and online platforms still face challenges raised by legal fragmentation in this field. There is not a unitary framework of users' rights or remedies, also considering that Member States enjoy margins of discretion in implementing such safeguards. Besides, safeguards in online content moderation have not been introduced horizontally to cover all content and situation. The Union has maintained a vertical approach based on specific categories of content (e.g. copyright content). The fragmentation of content moderation processes can lead to serious consequences for the freedom to conduct business of online platforms and, as a consequence, this uncertainty could produce chilling effects for users' freedom of expression. As analysed further in this chapter, the Digital Services Act provides an opportunity to complete this framework and provide a systematic horizontal approach to ensure more safeguards and remedies in the process of content moderation.¹⁷⁵

Therefore, it is time to focus on how the new phase of European digital constitutionalism can provide instruments to address the imbalance of power between users and online platforms in the field of content. There are two ways addressed in the next sections, which look respectively at the horizontal effect doctrine and at the regulation of content moderation as also driven by the Digital Services Act.

5.5 Horizontal Effect Filling Regulatory Gaps

Within this troubling framework for democratic values in the algorithmic society, the question is whether European constitutional law already owns the instruments to react, even without regulatory

¹⁷⁵ Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM(2020) 825 final.

intervention. Whereas proposing a regulatory solution would be a largely traditional approach, it is necessary to step back and wonder about the role of constitutional law in content moderation. Even if, in Europe, lawmakers have seemed to be prone to regulate online platforms, on the one hand, the interest of public actors to monitor online activities and enforce public policies online should not be neglected. On the other hand, online platforms aim to maintain their freedom to conduct business outside regulatory interferences. This apparently unrelated but converging interest leads to an invisible cooperation between public and private actors, thus creating a powerful brake to regulatory intervention. Such a situation could lead to potential conflicts of interest since political power could not regulate online platforms to protect forms of unaccountable cooperation.

To overcome this political impasse, one of the few ways to move further is to look beyond political powers and, precisely, at judicial power. In other words, it may be possible to rely on courts, and their independence, to ensure that the protection of fundamental rights is not locked down between political and business interests but is interpreted within the evolving information society. This approach would lead to wondering to what extent the horizontal effect doctrine of fundamental rights in Europe could be a solution to remedy the imbalance of power between users and online platforms exercising private powers on online speech.

The horizontal doctrine may promise to go beyond the public/private division by extending constitutional obligations even to the relationship between private actors (i.e. platform/user). Unlike the liberal spirit of the vertical approach, this theory rejects a rigid separation where constitutional rules apply vertically only to public actors to ensure the liberty and autonomy of private actors. Put another way, the horizontal doctrine is concerned with the issue of whether and to what extent constitutional rights can affect the relationships between private actors. As observed by Gardbaum, '[t]hese alternatives refer to whether constitutional rights regulate only the conduct of governmental actors in their dealings with private individuals (vertical) or also relations between private individuals (horizontal)'.¹⁷⁶

The horizontal effect can result from constitutional obligations on private parties to respect fundamental rights (i.e. direct effect) or their

¹⁷⁶ Stephen Gardbaum, 'The Horizontal Effect of Constitutional Rights' (2003) 102 *Michigan Law Review* 388.

application through judicial interpretation (i.e. indirect effect). Only in the first case, a private entity would have the right to rely directly on constitutional provisions to claim the violation of its rights vis-à-vis other private parties.¹⁷⁷ There is also a third (indirect) way through the positive obligations for states to protect human rights such as in the case of the Convention.¹⁷⁸

The horizontal application of fundamental rights could constitute a limitation to the expansion of power by social systems. According to Teubner, the emergence of transnational regimes shows the limits of constitutions as means of regulation of the whole society since social subsystems develop their own constitutional norms.¹⁷⁹ Therefore, the horizontal effects doctrine can be considered as a limit to self-constitutionalising private regulation. As a result, if the horizontal effect of fundamental rights is purely considered a problem of political power within society, an approach which excludes its application would hinder the teleological approach behind this doctrine, the aim of which is to protect individuals against unreasonable violations of their fundamental rights vis-à-vis private actors. As Tushnet underlined, if the doctrine of horizontal effect is considered ‘as a response to the threat to liberty posed by concentrated private power, the solution is to require that all private actors conform to the norms applicable to governmental actors’.¹⁸⁰

Nonetheless, the horizontal application of fundamental rights does not apply in the same way across the Atlantic. Within the US framework, the Supreme Court has usually applied the vertical approach where the application of the horizontal approach, known in the US as the ‘state action doctrine’, would be considered the exception.¹⁸¹ The First Amendment, and, more in general, US constitutional

¹⁷⁷ John H. Knox, ‘Horizontal Human Rights Law’ (2008) 102(1) *American Journal of International Law* 1.

¹⁷⁸ Daniel Augenstein and Lukasz Dziedzic, ‘State Responsibilities to Regulate and Adjudicate Corporate Activities under the European Convention on Human Rights’ (2017) EUI Working papers https://cadmus.eui.eu/bitstream/handle/1814/48326/LA_W_2017_15.pdf?sequence=1&isAllowed=y accessed 21 November 2021.

¹⁷⁹ Gunther Teubner, ‘The Project of Constitutional Sociology: Irritating Nation State Constitutionalism’ (2013) 4 *Transnational Legal Theory* 44.

¹⁸⁰ Mark Tushnet, ‘The Issue of State Action/Horizontal Effect in Comparative Constitutional Law’ (2003) 1(1) *International Journal of Constitutional Law* 79.

¹⁸¹ See *Shelley v. Kraemer* 334 U.S. 1 (1948). Mattias Kumm and Victor Ferreres Comella, ‘What Is So Special about Constitutional Rights in Private Litigation? A Comparative Analysis of the Function of State Action Requirements and Indirect Horizontal Effect’ in Andras Sajó and Renata Uitz (eds.), *The Constitution in Private Relations: Expanding*

rights,¹⁸² lack horizontal effect not only *in abstracto* but also in relation to online platforms.

