THE ONLINE HARMS WHITE PAPER: COMPARING THE UK AND GERMAN APPROACHES TO REGULATION[1]

A. INTRODUCTION

The internet has revolutionised our ability to communicate and connect across historic social, political and geographic divides. Where previously gatekeepers mitigated and negotiated access to mass media platforms, today potentially anyone – and any content – can reach millions of users in an instant. This development bears great opportunities for the democratisation of expression and the diversification of public discourse but has likewise broadened the impact of harm caused online. While hateful speech in an offline environment may attract police intervention, arrest and criminal sanctions, posting similar content on social media often appears less likely to result in comparable repercussions.[2] Enforcement of criminal law online is comparatively rare due to the significant investigatory challenges that relative online anonymity presents, limited policing resources and underreporting of potential offenses.[3] In addition, companies have historically been reluctant to vigorously and consistently enforce their own terms and conditions.

This raises the question how platforms and services can be regulated effectively to combat online harms without jeopardising free and open discourse. The paper explores the Online Harms White Paper published by the UK Government earlier this year and compares its

[1] The author would like to thank Oliver Butler, Kate Jones, Harriet Moynihan, Catherine O'Regan, and Jacob Rowbottom for helpful discussions and comments. The following paper is partially based on our joint response to the public consultation on the Online Harms White Paper, available here: *Response to the public consultation on the Online Harms White Paper* (Bonavero Report No 3/2019, 2019) <https://www.law.ox.ac.uk/sites/files/oxlaw/bonavero_response_online_harms_white_paper_-_3-2019.pdf> accessed 10 July 2019.

[2] 'Racist chants at Nottingham Trent University: Two men released' *BBC News* (9 March 2018) <http://www.bbc.co.uk/news/uk-england-nottinghamshire-43342058> accessed 9 March 2018; 'YouTube attacked over Neo-Nazi National Action video' *BBC News* (7 March 2018) <http://www.bbc.co.uk/news/technology-43319975> accessed 9 March 2018; 'My online stalker: The feeling he could hurt me never went away' *BBC News* (9 March 2018) <http://www.bbc.co.uk/news/uk-england-hereford-worcester-43291038> accessed 9 March 2018; Helen Warrell and Madhumita Murgia, 'Social media groups accused of terror fight failings' *Financial Times* (25 August 2016) <https://www.ft.com/content/5d608f40-6a16-11e6-ae5b-a7cc5dd5a28c> accessed 20 January 2019.

[3] The UK Government for instance conceded in 2016 that it was not in a position to provide statistics and figures for online hate crime, see *Action Against Hate - The UK Government's plan for tackling hate crime* (Home Office, 2016), p. 13 at [16].

regulatory approach with the infamous German Network Enforcement Law (*Netzwerkdurchsetzungsgesetz* - NetzDG).[4]

The NetzDG has attracted considerable media attention since fully entering into force on 1 January 2018. The law sparked some controversy in part due to a number of high profile deletions of content posted by German politicians.[5] The law offers the opportunity to study the practical effects of online regulation and should be considered carefully as UK legislation formulates its own regulatory approach. Regardless of what shape the legislation and regulation ultimately takes, human rights law should be at the forefront of considerations. It provides the appropriate framework to balance the myriad of competing interests in online spaces. In that spirit, the paper offers a preliminary appraisal of both the White Paper and the NetzDG based on the available data.

B. SCOPE

The White Paper proposes to cover an extremely wide range of companies offering a broad spectrum of platforms and services. The White Paper proposes to cover 'hosting, sharing and discovery of user-generated content' and 'facilitation of public and private online interaction between service users.'[6] This broad scope is equalled only by the breadth of online harms assigned to the regulator, ranging from terrorist content, hate crimes and intimidation to disinformation, violent content and excessive screen time.[7] At least with some of these harms, for instance screen time, the evidence basis for harmfulness appears doubtful.[8] Jointly, the broad scope of companies and online harms covered risks setting any regulator an impossible task

[4] *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz)*, Act to Improve Enforcement of the Law in Social Networks of 1 September 2017 (BGBl. I 3352); an English translation can be found on the website of the Federal Ministry of Justice: 'Act to Improve Enforcement of the Law in Social Networks' (*Bundesministerium der Justiz und für Verbraucherschutz*, 12 Juli 2017) <http://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf;jsessionid=E23D49344654C18787B25C74A2E3D9C9.1_cid334?__blob=publicationFile&v=2> accessed 1 July 2018.

