Pandemics and Platforms: Private Governance of (Dis)Information in Crisis Situations

Matthias C. Kettemann and Marie-Therese Sekwenz

in: Matthias C. Kettemann (ed.), How Platforms Respond to Human Rights Conflicts Online. Best Practices in Weighing Rights and Obligations in Hybrid Online Orders (Hamburg: Verlag Hans-Bredow-Institut, 2022)

EU COST Action – CA19143 – Global Digital Human Rights 2022

Pandemics and Platforms: Private Governance of (Dis)Information in Crisis Situations

Matthias C. Kettemann and Marie-Therese Sekwenz[1]
UNIVERSITY OF INNSBRUCK, LEIBNIZ INSTITUTE FOR MEDIA RESEARCH | HANS-BREDOW-INSTITUT

Introduction

What role do online platforms play in managing and governing information during the pandemic? Chinese platforms cooperated substantially with the governments' message (and message control) on COVID-19, but also US-based platforms like Twitter and Facebook that had employed a hands-off approach to certain types of disinformation in the past invested considerably in the tools necessary to govern online disinformation more actively. Facebook, for instance, deleted Facebook events for anti-lockdown demonstrations while Twitter had to rely heavily on automated filtering (with human content governance employees back at home). This contribution will assess these practices, their impact and permanence in light of the author's research on the important role of intermediaries as normative actors, including their establishment, through terms of service and content governance practices, of a private order of public communication.

State responsibilities and private duties regarding online communication

Online just as offline, states have an obligation to respect, protect and ensure human rights for everyone on their territory or under their control.[2] This extends the duties states have from the analog world into the digital one, especially as being 'online' is now the new normal and the internet of platforms and contents is enriched by an internet of things (like smart cars) and an internet of bodies (like intelligent wearables). Even as new approaches to norm entrepreneurship online emerge,[3] rights that people have offline are still their rights in online environments.

Online just as offline, states have a primary responsibility and ultimate obligation to protect human rights and fundamental freedoms.[4] But what are these requirements international law imposes on states to ensure rights online? A key international legal basis for freedom of expression is Article 19 of the Universal Declaration of Human Rights, which is largely considered to reflect customary law. In addition, in 1976 the International Covenant on Civil and Political Rights (ICCPR) was adopted, which in its Article 19

[1] This contribution was first published in Matthias C. Kettemann and Konrad Lachmayer (eds.), Pandemocracy in Europe. Power, Parliaments and People in Times of Covid-19 (London: Hart, 2021).

[2] This section draws on Kettemann/Benedek, Freedom of expression online, in Mart Susi (Hrsg.), Human Rights, Digital Society and the Law. A Research Companion (London: Routledge, 2019), 58-74 and Benedek/Kettemann, Freedom of Expression on the Internet (Strasbourg: Council of Europe, 2014, 2nd ed. 2020).

[3] Radu/Kettemann/Meyer/Shahin, 'Normfare: Norm entrepreneurship in internet governance', Telecommunications Policy, Volume 45, Issue 6, 2021, https://doi.org/10.1016/j.telpol.2021.102148.

[4] Just see European Court of Human Rights, Beizaras and Levickas v. Lithuania, (Application no. 41288/15), 15 January 2020.

reiterates the text of the Universal Declaration and then clarifies (in para. 2) that everyone "shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice."

Accordingly, the right goes beyond the freedom of the press and the freedom of the media to include individual expression in the widest sense. However, the right, with the exemption of the freedom of opinion, is not absolute or without limits. Under certain clearly defined conditions it can be restricted. In its biannual resolution on human rights on the internet in 2012, 2014 and 2016, the Human Rights Council affirmed, with references to Articles 19 of the UDHR and the ICCPR, the special role of freedom of expression online: "the same rights that people have offline must also be protected online, *in particular freedom of expression*, which is applicable regardless of frontiers and through any media of one's choice [...]."[5]

An evaluation of freedom of expression standards in international law from a European perspective (must) also consider similar regional standards such as the protections of Article 10 (1) of the European Convention on Human Rights (ECHR), enshrining "the right to freedom of expression. This right shall include the freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers." Note the reference to the non-interference "by public authority": States are obliged to protect freedom of expression both as a free-standing right and as an essential "enabler" of other rights through the internet. As former UN Special Rapporteur for Freedom of Expression, Frank La Rue, wrote, "by acting as a catalyst for individuals to exercise their right to freedom of opinion and expression, the internet also facilitates the realisation of a range of other human rights".[6]

The ECtHR case of K.U. v. Finland[7] confirms that states have an obligation, under the European Convention of Human Rights, to ensure that the human rights of persons under their jurisdiction are protected – offline just as online. If social network service providers fail to introduce safeguards (in the case of K.U. v. Finland, to protect the privacy rights of a child), states need to enforce a legal protection framework.[8] Just as real as the primary responsibility of states, however, is the observation that a lot of the discourse relevant for the constant opinion-forming work of democratic modernity takes place in private spaces.

The key questions regarding how to enable, moderate and regulate speech today therefore have to be asked and answered with a view to digital and private spaces.

The vast majority of communicative spaces on the internet are privately held and owned.[9] This is due to the powerful role of intermediaries, companies that enable our online activity.[10] States are therefore not the

[5] Human Rights Council Resolution 32/13, The promotion, protection and enjoyment of human rights on the Internet, UN Doc. A/HRC/RES/32/13 of 18 July 2016, para. 1 (emphasis added).