Even if scholars have tried to propose new ways to go beyond such a rigid verticality,¹⁸³ the Supreme Court has been clear about the limits of this doctrine when addressing the possibility that a non-profit corporation designated by New York City to run a public access television network limit users' speech.¹⁸⁴ In an ideological 5–4 ruling, the court rejected the idea that the TV station in question could be considered a state actor, and, therefore, there was no reason to focus on the violation of the First Amendment. Although this case concerned public access channels, the property-interest arguments could have a broad impact in the information society, precisely on the protection of speech over online platforms'. This would lead towards Balkin's warning about the limit of 'judge-made doctrines' of the First Amendment.¹⁸⁵

The horizontal extension of fundamental rights is less rigid in the European environment,¹⁸⁶ and it is characterised by different models.¹⁸⁷ As already underlined in Chapter 1, one of the primary explanations for the extension of constitutional values beyond a vertical dimension lies in the roots of European constitutionalism, precisely in the protection of human dignity.¹⁸⁸ This approach is also reflected in the social democratic openness of Member States and the

Constitutionalism 265 (Eleven 2005); Mark Tushnet, 'Shelley v. Kraemer and Theories of Equality' (1988) 33 *New York Law School Law Review* 383.

¹⁸² The prohibition on slavery as provided for by the Thirteenth Amendment applies to public and private actors. Gardbaum (n. 176) 388; George Rutherglen, 'State Action, Private Action, and the Thirteenth Amendment' (2008) 24(6) *Virginia Law Review* 1367.

¹⁸³ Jonathan Peters, 'The "Sovereigns of Cyberspace" and State Action: The First Amendment's Application (or Lack Thereof) to Third-Party Platforms' (2018) 32 *Berkeley Technology Law Journal* 988; Lyriisa B. Lidsky, 'Public Forum 2.0' (2011) *Boston University Law Review* 1975; Paul S. Berman, 'Cyberspace and the State Action Debate: The Cultural Value of Applying Constitutional Norms to "Private" Regulation' (2000) 71 *University of Colorado Law Review* 1263.

¹⁸⁴ *Manhattan Community Access Corp. v. Halleck*, No. 17–1702, 587 U.S. ____ (2019).

¹⁸⁵ Jack M. Balkin, 'The Future of Free Expression in a Digital Age' (2009) 36 *Pepperdine Law Review* 427, 443–4.

¹⁸⁶ Regarding the horizontal effect of fundamental rights in the EU framework, see Eleni Frantziou, *The Horizontal Effect of Fundamental Rights in the European Union. A Constitutional Analysis* (Oxford University Press 2019); Sonya Walkila, *Horizontal Effect of Fundamental Rights in EU Law* (European Law Publishing 2016).

¹⁸⁷ Aurelia Colombi Ciacchi, 'Judicial Governance in European Private Law: Three Judicial Cultures of Fundamental Rights Horizontality' (2020) 4 *European Review of Private Law* 931.

¹⁸⁸ Catherine Dupré, *The Age of Dignity: Human Rights and Constitutionalism in Europe* (Hart 2015).

European area which is far from the liberal approach of the US framework. According to Tushnet, states which are more oriented to develop welfare systems and provide social rights in their constitutions more readily apply the horizontal effect doctrine. This position should not surprise since it is the natural consequence of how rights and freedoms are conceived in welfare states. The positive and programmatic nature of some constitutional rights leads to a broader role for lawmakers but, especially, for courts to define the limits of these rights. It is not by chance that, in the European framework, the doctrine of the horizontal effect has found an extensive application in the field of labour law.¹⁸⁹

The European horizontal effect doctrine is far from being locked just in the field of social rights. Traditionally, the effects of the rights recognised directly under EU primary law have been capable of horizontal application. The ECJ has applied both the horizontal effect and the positive obligation doctrines regarding the four fundamental freedoms and general principles.¹⁹⁰ In the *Van Gend En Loos* case, the ECJ stated: 'Independently of the legislation of Member States, Community law not only imposes obligations on individuals but is also intended to confer upon them rights which become part of their legal heritage'.¹⁹¹ This definition remained unclear until the court specified its meaning in *Walrave*,¹⁹² which, together with *Bosman*¹⁹³ and *Deliege*,¹⁹⁴ can be considered the first acknowledgement of the horizontal effect of the EU fundamental freedoms.¹⁹⁵

Likewise, since the Charter acquired the same legal value of a Treaty,¹⁹⁶ judicial activism has also been extended to the Charter.¹⁹⁷ Recently, in

¹⁸⁹ See Case 43/75 *Defrenne v. Sabena (No 2)* (1976) ECR 455. More recently, Case C-555/07 *Kücükdeveci v. Sweden* (2010) ECR I-365; Case C-144/04 *Mangold v. Rüdiger Helm* (2005) ECR I-9981. But see Case C-176/12 *Association de médiation sociale v. Union locale des syndicats CGT* (2014).

¹⁹⁰ Elena Gualco and Luisa Lourenço, "Clash of Titans". General Principles of EU Law: Balancing and Horizontal Direct Effect' (2016) 1(2) European Papers 643.

¹⁹¹ Case 26/62 *van Gend & Loos v. Netherlands Inland Revenue Administration* (1963) ECR 1.

¹⁹² Case 36/74 *Walrave v. Association Union cycliste internationale* (1974) ECR 1405.

¹⁹³ Case C-415/93 *Union royale belge des sociétés de football association v. Bosman* (1995) ECR 4921.

¹⁹⁴ Case C-51/96 *Deliege v. Ligue francophone de judo et disciplines associées* (2000) ECR I-2549.

¹⁹⁵ Among the other decisions, see Case C-281/98 *Angonese v. Cassa di Risparmio di Bolzano* (2000) ECR I-2055; Case C-103/08 *Gottwald v. Bezirkshauptmannschaft Bregenz* (2009) ECR I-9117; Case C-223/09 *Dijkman v. Belgische Staat* (2010) ECR I-6649.

¹⁹⁶ Consolidated version of the Treaty on European Union (2012) OJ C 326/13, Art. 6(1). Grainne De Burca and Jo B. Aschenbrenner, 'The Development of European Constitutionalism and the Role of the EU Charter of Fundamental Rights' (2003) 9 Columbia Journal of European Law 355.

¹⁹⁷ Dorota Leczykiewicz, 'Horizontal Application of the Charter of Fundamental Rights' (2013) 38(3) European Law Review 479.