[5] Andrea Diener, 'Storch-Satire ist nicht regelkonform' *Frankfurter Allgemeine Zeitung* (3 January 2018) <http://www.faz.net/aktuell/feuilleton/medien/twitter-sperrt-titanic-magazin-wegen-storch-satire-15371919.html> accessed 1 July 2018; 'Maas-Tweet über Thilo Sarrazin gelöscht' *Spiegel Online* (8 January 2018) <http://www.spiegel.de/netzwelt/netzpolitik/netzdg-heiko-maas-tweet-ueber-thilo-sarrazin-verschwunden-a-1186747.html> accessed 1 July 2018.

[6] *Online Harms White Paper* (HM Government, 2019), p.49.

[7] Ibid, p.31

[8] Amy Orben, Tobias Dienlin and Andrew K. Przybylski, 'Social media's enduring effect on adolescent life satisfaction' (2019) 116 Proceedings of the National Academy of Sciences 10226.

that is likely to lead either to excessively vague standards or highly selective oversight and enforcement actions: both raise serious legitimacy and rule of law concerns.

I. Services and platforms

If a regulator is only capable of providing oversight and enforcement for a few high-profile companies, whether for practical purposes or due to a lack of resources, then this risks undermining its broader mission and legitimacy. The danger is that medium and lower profile companies, while technically covered by the regulatory regime, might in practice not be monitored or only infrequently subject to enforcement actions. However, even with a narrowly tailored regulatory objective, companies require more details on the codes of practice and oversight and enforcement measures they can reasonably expect from the regulator. It is not difficult to imagine that smaller companies and newcomers could be pushed out of the market if the compliance burden is too great, especially if user-generated content is not core to their business model. The potential for harm is plainly greater on a popular social media platform than on a message board run by a local canoe club.

For good reasons, the Network Enforcement Law in Germany draws sharper lines as to which services and platforms are regulated. The NetzDG applies only to 'social media networks' with at least two million users within Germany. Social media networks are defined as internet platforms that seek to profit from providing users with the opportunity to share content with other users and the broader public. Platforms which provide individualised communication services, such as email or messaging apps, as well as platforms providing editorialised content, such as news websites, are excluded from the scope of the law (§ 1 (1) NetzDG).

While it is reasonable to seek a broader regulation of platforms and services than envisioned by NetzDG, it is nonetheless crucial that a regulator is equipped to deliver its regulatory mission. An overburdened and under-resourced regulator that resorts to either loose standards that most companies already meet or to highly selective enforcement actions driven by public attention would have little impact on online harms across the board and damage its legitimacy and the rule of law in the process. Starting smaller with a narrower scope is one viable technique to address these concerns, but at the very least, legislation should consider exempting certain companies partially or entirely from regulation. Legislation should carefully consider whether a broad scope for regulation as proposed by the White Paper is realistically achievable and desirable based on existing human rights obligations under the Human Rights Act 1998.

II. Online harms

In terms of content regulated, NetzDG is also significantly more modest than the White Paper. The White paper covers a broad range of online harms that cover certain illegal conduct, such as distributing terrorist content, harassment and hate crimes, as well as legal conduct that is deemed harmful, for instance disinformation, trolling, intimidation and excessive screen time.[9]

By contrast, under NetzDG content is designated illegal only if it falls under one of the enumerated provisions of the German criminal code.[10] Collectively, the enumerated provisions simultaneously cover more and less than some of the online harms mentioned in the White Paper. For instance, in Germany one might attract criminal liability for defamation when describing a specific abortion doctor's work as 'babycaust' even though this does not fall under most definitions of hate speech: it is not based on attributes such as race, religion, ethnic origin, sexual orientation, disability, or gender.[11] Conversely, election posters by a far right party depicting ethnically stereotyped people on a flying carpet with the caption 'Have a good flight home' did not attract criminal liability even though it at least arguably constitutes hate speech on the grounds of race, religion and ethnic origin.[12]

The idea of the White Paper to extend regulation beyond content which is prohibited under criminal law is not wrong in principle. However, there is a danger that the harms captured are overly broad and vague, which makes regulation vulnerable to abuses like the targeting of unpopular as opposed to harmful speech. In time, the notion of an online harm may well be defined more strongly by reference to media attention and political opportunism than evidence and principle. Many of the harms mentioned in the White Paper are difficult to define clearly and at times it is debatable whether there is sufficient evidence to consider them harms at all, thus removing a core justification for regulation under human rights law.