Egenberger,¹⁹⁸ the ECJ extended horizontal application to the right of non-discrimination and the right to an effective remedy and to a fair trial, respectively enshrined in Articles 21 and 47 of the Charter, in a case involving compensation for discrimination on the grounds of religion suffered in a recruitment procedure. Likewise, in *Bauer*,¹⁹⁹ the Court went even further. The ECJ did not only extend horizontal effects to the right to limitation of maximum working hours as fair and just working condition,²⁰⁰ but also overcame its precedents in *Association de médiation sociale*, where it rejected horizontal effects to the workers' right to information and consultation.²⁰¹ In *Bauer*, the ECJ clarified that the narrow scope of Article 51(1) does not deal with whether individuals, or private actors, may be directly required to comply with certain provisions of the Charter.²⁰²

With regard to the right to freedom of expression as enshrined in the Charter,²⁰³ the ECJ has not still provided its guidance. A literal interpretation of Article 11 of the Charter could constitute a barrier to any attempt to extend its scope of application. Likewise, Article 51(1) of the Charter seems to narrow down the scope of application of the Charter to EU institutions and Member States in their implementation of EU law.²⁰⁴ Brkan warned about the risk for the system of European competences relating to the introduction of a positive obligation in the field of freedom of expression to fill the legislation gap.²⁰⁵ Indeed, 'in creating such a positive obligation, the CJEU would not only have to observe the principles of conferral and subsidiarity, but also pay attention not to overstep its own competences by stepping into the shoes of a legislator'.²⁰⁶ This issue, however, has not discouraged the ECJ to underline the

¹⁹⁸ Case C-414/16 *Vera Egenberger v. Evangelisches Werk für Diakonie und Entwicklung e.V.* (2018).

¹⁹⁹ Case C-569/16 *Stadt Wuppertal v. Maria Elisabeth Bauer and Volker Willmeroth v. Martina Broßmann* (2018).

²⁰⁰ Charter (n. 40), Art. 31(2).

²⁰¹ C-176/12 (n. 189), 51.

²⁰² C-569/16 (n. 199), 87.

²⁰³ Charter (n. 40), Art. 11.

²⁰⁴ *Ibid.* According to Art. 51(1): 'The provisions of this Charter are addressed to the institutions and bodies of the Union with due regard for the principle of subsidiarity and to the Member States only when they are implementing Union law. They shall therefore respect the rights, observe the principles and promote the application thereof in accordance with their respective power'.

²⁰⁵ Maja Brkan, 'Freedom of Expression and Artificial Intelligence: On Personalisation, Disinformation and (Lack Of) Horizontal Effect of the Charter' SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3354180 accessed 21 November 2021.

²⁰⁶ *Ibid.*

relevance of the right to freedom of expression online in private litigations.²⁰⁷ The court underlined that interferences with freedom of expression would not be justified in case the measures adopted by the provider are not 'strictly targeted, in the sense that they must serve to bring an end to a third party's infringement of copyright or of a related right but without thereby affecting internet users who are using the provider's services in order to lawfully access information'.²⁰⁸

The reasons for an alleged lack of horizontality are not only rooted in the separation between judicial and political power but also depend on the constitutive difference between negative liberties and positive rights. As Beijer underlined, in the Union framework, there is less pressure to rely on positive obligations based on the violation of fundamental rights since obligations are horizontally translated in acts of EU law.²⁰⁹ The approach of the ECJ does not surprise since the field of labour law can be considered as one of the primary expressions of the welfare conception. The extension of such a rule to the principle of non-discrimination aims to ensure not only formal but also substantive equality between individuals. In this framework, the right to freedom of expression is instead conceived within the framework of negative liberties which only consider public actors as a threat. In other words, it is not just a matter of literal interpretation of Article 11 of the Charter but also of theoretical distance, even if the common matrix of human dignity in European constitutionalism could provide that constitutional ground to extend the (horizontal) effects to freedom of expression.

Besides, within the complexity of the horizontal effect doctrine,²¹⁰ it is worth highlighting at least a primary drawback, which can also be applied to content moderation. While the horizontal effect doctrine could be a constitutional instrument to generally mitigate the exercise of private powers on freedom of expression, nonetheless, the extension of obligations to respect constitutional rights to online platforms would raise several concerns. Applying this doctrine extensively could lead to negative effects for legal certainty. Every private conflict can virtually be represented as a clash between different fundamental rights. This

²⁰⁷ Case C-314/12 *UPC Telekabel Wien GmbH v. Constantin Film Verleih GmbH and Wega Filmproduktionsgesellschaft mbH* (2014); Case C-484/14 *Tobias Mc Fadden v. Sony Music Entertainment Germany GmbH* (2016).

²⁰⁸ *Telekabel* (n. 207), 56. See also *McFadden* (n. 207), 93.

²⁰⁹ Malu Beijer, *The limits of Fundamental Rights Protection by the EU: The Scope for the Development of Positive Obligations* 297 (Intersentia 2017).

²¹⁰ Robert Alexy, *A Theory of Constitutional Rights* (Oxford University Press 2002).

approach could lead to the extension of constitutional obligations to every private relationship, thus hindering any possibility to foresee the consequences of a specific action or omission. Fundamental rights can be applied horizontally only *ex post* by courts through the balancing of the rights in question.

It cannot be excluded that this approach could be even more multifaceted in civil law countries where judges are not legally bound by precedents, but can take their own path to decide whether constitutional obligations apply to private litigations or not.²¹¹ In Chapter 2, the judicial activism of the ECJ has already shown the role of courts in ensuring that the protection of fundamental rights is not frustrated in the digital environment. The further empowerment of judicial over political power could lead to increasing fragmentation and uncertainty about content moderation obligations, thus undermining the principle of the separation of powers and rule of law. This is not something far from reality. While, in the US, courts continue to ban any users' complaints against the removal of content,²¹² some cases in Europe have shown how courts have already dealt with the horizontal extension of constitutional rights in private litigations between users and online platforms, also leading to different outcomes.²¹³

These concerns around judicial power could be partially overcome by limiting the application of the horizontal effect only to those private actors exercising delegated public functions, as seen in Chapter 3. In the case of platforms, although these entities cannot be considered public actors *per se*, their delegated public functions to moderate content (e.g. obligation to remove illicit content in case of awareness) could be subject to the safeguards applying to the public

²¹¹ The difference between common law and civil law should not be considered rigid. Nonetheless, the constitutive differences in the role of courts deserve to be mentioned when focusing on the limits of the horizontal effect doctrine. See, generally, Paul Brand and Joshua Getzler (eds.), *Judges and Judging in the History of the Common Law and Civil Law: From Antiquity to Modern Times* (Cambridge University Press 2015); Joseph Dainow, 'The Civil Law and the Common Law: Some Points of Comparison' (1966–7) 15(3) *American Journal of Comparative Law* 419.

²¹² Niva Elkin Koren, Giovanni De Gregorio and Maayan Perel, 'Social Media as Contractual Networks: A Bottom up Check on Content Moderation' *Iowa Law Review* forthcoming; Daphne Keller, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech' (2019) Aegis Series Paper No. 1902 (29 January 2019) www.hoover.org/sites/default/files/research/docs/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech_0.pdf accessed 21 November 2021.

²¹³ See, e.g., Court of Rome, *CasaPound v. Facebook* (2019); German Federal Constitutional Court, *Der Dritte Weg v. Facebook Ireland Ltd.* (2019).

sector (e.g. transparency). In other words, constitutional law would extend its horizontal boundaries only where public actors entrust private actors with quasi-public functions through delegation of powers. Users have a legitimate expectation that, when public actors have entrusted private ones to pursue public tasks, the latter should be held accountable for violation of constitutional rights and freedoms. On the contrary, where platforms exercise autonomous powers, a broad extension of the horizontal effect doctrine would transform these entities into public actors by default. This approach would provide users with the right to bring claims related to violations of, for example, freedom of expression directly against platforms as entities performing delegated public functions.

At first glance, this mechanism would allow fundamental rights to become horizontally effective against the conduct or omission of actors evading their responsibilities under a narrative based on freedoms and liberties. However, a closer look could reveal how empowering users to challenge online platforms could lead to a compression of the freedom to conduct business of these actors. Such interference could not be tolerated under a European constitutional perspective. Freedom of expression is not an absolute right with the result that its protection cannot lead to the destruction of other constitutional interests.

Besides, requiring online platforms not to censor content or generally avoid interferences with freedom of expression (e.g. must-carry obligations) could affect the process of content moderation, thus making platforms' spaces more exposed to objectionable content. This situation would undermine not only the freedom to conduct business of online platforms which would lose advertising revenues but also democratic values online since users would be more exposed to harmful content, thus reducing their freedom to share opinions and ideas online.

Therefore, the horizontal effect doctrine cannot always provide a stable solution for the imbalances between public and private power in the algorithmic society. It could be a reactive remedy which would not be able to comprehensively mitigate the challenges of content moderation. This consideration does not imply that judges could not play a critical role in protecting constitutional values from technological annihilation.²¹⁴ On the one hand, this doctrine would perfectly

²¹⁴ Oreste Pollicino, *Judicial Protection of Fundamental Rights on the Internet: A Road Towards Digital Constitutionalism?* (Hart 2021).

match with the reactive side of European digital constitutionalism. On the other hand, it would fail to provide the other side of this constitutional phase, namely a normative framework based on the injection of democratic values online to deal with private powers in the long run.

There is also another chance for freedom of expression to mitigate and remedy the challenges of content moderation. By moving from a negative to a positive dimension, it is possible to look at freedom of expression not only as a negative liberty but also as a positive right. This is not a call to define the welfare of freedom of expression but to understand how to foster media pluralism in the digital environment. Likewise, this system would not just focus on directly empowering users to decide on the removal of content. As observed by Rosen, 'a user-generated system for enforcing community standards will never protect speech as scrupulously as unelected judges enforcing strict rules'.²¹⁵ The approach of European digital constitutionalism focuses on transparency and procedural safeguards to ensure more autonomy and diversity of online content.

The role of digital constitutionalism is not just to provide new solutions but also to reframe old categories into the evolving technological scenario. As the next section suggests, in order to limit the significant power of online platforms over constitutional rights and freedoms, it is not necessary to provide more access but to understand how to foster media pluralism in the algorithmic society by promoting diversity and transparency in content moderation.

5.6 Rethinking Media Pluralism in the Age of Online Platforms

The challenges of content moderation at the European level would require a more comprehensive strategy which is not only reactive but also promotes the development of a democratic public sphere. The fragmentation of substantive obligations and procedural safeguards and the limit of the horizontal effect does not seem to provide a stable framework to remedy platform power. Even if the first steps of European digital constitutionalism have led to a shift in the European approach to content moderation, still the lack of systemic remedies

²¹⁵ Jeffrey Rosen, 'The Deciders: The Future of Privacy and Free Speech in the Age of Facebook and Google' (2012) 80 *Fordham Law Review* 1525.

could increase uncertainty, thus undermining not only fundamental rights but also the principle of the rule of law. This consideration does not mean that the path of European digital constitutionalism has not designed a turning point, but the fragmentation of legal regimes influencing content moderation would introduce more risks than advantages, even for online platforms.

Consequently, it is worth wondering how European constitutional law can provide other ways to remedy the challenges to the right of freedom of expression in a public sphere which is characterised by opacity and lack of accountability. In the context of traditional media outlets, media pluralism would have been one of the primary ways to ensure more diversity and transparency, thus fostering the positive and passive dimension of the right to freedom of expression.²¹⁶ Together with media freedom, pluralism is a precondition for an open and dialectic debate in a democratic society. Granting access to vast and diversified sources of information increases individual exposure to different ideas and opinions contributing to a democratic public sphere. In the digital age, even if there are multiple definitions of media pluralism online,²¹⁷ and especially how to measure its effect,²¹⁸ users are exposed to content subject to the opaque governance of online platforms which do not provide users with any instruments to understand how their expressions are moderated online.

In order to fix the challenges of the algorithmic public sphere, it is worth understanding how to reframe media pluralism in the age of online platforms. In particular, ensuring access and diversity of information online contributes to ensuring that individuals are not just exposed to polarised information or harmful content. This approach is critical to ensure individual autonomy and dignity while promoting a dialectic relationship in a democratic society.

²¹⁶ Damian Tambini, *Media Freedom* (Wiley 2021).

²¹⁷ Judit Bayer and Sergio Carrera, 'A Comparative Analysis of Media Freedom and Pluralism in the EU Member States' (2016) Study for the LIBE Committee [www.europarl.europa.eu/RegData/etudes/STUD/2016/571376/IPOL_STU\(2016\)571376_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571376/IPOL_STU(2016)571376_EN.pdf) accessed 20 November 2021; Peter Barron and Simon Morrison, 'Pluralism after Acarcity: The Benefits of Digital Technologies' LSE Media Policy Project blog (18 November 2014) <http://blogs.lse.ac.uk/mediapolicyproject/2014/11/18/pluralism-after-scarcity-the-benefits-of-digital-technologies/> accessed 20 November 2021.

²¹⁸ Kari Karppinen, 'The Limits of Empirical Indicators: Media Pluralism As an Essentially Contested Concept' in Peggy Valcke and others (eds.), *Media Pluralism and Diversity: Concepts, Risks and Global Trends* 287 (Springer 2015).