Instead, legislation should consider a two-tier approach to regulation which differentiates between: (a) harms with a strong evidence basis and a reasonably clear definition ('definite harms') and (b) harms with a weaker evidence basis or with a less clear and contingent

[9] *Online Harms White Paper*, p.31.
[10] Criminal Code in the version promulgated on 13 November 1998 (BGBl. I 3322), last amended by Article 1 of the Law of 24 September 2013 (BGBl. I 3671) and Article 6(18) of the Law of 10 October 2013 (BGBl. I 3799)
[11] Federal Constitutional Court, 1 BvR 49/00, 24 May 2006, BVerfGK 8, 89; the conviction for defamation was found compatible with Article 10 ECHR, see *Hoffer and Annen v Germany*, 397/07; 2322/07, Court, 13 January 2011 at [49].
[12] Higher Regional Court Munich, 5St RR (II) 9/10, 9 February 2010, NJW 2010, 2150.

definition ('contextual harms'), for instance harms relating to online screen time and disinformation, respectively. To an extent this idea is reflected in the White Paper's suggestion that there will be a stricter code of practice for terrorist and child abuse content than other online harms.[13] However, a two-tier approach would take this idea further by insisting on a prescriptive regulatory approach for definite harms, and an expressly more flexible model for contextual harms. Crucially, there are often a multitude of plausible interventions that might address contextual harms.

The regulator may therefore wish to focus its attention on ensuring consistent enforcement of existing legal obligations as well as terms and conditions of companies. The latter often already make provision for many of the contextual harms raised in the White Paper: the core problem more often lies with consistent, fair, effective and efficient enforcement. As part of a flexible model for contextual harms, the regulator could set companies certain targets based on human rights norms that they must accomplish and require the publication of details on their complaints management system and procedures. Such targets could be user focused, for instance requiring a social media company to provide account protections and support for individuals facing targeted harassment, or company focussed, requiring procedures that action reported content within a given timeframe. As we shall see below when discussing obligations and sanctions, this model has already produced some interesting developments in Germany under NetzDG.

C. OBLIGATIONS

The White Paper broadly seeks to make online companies more responsible and accountable to users, especially given the asymmetric nature of their contractual relationship. However, as Jacob Rowbottom has pointed out in the joint response to the Online Harms White Paper, it is not clear whether this goal is captured well by the terminology of a 'duty of care'.[14] The envisioned framework would not grant any individual an action or remedy in the event that a duty has been breached: this would be the conventional understanding of a duty of care in negligence law. Instead, the White Paper proposes codes of practice, which companies are expected to follow and any penalties for systemic failures would be imposed by the regulator.

[13] *Online Harms White Paper*, pp.11-13.
[14] Ibid, pp.3-4.

Such an approach appears more closely related to a conventional model of statutory regulation, such as the one employed by NetzDG.

The core obligations under NetzDG centre on establishing an effective and transparent complaints management infrastructure (§ 3 NetzDG) and compiling bi-annual reports on its activity (§ 2 NetzDG). Especially the reporting obligations are quite detailed and include provisions that set out reviewer training and oversight requirements. The complaints management infrastructure must chiefly ensure that the social networks delete or block 'illegal content' within a specified timeframe. Deletion results in a global removal of the content from the platform, while blocking merely makes it unavailable in Germany.[15] In this context, it is important to note that the obligation to delete is not novel. NetzDG merely enforces compliance with existing legal obligation under § 10 of the Telemedia Act (TMG), which was introduced to implement the e-Commerce Directive in Germany.[16] Under that provision, social media platforms are liable under private and criminal law once they are made aware of illegal content and if they fail to delete it without undue delay.

With respect to the deadlines, NetzDG distinguishes between 'manifestly illegal' and 'illegal' content, prescribing different deadlines for deletion. Manifestly illegal content must be deleted within 24 hours of a receiving a complaint, while merely illegal content must be actioned within seven days. Manifestly illegal content is not defined, but legislative commentary suggests that only the most obvious of cases would be covered: when in doubt, the seven-day deadline is applicable.[17] Separately from the duty to delete content platforms must preserve evidence of criminal conduct for up to ten weeks (§ 3 (2) NetzDG) for law enforcement purposes.