Therefore, the point is about complementing the negative and active sides of freedom of expression with a positive and passive approach. In other words, rather than focusing on protecting users from public interferences (i.e. negative side) and allowing them to freely share ideas and opinion (i.e. active side), the question is about the role of public actors in providing users with tools to check and complain against private interferences (i.e. positive approach) and ensure information quality and diversity (i.e. passive approach). As examined by the next subsections, the two approaches are strictly interconnected. The positive and passive approaches to freedom of expression encourage public actors to regulate content moderation by injecting safeguards strengthening exposure and diversity.

5.6.1 The Positive Side of Freedom of Expression

Once again, European constitutional law owns the instruments to reach this aim. Serious threats for fundamental rights can be considered the triggers of the positive obligation of states to regulate private activities to protect fundamental rights as underlined by the Strasbourg Court,²¹⁹ also in relation to the right to be informed.²²⁰ As the Council of Europe underlined, '[a]s the ultimate guarantors of pluralism, States have a positive obligation to put in place an appropriate legislative and policy framework to that end. This implies adopting appropriate measures to ensure sufficient variety in the overall range of media types, bearing in mind differences in terms of their purposes, functions and geographical reach'.²²¹ As the former UN special rapporteur on freedom of expression observed regarding the use of artificial intelligence technologies, 'human rights law imposes on States both negative obligations to refrain from implementing measures that interfere with the exercise

²¹⁹ See, e.g., *Von Hannover v. Germany* (2005) 40 EHRR 1; *Verein gegen Tierfabriken Schweiz (VgT) v. Switzerland* (2001) 34 EHRR 159. See Lech Garlicki, 'Relations between Private Actors and the European Convention on Human Rights' in Sajó and Uitz (n. 181), 129.

²²⁰ See, e.g., *Österreichische Vereinigung zur Erhaltung, Stärkung und Schaffung v. Austria* (2013); *Youth Initiative for Human Rights v. Serbia* (2013); *Társaság a Szabadságjogokért v. Hungary* (2009); *Sdruženi Jihočeské Matky v. the Czech Republic* (2006); *Bladet Tromsø and Stensaas v. Norway* (1999).

²²¹ Recommendation CM/Rec(2018)1 of the Committee of Ministers to Member States on media pluralism and transparency of media ownership (7 March 2018).

of freedom of opinion and expression and positive obligations to promote the rights to freedom of opinion and expression and to protect their exercise'.²²²

The Strasbourg Court has not only underlined the democratic role of the media,²²³ or the prohibition for states to interfere with freedom of expression. It went even further by recognising that Article 10 can lead to positive obligations.²²⁴ For instance, in *Dink v. Turkey*,²²⁵ the court addressed a case concerning the protection of journalists' expressions clarifying that states have a positive obligation 'to create ... a favourable environment for participation in public debate by all the persons concerned enabling them to express their opinions and ideas without fear, even if they run counter to those defended by the official authorities or by a significant part of public opinion, or even irritating or shocking to the latter'.²²⁶ More recently, in *Khadija Ismayilova v. Azerbaijan*,²²⁷ the Strasbourg Court recognised that states are responsible for protecting investigative journalists. Besides, the protection of the right to freedom of expression under the Convention safeguards not only the right to inform but also the right to receive information.²²⁸ The Strasbourg Court has further clarified the characteristics of such a positive obligation in *Appleby and Others v. UK*, precisely considering the nature of expression at stake and its role for public debates.²²⁹

With regard to the digital environment, the Strasbourg Court recognised the role of the Internet in 'enhancing the public's access to news and facilitating the dissemination of information in general',²³⁰ underlining also that 'the internet has now become one of the principal means by which individuals exercise their right to freedom of expression and information, providing as it does essential tools for participation in activities and discussions concerning political issues

²²² Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2018) <https://undocs.org/A/73/348> accessed 21 November 2021.

²²³ See, e.g., *Barthold v. Germany* (1985) 7 EHRR 383; *Lingens v. Austria* (1986) 8 EHRR 407.

²²⁴ See, e.g., *Fuentes Bobo v. Spain* (2001) 31 EHRR 50; *Özgür Gündem v. Turkey* (2001) 31 EHRR 49.

²²⁵ *Dink v. Turkey* (2010).

²²⁶ *Ibid.*, 137.

²²⁷ *Khadija Ismayilova v. Azerbaijan* (2019).

²²⁸ *Sunday Times v. the United Kingdom (No. 1)* (1979) 2 EHRR 245, 66.

²²⁹ *Appleby and Others v. UK* (2003).

²³⁰ *Cengiz and Others v. Turkey* (2015), 49, 52.

and issues of general interest'.²³¹ Nonetheless, the court just addressed the problem of accessing information without scrutinising the criteria according to which information should be organised. Even if there are different views about how the introduction of artificial intelligence technologies in content moderation affects the right to receive information,²³² users still cannot access information about this process not only to understand the source and reliability of content they access but also remedy harms coming from the block of accounts or the removal of content.

In the European framework, positive obligations in the field of content moderation would also derive from the need to ensure users the right to access remedies against the violations of their fundamental rights. According to Article 13 ECHR, 'everyone whose rights and freedoms as set forth in this Convention are violated shall have an effective remedy before a national authority notwithstanding that the violation has been committed by persons acting in an official capacity', along with the requirements of Article 1 on the obligation to respect human rights and Article 46 on the execution of judgments of the Strasbourg Court. This provision requires contracting parties not just to protect the rights enshrined in the Convention but especially avoid that the protection of these rights is frustrated by lack of domestic remedies. As observed by the Strasbourg Court, 'where an individual has an arguable claim to be the victim of a violation of the rights set forth in the Convention, he should have a remedy before a national authority in order both to have his claim decided and, if appropriate, to obtain redress'.²³³ Similarly, Article 47 of the Charter provides even broader protection of this right being recognised by a general principle of EU law.²³⁴

Moving from the Convention to the Charter, it is worth recalling that Article 11 does not only protect the negative dimension of freedom of expression, but also the positive dimension of media pluralism when it states that '[t]he freedom and pluralism of the media shall be respected'.²³⁵ To achieve this purpose, Member States are required to

²³¹ *Ahmet Yıldırım v. Turkey* (2012), 54.

²³² Sarah Eskens and others, 'Challenged by News Personalisation: Five Perspectives on the Right to Receive Information' (2017) 9(2) *Journal of Media Law* 259.

²³³ *Leander v. Sweden* (1987), 77.

²³⁴ Case 222/84 *Johnston v. Chief Constable of the Royal Ulster Constabulary* (1986) ECR 1651; Case 222/86 *Union nationale des entraîneurs et cadres techniques professionnels du football (Unectef) v. Georges Heylens and others* (1987) ECR 4097; Case C-97/91 *Oleificio Borelli SpA v. Commission of the European Communities* (1992) ECR I-6313.

²³⁵ Charter (n. 40), Art. 11(2).

ensure not only that interferences with the right to freedom of expression are avoided (i.e. negative dimension), but also that diverse and plural access to content is guaranteed (i.e. positive dimension). In *Sky Österreich*,²³⁶ the ECJ dealt with a case involving the protection of media pluralism relating to the financial conditions under which the provider is entitled to gain access to the satellite signal to make short news reports. In this case, the ECJ underlined the protection of the right to be informed or receive information guaranteed by Article 11 of the Charter as a limit to the freedom to conduct a business. In this case, by balancing the two fundamental rights in question, the ECJ gave priority to public access to information over contractual freedom. Nonetheless, once more, this case deals with access and not quality. It is also not clear whether the EU framework could be influenced by the positive obligations of the Convention. It is true that the Charter provides a bridge between the two systems by stating that ‘the meaning and scope of [Charter’s] rights shall be the same as those laid down by the said Convention’.²³⁷

Despite different interpretations, as observed by Kuczerawy, ‘the duty to protect the right to freedom of expression involves an obligation for governments to promote this right and to provide for an environment where it can be effectively exercised without being unduly curtailed’.²³⁸ In the field of algorithmic technologies, the Council of Europe has underlined the importance of ensuring different safeguards such as contestability and effective remedies in relation to public and private actors.²³⁹ Precisely, states should ensure ‘equal, accessible, affordable, independent and effective judicial and non-judicial procedures that guarantee an impartial review, in compliance with Articles 6, 13 and 14 of the Convention, of all claims of violations of Convention rights through the use of algorithmic systems, whether stemming from public or private sector actors’.²⁴⁰

²³⁶ Case C-283/11 *Sky Österreich GmbH v. Österreichischer Rundfunk* (2013).

²³⁷ Charter (n. 40), Art. 52(3).

²³⁸ Aleksandra Kuczerawy, ‘The Power of Positive Thinking. Intermediary Liability and the Effective Enjoyment of the Right to Freedom of Expression’ (2017) 3 *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 182, 186–7.

²³⁹ Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems (8 April 2020) www.statewatch.org/media/documents/news/2020/apr/coe-recommendation-algorithms-automation-human-rights-4-20.pdf accessed 21 November 2021.

²⁴⁰ *Ibid.*

Therefore, the potential regulation of content moderation would not just result from the need to balance other constitutional interests. Injecting democratic safeguards in the process of content moderation would aim to enhance the effective protection of the right to freedom of expression rather than undermining it. Besides, it is not only the right to freedom of expression but also the freedom to conduct business which is limited by the prohibition of abuse of rights.²⁴¹ In other words, the positive obligations of public actors should lead to limit platform powers to define the protection of freedom of expression online, thus balancing constitutional rights and freedoms.

5.6.2 The Passive Side of Freedom of Expression

The logics of moderation limit the transparency and accountability of online platforms, thus marginalising users from understanding how content is processed in the digital environment. Since users cannot generally rely on horizontal and general rights vis-à-vis online platforms, this situation leaves these actors free to decide how to balance and enforce fundamental rights online without any public guarantee. Since the liberal approach to free speech (i.e. the free marketplace of ideas) has led to collateral effects in the digital environment, the protection of the negative side of this freedom is not enough to protect constitutional rights any longer. Therefore, in order to reduce the power of online platforms moderating content on a global scale, it is worth proposing a positive dimension of freedom of expression, triggering a new regulatory intervention towards the adoption of substantive rights and procedural safeguards. This approach contributes to filling the gap of safeguards in content moderation.

At first glance, addressing this issue could lead to changing the liability system of online platforms to increase their degree of responsibility in online content moderation. Nevertheless, this kind of regulatory approach could undermine the economic freedoms of online platforms, which would be overwhelmed by disproportionate obligations. Moreover, changing the safe harbour system would not solve the issue of transparency and accountability in online content moderation. Increasing legal pressure on online platforms by introducing monitoring obligations would result in 'overly aggressive, unaccountable self-policing, leading to arbitrary and

²⁴¹ Convention (n. 41), Art. 17; Charter (n. 40), Art. 54.

unnecessary restrictions on online behavior'.²⁴² This risk, known as collateral censorship, could have strong effects on democracy, thus requiring regulators to avoid threatening online platforms for failing to correctly police content.²⁴³ Due to the ability to govern their digital spaces through content moderation, governments find themselves stuck in cooperating with online platforms.

Apart from the risks of surveillance, even the best-equipped public body would be overwhelmed when handling all the content that platforms moderate. It is true that, in a perfect world, decisions about rights and freedoms should be covered by safeguards and guaranteed by independent public bodies. Nonetheless, reality shows that the fight against illegal content would be hard without online platforms. This consideration does not mean that constitutional democracies should renounce protecting constitutional values but that they should recognise the limits of public enforcement in the digital environment. Therefore, as underlined in Chapter 7, the match is not between private and public enforcement but it is about how to put together the two systems by injecting democratic safeguards in the relationship between public and private actors.

The aim of this new positive and passive approach is not to make platforms liable for their conducts, but responsible for protecting democratic values through more transparent and user-driven procedures. A solution could consist of regulating diversity.²⁴⁴ Some algorithms can be designed to increase diversity and operate adversarial to profiling. In other words, algorithms could also be a support to ensure pluralism and fight the process of targeting based on users' interaction and network (e.g. echo chambers), thus reaching serendipity.²⁴⁵ The European Commission's Code of Practice on Disinformation has encouraged platforms to conduct a process of dilution to tackle

²⁴² Milton Mueller, 'Hyper-Transparency and Social Control: Social Media as Magnets for Regulation' (2016) 39(9) *Telecommunications Policy* 804, 809.

²⁴³ Jack M. Balkin, 'Free Speech and Hostile Environments' (1999) 99 *Columbia Law Review* 2295.

²⁴⁴ Maria Luisa Stasi, 'Ensuring Pluralism in Social Media Markets: Some Suggestions' (2020) EUI Working Paper RSCAS 2020/05 https://cadmus.eui.eu/bitstream/handle/1814/65902/RSCAS_2020_05.pdf?sequence=1&isAllowed=y accessed 20 November 2021.