The most important exception to the seven-day deadline is triggered if operators refer the decision on deletion to a body of industry self-regulation (§ 3 (3) (b) NetzDG). This procedure removes the final decision from affected companies, thus mitigating concerns over private censorship, and is arguably the preferred regulatory approach of NetzDG. Social media platforms may choose to collectively setup and fund such bodies, provided they are independent and issue binding decisions. Certain conditions of operation apply (§ 3 (6) NetzDG), including that such bodies must be accredited by the Federal Ministry of Justice.

---

[15] Although blocking and deleting are thus distinct, they will be collectively referred to as 'deleting' throughout this paper as they have identical requirements.

[16] Telemedia Act of 26 February 2007 (BGGl. I 179), last amended by Article 1 of the Act of 28 September 2017 (BGBl. I 3530);

[17] *Entwurf eines Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz – NetzDG)* (Drucksache 18/12356, 2017) <http://dipbt.bundestag.de/dip21/btd/18/123/1812356.pdf#page=18> accessed 1 July 2018.

Industry self-regulation has not been favoured by social media companies like Facebook and twitter, but is otherwise a common feature in the German media regulatory landscape. It has been setup amongst others by the movie, tv, and video games industries to rate the age appropriateness of content and these bodies wield considerable power affecting sales and marketing rules under youth protection laws.[18] For instance, the self-regulatory body of the video games industry (USK) only last year overturned a long standing ban on Nazi symbolism, despite well-established exceptions under German criminal law for artistic and scientific use of such symbols.[19]

D. SANCTIONS

Broadly speaking the White Paper has suggested a number of possible sanctions and enforcement powers for the codes of practice. These powers may range from publication of reports and compliance rankings, to administrative fines and on towards more serious disruptions of the business activities for persistently non-compliant companies. Powers to block access to websites entirely come with great risks and should therefore only be imposed as a last resort. There are many evasion mechanisms (for instance, virtual private networks, mirrors of banned websites, darknet websites) that are accessible for users with basic technological competence. Furthermore, blocking websites may cause significant collateral harm to users engaged in legitimate use and will make it difficult to justify such a sanction as proportionate under human rights law. Turkey was found to have violated Article 10 of the European Convention on Human Rights (ECHR) when it blocked access to all sites of a popular search engine provider as well as to a popular video sharing platform because content available through their services allegedly insulted the memory of Mustafa Kemal Atatürk.[20]

Under NetzDG the primary sanction mechanism are fines. Occasional, non-systematic mistakes by social media platforms in an otherwise lawful complaints management infrastructure would not typically attract fines under NetzDG. Contrary to the impression given by some reports, no

[18] 'Freiwillige Selbstkontrolle der Filmwirtschaft - FSK' (*Freiwillige Selbstkontrolle der Filmwirtschaft GmbH*) <https://www.spio-fsk.de/> accessed 1 July 2018 [self-regulation body of the film industry]; 'Unterhaltungssoftware Selbstkontrolle - USK' (*Freiwillige Selbstkontrolle Unterhaltungssoftware GmbH*) <http://www.usk.de/> accessed 1 July 2018 [self-regulation body of the video games industry].

[19] 'USK berücksichtigt bei Altersfreigabe von Spielen künftig Sozialadäquanz' (Berlin, 9 August 2018) <http://www.usk.de/service/presse/details-zum-presseartikel/article/usk-beruecksichtigt-bei-altersfreigabe-von-spielen-kuenftig-sozialadaequanz/> accessed 1 August 2018.

[20] *Yildirim v Turkey*, 3111/10, Court, 18 December 2012; *Cengiz and others v Turkey*, 48226/10; 14027/11, Court, 1 December 2015.

fines can attach to decisions in individual cases. Instead, fines require a systemic and persistent failure in the complaints management infrastructure which must be substantiated through content that has been ruled illegal by a court in a separate proceeding (§ 4 (5) NetzDG). For instance, a failure to name an agent or lack of response from a named agent may attract a fine of up to 500.000 Euros, while other failures to implement a NetzDG compliant management scheme may result in fines of up to 5 million Euros. The latter increases to 50 million Euros for legal persons and corporations under § 30 (2) of the Act on Regulatory Offences, pursuant to § 4 (2) NetzDG.[21] Facebook has recently become the first social media platform to be fined 2 million Euros by the regulator for shortcomings in its reporting obligations.[22] The regulator highlighted incomplete and missing data in the bi-annual transparency reports, and the difficulty for users to access the NetzDG specific complaint mechanism, as well as deficits in reporting on management and training of content moderators. Crucially, however, the fine was not issued due to a failure to action specific illegal content and Facebook can still appeal the fine.

E. OVERBLOCKING

Critics of online regulation often focus on the potential for unintended consequences for free expression.[23] They allege that regulation will promote an overly aggressive deletion policy that targets content that is both legal under relevant laws and permissible under community standards (this is referred to as 'overblocking'). Overblocking could arise due to the structure of the fines, which notably do not apply for systemic failures that lead to the deletion of too much legal content. A prudent social media platform would, so goes the argument, when confronted with a high volume of reports, delete content that is questionable, rather than risk a fine. Indeed, if overblocking is a prevalent phenomenon beyond the occasional erroneous decision, for instance due to the structure of the fines under NetzDG, it risks producing a 'chilling effect' on freedom of expression. This seems a clear risk for those users whose content

[21] Act on Regulatory Offences of 19 February 1987 (BGGl. I 602), last amended by Article 4 of the Act of 13 May 2015 (BGBl. I 706).

[22] *Federal Office of Justice Issues Fine against Facebook* (2019) <https://www.bundesjustizamt.de/DE/Presse/Archiv/2019/20190702_EN.html;jsessionid=AE3371DA35E5EEFEEB5A8BFF7F52AFCA.1_cid392?nn=3449818> accessed 20 July 2019.

[23] *Response to Consultation on the Online Harms White Paper* (Open Rights Group, 2019) <https://www.openrightsgroup.org/assets/files/reports/report_pdfs/Online_Harms_Consultation_Response.pdf> accessed 20 July 2019, p.7; Mathias Hong, 'Das NetzDG und die Vermutung für die Freiheit der Rede' (*Verfassungsblog*, 9 January 2018) <https://verfassungsblog.de/das-netzdg-und-die-vermutung-fuer-die-freiheit-der-rede/> accessed 1 July 2018; Diana Lee, 'Germany's NetzDG and the Threat to Online Free Speech' (*Case Disclosed - Media Freedom and Information Access Clinic Blog*, 10 October 2017) <https://law.yale.edu/mfia/case-disclosed/germanys-netzdg-and-threat-online-free-speech> accessed 1 July 2018.

is removed and whose accounts are suspended, but also for those indirectly affected, as it creates an atmosphere where fear of sanctions reduces individual willingness to exercise free expression. Overall, this may contribute to stifling of lawful expression and public debates.[24] However, there are good reasons to be sceptical that overblocking is an inevitable consequence of online regulation, and this scepticism is supported by available data on NetzDG.

For one, it is not clear that a chilling effect is inevitable. Most obligations under the law do not directly relate to deletion at all: filling bi-annual reports, naming an agent who can receive complaints and training of staff members who evaluate reported content are among many examples. Social media platforms also have the ability to delegate their responsibilities to a recognised body of industry self-regulation. The law does not require the extensive investment into platform exclusive content moderators currently taking place at some companies. In fact, NetzDG specifically permits (and arguably encourages) platforms to pool their resources by providing more generous deadlines for industry self-regulation. As such, it is regrettable that no major social media platform appears to be currently considering this option, but this can be partially explained through the history of content moderation policies.

Generally speaking, social media platforms have an unfortunate track record of failing to recognise and delete hate speech even where it unequivocally violates their terms and conditions, at times thereby enabling violence and lynching.[25] The risk of criminal and civil litigation is comparatively rare and social media platforms apparently judged the significant costs of compliance as unnecessary expenses. This is evidenced by Facebook and Twitter hiring significantly more content moderators following the passing of NetzDG, and reports now indicate that German speakers account for up to a sixth of the global content moderation team.[26] It goes without saying that this vastly over represents German speakers in the global customer base of social media companies.

---

[24] For a more in-depth look at the constitutionality of NetzDG under the German constitution, see Stefan Theil, 'The German NetzDG: A Risk Worth Taking?' (*Verfassungsblog*, 8 February 2018) <https://verfassungsblog.de/the-german-netzdg-a-risk-worth-taking/> accessed 6 May 2019.