²⁴⁵ Judith Möller and other, 'Do not Blame it on the Algorithm: An Empirical Assessment of Multiple Recommender Systems and their Impact on Content Diversity' (2018) 21(7) *Information, Communication & Society* 959.

disinformation by improving the findability of trustworthy content. This solution would be a way to frame the role of algorithms not only as a risk but also as a support for democratic values.²⁴⁶ In other words, such a new positive framework of freedom of expression would address the process of moderation without regulating content or changing platform immunities.

Therefore, the issue to solve is not just that relating to the liability of online intermediaries but also that concerning the injection of transparency obligations and procedural safeguards.²⁴⁷ Here, the proposal for a positive framework of freedom of expression is focused on the proceduralisation of content moderation which would not affect platform immunity. The Council of Europe stressed the relevance of the positive obligation to ensure the protection of rights and freedoms through the horizontal effect of human rights and the introduction of regulatory measures. In this case, 'due process guarantees are indispensable, and access to effective remedies should be facilitated vis-à-vis both States and intermediaries with respect to the services in question'.²⁴⁸

Without regulating online content moderation, it is not possible to expect that platforms will turn their business interests driven by profit maximisation to a constitutional oriented approach. New substantive rights and procedural rules would provide users with a set of remedies against the potential violation of their fundamental rights resulting from discretionary decisions by platforms concerning online content while providing proportionate obligations in the field of content moderation.

Besides, this positive approach to freedom of expression could also advantage online platforms. A harmonised regulatory framework of content moderation would reduce the costs of compliance while

²⁴⁶ Brigit Stark and others, 'Are Algorithms a Threat to Democracy? The Rise of Intermediaries: A Challenge for Public Discourse' Algorithm Watch (26 May 2020) <https://algorithmwatch.org/wp-content/uploads/2020/05/Governing-Platforms-communications-study-Stark-May-2020-AlgorithmWatch.pdf> accessed 22 November 2021.

²⁴⁷ Aleksandra Kuczerawy, 'Safeguards for Freedom of Expression in the Era of Online Gatekeeping' (2018) 3 *Auteurs & Media* 292; Kate Crawford and Jason Schultz, 'Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms' (2014) 55 *Boston College Law Review* 93; Danielle K. Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89 *Washington Law Review* 1.

²⁴⁸ Recommendation CM/Rec(2018)2 of the Committee of Ministers to Member States on the roles and responsibilities of internet intermediaries (2018).

enhancing legal certainty and their freedom to conduct business. The liability regime established by the e-Commerce Directive could be replaced by a uniform system of rules and safeguards to increase harmonisation in the internal market. It should not be forgotten that the market is not made only of large online platforms able to comply with any regulation. Therefore, the regulation of content moderation should provide a layered scope of application which takes into consideration small and medium-sized businesses. Otherwise, the risk is to create a legal barrier in the market, fostering the power of some online platforms over the others. A new set of rules on procedural transparency and accountability would reduce the challenges raised by regulatory fragmentation and legal uncertainty which platforms face when moderating content. Even the complementary introduction of a 'Good Samaritan' clause could increase legal certainty by breaking the distinction between active and passive providers and encourage platforms to take voluntary measures. Nonetheless, the solution of European digital constitutionalism would lead to increasing transparency and accountability in the process of content moderation, and maintaining the exception of liability of online platforms.

Therefore, the regulation of online content moderation should be based on four general principles: ban of general monitoring obligation; transparency and accountability; proportionality; availability of human intervention. Precisely, according to the first principle, Member States should not oblige platforms to generally moderate online content. This ban is crucial to safeguard fundamental rights such as freedom to conduct business, privacy, data protection and, last but not least, freedom of expression.²⁴⁹ Secondly, content moderation rules should be assessed and explained to users *ex ante* in a transparent and user-friendly way and *ex post* when content is removed or blocked. In this case human rights impact assessment and transparent notice including the guidelines and criteria used by online platforms to moderate content can ensure that risks for fundamental rights are mitigated and decisions are as predictable as possible. The third principle aims to strike a fair balance between the rights of users and the obligations of platforms. Although the lack of transparent and accountable procedures relegates users in a position of *subjectionis*, the enforcement of users'

²⁴⁹ See, for example, Case C-70/10 *Scarlet Extended SA v. Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM)* (2011) ECR I-11959; Case C-360/10 *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v. Netlog NV* (2012).

rights should not nonetheless lead to a disproportionate limitation of the right and freedom of online platforms to perform their business, especially if we want to protect new or small platforms. The fourth principle is based on introducing the principle of human-in-the-loop in content moderation. The role of humans in this process could be an additional safeguard allowing users to rely on a human translation of the procedure subject to specific conditions.

5.6.3 The Digital Services Act

The adoption of the Digital Services Act constitutes a primary step towards the normative framework supported by the rise of European digital constitutionalism. The Digital Services Act is just a piece of a broader European strategy reviewing the objectives of the Digital Single Market to shape the European digital future.²⁵⁰ As examined in Chapter 7, the proposal for a regulation on artificial intelligence technologies is another example of this European strategy which aims to face the challenges of the algorithmic society.²⁵¹

The adoption of the Digital Services Act can be considered a milestone of the European constitutional strategy. In order to face the challenges raised by platform power, together with the Digital Markets Act,²⁵² the Digital Services Act plays a critical role in providing a supranational and horizontal regime to mitigate the challenges raised by the power of online platforms in content moderation. This legal package promises to update a regulatory framework that dates to the e-Commerce Directive by providing a comprehensive approach to increase transparency and accountability of online platforms in content moderation. Also, the Digital Services Act takes into account the different sizes of online providers by establishing that its scope extends to micro or small enterprises pursuant to the annex to Recommendation 2003/361/EC.²⁵³ Besides, it will provide a horizontal framework for a series of other

²⁵⁰ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Shaping Europe's digital future. COM(2020) 67 final.

²⁵¹ Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts COM(2021) 206 final.

²⁵² Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) COM(2020)842 final.