[25] Matt Rynolds, 'Facebook could block far-right hate speech, so why isn't it?' *Wired* (20 December 2017); Michael Safi, 'Sri Lanka accuses Facebook over hate speech after deadly riots ' *Guardian* (14 March 2018) <https://www.theguardian.com/world/2018/mar/14/facebook-accused-by-sri-lanka-of-failing-to-control-hate-speech> accessed 1 August 2018; Shaikh Azizur Rahman, ''Fake news often goes viral': WhatsApp ads warn India after mob lynchings ' *Guardian* (13 July 2018) <https://www.theguardian.com/world/2018/jul/13/fake-news-whatsapp-ads-india-mob-lynchings> accessed 1 August 2018.

[26] 'Tough new German law puts tech firms and free speech in spotlight' *Guardian* (5 January 2018) <https://www.theguardian.com/world/2018/jan/05/tough-new-german-law-puts-tech-firms-and-free-speech-in-spotlight> accessed 2 February 2018.

Finally, data published in the bi-annual reports required under NetzDG do not indicate widespread overblocking. Instead, they reveal that both Facebook and Twitter have elected not to delete the vast majority of content reported under a separately established NetzDG mechanism. Facebook received 886 reports under this mechanism and chose to delete content based on 218 reports between 1 January and 30 June 2018, or roughly 25 percent of cases.[27] This notably excludes reports under the Facebook community standards, which are provided separately. In the same timeframe twitter received 264,818 reports which resulted in deletion in 28,645 cases, or roughly 11 percent of cases.[28] It should be noted that Twitter does not provide separate data on NetzDG based reports, and instead released aggregate data for all complaints. These figures have only marginally shifted in the second annual reports.[29] Facebook shows a total of 500 NetzDG reports, 159 of these led to the deletion of content, which equates to roughly 33.8 percent of cases, while Twitter shows a total of 256,462 reports of which 23,165 led to the deletion of content, roughly 9 percent of cases. The trends are broadly similar: the vast majority of reports do not result in deletion of content. This is problematic for critics of online regulation as a cornerstone of their argument relies on demonstrating that overblocking is more than a theoretical possibility.

F. CONCLUSION

Online regulation, when crafted with care, can address online harms and may have civilizing influence on online expression instead of foreshadowing the end of a free and open discourse.

In that vein, the broad scope of companies and online harms targeted by the UK Online Harms White Paper is particularly concerning. It carries the risk that the regulator will be overburdened and resort to excessively vague standards or to highly selective oversight and enforcement actions: both would undermine the legitimacy of the regulator and adversely impact the rule of law. The idea mooted in the White Paper to extend regulation beyond content which is

---

[27] *Facebook NetzDG-Transparenzbericht Juli 2018* (Facebook, 2018) <https://fbnewsroomus.files.wordpress.com/2018/07/facebook_netzdg_juli_2018_deutsch-1.pdf> accessed 10 July 2019, p.7.

[28] *Twitter Netzwerkdurchsetzungsgesetzbericht: Januar - Juni 2018* (Twitter, 2018) <https://cdn.cms-twdigitalassets.com/content/dam/transparency-twitter/data/download-netzdg-report/netzdg-jan-jun-2018.pdf> accessed 10 July 2019, p.7.

[29] *Facebook NetzDG-Transparenzbericht Januar 2019* (Facebook, 2019) <https://fbnewsroomde.files.wordpress.com/2019/01/facebook_netzdg_januar_2019_deutsch52.pdf> accessed 10 July 2019; *Twitter Netzwerkdurchsetzungsgesetzbericht: Juli - December 2018* (Twitter, 2018) <https://cdn.cms-twdigitalassets.com/content/dam/transparency-twitter/data/download-netzdg-report/current-report.pdf> accessed 10 July 2019.

prohibited by the criminal law is not wrong in principle. However, there is a danger of being over inclusive and that what counts as online harms worthy of regulation is defined chiefly by reference to media attention and political opportunism. Legislation might therefore consider a two-tier approach to regulation and offer exemptions for certain companies. In any case, human rights law should be at the forefront of considerations because it provides the appropriate framework to balance the myriad of competing interests in online spaces.

NetzDG presents a good case study that should be closely considered in the legislative process. The viability of its more limited regulatory approach depends considerably on its practical effects: should NetzDG produce systematic and significant overblocking that leads to a chilling effect, then it will be difficult to justify under human rights norms. The available data from the two bi-annual reports certainly does not provide much evidence of overblocking. Nonetheless, close scrutiny of any online regulation under human rights law remains essential as there are sound historical, legal and normative arguments that caution against heavy handed interference with freedom of expression.