²⁵³ Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises C(2003) 1422.

measures adopted in recent years which are instead defined as *lex specialis* like the Copyright Directive or the AVMS Directive.²⁵⁴

The title of the proposal reveals how the Digital Services Act will affect the regulatory framework envisaged by the e-Commerce Directive. While maintaining the exemption of liability for online platforms,²⁵⁵ and the ban for Member States to impose general monitoring obligations,²⁵⁶ the Digital Services Act overcomes the issue of neutrality by adopting a Good Samaritan clause. This approach contributes to overcoming the legal uncertainty relating to the definition of passive providers. Online platforms will be free to take 'voluntary own initiative investigations or other activities aimed at detecting, identifying and removing, or disabling of access to, illegal content' without fearing to be sanctioned for failing to comply with their exemption of liability.²⁵⁷ Nonetheless, the Digital Services Act is different from the Communications Decency Act since it limits platform power by providing substantial obligations and procedural safeguards which require platforms to disclose information, assess the risk for fundamental rights and provide redress mechanisms. Additionally, it also maintains (and clarifies) the role of courts or administrative authorities, by requiring an intermediary service provider to terminate or prevent a specific infringement by proceduralising the process to follow for orders to act against illicit content,²⁵⁸ or provide information.²⁵⁹

Even if the proposal maintains the rules of exemption of liability for online intermediaries and extends their freedom to take voluntary measures, it introduces some (constitutional) adjustment which aims to increase the level of transparency and accountability of online platforms. Since the first Recitals, the Digital Services Act complements the goals of the internal market with a constitutional-oriented approach. In particular, it clarifies that providers of intermediary services shall behave responsibly and diligently to allow Union citizens and other persons to exercise their fundamental rights guaranteed in the Charter of Fundamental Rights of the European Union, in particular

²⁵⁴ Caroline Cauffman and Catalina Goanta, 'A New Order: The Digital Services Act and Consumer Protection' (2021) *European Journal of Risk Regulation* 1.

²⁵⁵ Digital Services Act (n. 175), Arts. 3–5.

²⁵⁶ *Ibid.*, Art. 7.

²⁵⁷ *Ibid.*, Art. 6.

²⁵⁸ *Ibid.*, Art. 8.

²⁵⁹ *Ibid.*, Art. 9.

the freedom of expression and information and the freedom to conduct a business, and the right to non-discrimination.²⁶⁰

To achieve this purpose, the Digital Services Act introduces transparency requirements and provides users with the possibility to access redress mechanisms.²⁶¹ In other words, without regulating content, it requires online platforms to comply with procedural safeguards, thus making the process of content moderation more transparent and accountable. These obligations lead online platforms to consolidating their bureaucracy of online content which designs the administrativisation of content moderation. The procedural rules on the notice-and-takedown or on reasons about content removal are just two primary examples of how the Union is trying to require online platforms to be more transparent and accountable.

This approach however has not been enough since the Digital Services Act provides additional obligations which only apply to those platforms falling within the notion of 'very large online platforms'.²⁶² In this case, the proposal sets a higher standard of due diligence, transparency and accountability. These platforms are required to develop appropriate tools and resources to mitigate the systemic risks associated with their activities. And to make this system more effective, the Digital Services Act introduces sanctions applying to all the intermediaries up to 6 per cent of turnover on a global scale in the previous year.²⁶³

This framework underlines how the Commission aims to provide a new legal framework for digital services that is capable of strengthening the Digital Single Market while protecting the rights and values of the Union which are increasingly challenged by the governance of online platforms in the information society. This approach should not surprise since it perfectly fits within the path of European digital constitutionalism whose roots, based on human dignity, do not tolerate the exercise of private power threatening fundamental rights and democratic values while escaping public oversight.

The Digital Services Act shows the resilience of the European constitutional model reacting to the threats of private powers to freedom of expression. Even if some of these rules could be improved during the process of adoption, it is possible to underline that this proposal

²⁶⁰ *Ibid.*, Recital 3.

²⁶¹ *Ibid.*, Arts. 12–24.

²⁶² *Ibid.*, Arts. 25–33.

²⁶³ *Ibid.*, Art. 42.

provides a horizontal regulatory framework to limit platform power in the field of content, thus showing the positive and passive side of European freedom of expression. This new phase should not be seen merely as a turn towards regulatory intervention or as an imperialist extension of European constitutional values. It is more a normative reaction of European digital constitutionalism promoting the positive and passive side of freedom of expression to address the challenges of the algorithmic society.

5.7 Expressions and Personal Data

The relevance of European constitutional law in the field of content moderation should be unveiled at this time. While constitutional provisions have been conceived as limits to the coercive power of the state, in the algorithmic society an equally important and pernicious threat for freedom of expression comes from online platforms making decisions on expression based on their ethical, economic and self-regulatory framework. This situation leads European constitutional law to react to protect constitutional rights and liberties, thus designing a strategy in the long run. This approach does not mean that public actors' interferences with the right to freedom of expression are not relevant any longer, but that it is necessary to look at the limitations to the exercise of freedoms as the result of platform power.

The current opacity of content moderation constitutes a challenge for democratic societies. If individuals cannot understand the reasons behind decisions involving their rights, primarily when automated decision-making systems are involved, the pillars of autonomy, transparency and accountability on which democracy is based are destined to fall. While, in the past, the liberal approach to free speech fitted with the purpose to safeguard democratic values in the digital environment, today, the emergence of new powers governing the flow of information may require a shift from a negative dimension to a positive approach by regulating content moderation. The liberal approach transplanted in the Union from the western side of the Atlantic in the aftermath of the Internet has led online platforms to impose their authoritative regime on content based on a mix of technological and contractual instruments. The result of this situation has led users in a status of *subjectionis* where they find themselves forced to comply with standards of freedom of expression autonomously determined by online platforms.

Within this framework, the Union has started to focus on introducing mechanisms of transparency and accountability in online content moderation. For example, the rights to obtain motivation or human intervention are still unripe but important steps towards a more democratic digital environment. These user rights should not be considered only as instruments to improve transparency and accountability but also as tools to limit the discretion of online platforms operating as private powers outside any constitutional boundary. Nevertheless, it is necessary to observe that Union efforts are still not enough to ensure a path towards the democratisation of the digital environment. The multiple legal regimes regulating online content moderation are increasingly intertwined. This approach could also affect the platform freedom to conduct business since it requires these actors to set different regimes of content moderation.

Nonetheless, the approach of the Union underlines the talent of European digital constitutionalism to react against new forms of powers undermining democratic values. As in the field of data, as examined in Chapter 6, the Union has started to pave the way towards the regulation of platform powers, thus leading to an increasing convergence of safeguards in the field of data and content. In other words, the Union's approach can be considered a first crucial step towards a new approach to content moderation where online platforms are required to operate as responsible actors in light of their gatekeeping role in the digital environment.

Still, the challenges to freedom of expression are not isolated. They are intimately intertwined with the protection of privacy and personal data. Content and data are the two sides of the same coin of digital capitalism. For example, this relationship is evident in content moderation where the content shared by users is also processed as data to provide tailored advertising services. More generally, the challenge concerns the intimate relationship between algorithmic technologies and the processing of (personal) data. Therefore, it is time to focus on the field of data to underline the role of European digital constitutionalism in protecting fundamental rights and democracy.