[6] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN Doc. A/HRC/17/27 of 16 May 2011, paras. 22 and 23. But the internet also brings about new challenges to these same human rights.

[7] ECtHR, K.U. v. Finland (2 December 2008), Application No. 2872/02.

[8] See Benedek/Kettemann, Freedom of Expression on the Internet (Strasbourg: Council of Europe, 2014, 2nd ed. 2020), pp. 92, 110.

[9] On why would need public social media, too, see Lukas B. Wieser Social Media im demokratischen Verfassungsstaat – Warum wir öffentlich-rechtliche soziale Medien brauchen. In M. Becker, M. Hofer, E. Paar, & C. Romirer (Hrsg.), Gesellschaftliche Herausforderungen – Öffentlich-rechtliche Möglichkeiten (S. 239-288). Wien: Verlag Jan Sramek.

[10] Cf. Kettemann/Schulz, Setting Rules for 2.7 Billion. A (First) Look into Facebook's Norm-Making System: Results of a Pilot Study (Hamburg: Working Papers of the Hans-Bredow-Institut, Works in Progress # 1, January 2020), https://leibniz-hbi.de/uploads/media/Publikationen/cms/media/5pz9hwo_AP_WiP001InsideFacebook.pdf.

only actors in ensuring human rights online. As the 2018 Recommendation of the Council of Europe on internet intermediaries notes, a "wide, diverse and rapidly evolving range of players, commonly referred to as "internet intermediaries", facilitate interactions on the internet between natural and legal persons by offering and performing a variety of functions and services. Some connect users to the internet, enable the processing of information and data, or host web-based services, including for user-generated content. Others aggregate information and enable searches; they give access to, host and index content and services designed and/or operated by third parties."[11]

Network effects and mergers have led to the domination of the market by a relatively small number of key intermediaries. As the 2018 Recommendation warned, these few companies have growing power: "[the] power of such intermediaries as protagonists of online expression makes it imperative to clarify their role and impact on human rights as well as their corresponding duties and responsibilities, including as regards the risk of misuse by criminals of the intermediary's services and infrastructure."[12]

Internet intermediaries have duties under international and national law. In line with the UN Guiding Principles on Business and Human Rights and the "Protect, Respect and Remedy" Framework, intermediaries should respect the human rights of their users and affected parties in all their actions. This includes the responsibility to act in compliance with applicable laws and regulatory frameworks. Internet intermediaries also develop their own rules, usually in form of terms of service or community standards that often contain content-restriction policies. This responsibility to respect within their activities all internationally recognized human rights, in line with the United Nations Guiding Principles on Business and Human Rights, exists independently of the states' ability or willingness to fulfil their own human rights obligations.[13]

States have also misused intermediaries in the past to introduce filters and enforce laws that violate international human rights commitments. Therefore, as the Recommendation notes, any norms applicable to internet intermediaries, regardless of their objective or scope of application, "should effectively safeguard human rights and fundamental freedoms, as enshrined in the European Convention on Human Rights, and should maintain adequate guarantees against arbitrary application in practice."[14]

Due to the multi-layered nature of the regulatory framework governing services provided by or through intermediaries, their regulation is challenging. As they operate in many countries and data streams, especially for cloud-based services, and often cross many countries and jurisdictions, different and conflicting laws may apply.[15] This is exacerbated by, as the 2018 Council of Europe recommendation identified, "the global nature of the internet networks and services, by the diversity of intermediaries, by the volume of internet communication, and by the speed at which it is produced and processed."[16]

In line with the UN Guiding Principles on Business and Human Rights and the 'Protect, Respect and Remedy' Framework ('Ruggie Principles'), a convincing approach posits that intermediaries need to behave

[11] Council of Europe, Recommendation CM/Rec (2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries, preambular para. 4.

[12] Ibid., preambular para. 7.

[13] Ibid., para. 2.1.1.

[14] Ibid., para. 2.1.2.

[15] Ibid., preambular para. 6.

[16] Ibid., preambular para. 9.

in a certain way to keep their 'social licence' to operate the quasi-public sphere. Such a 'licence' necessitates commitments to human rights of their users and affected parties in all their actions (including the formulation and application of terms of service) in order to address and remedy negative human rights impacts directly. For example, in order to identify and prevent adverse human rights impacts, business enterprises need to carry out human rights-due diligence. This should involve meaningful consultation with potentially affected groups and other relevant stakeholders, taking appropriate action, monitoring the effectiveness of the response and communicating their action as part of their accountability obligations.[17]

There is substantial literature on the duties of private entities in international law, especially with regard to the duties of transnational corporations[18] and private military contractors.[19] Much of it is applicable to internet standard-setters, but also to internet content companies, such as search engine providers and social networking services.[20]

Platforms in Pandemic Times

In a study[21] and subsequent analysis[22] of platform behaviour during the year of the rising Covid-19 pandemic 2020, we have identified a number of key shared commonalities among more than 40 states. Dominant platforms have been able to defend, or even solidify, their position, but communicative practices on those platforms are changing. State authorities increasingly use platforms to communicate and inform, and platforms support these approaches willingly. In the following, we look specifically at selected platforms and study their reaction to (dis)information related to Corona to assess whether we can see an emergence of a cross-platform commitment to counter Corona-related disinformation.

Facebook

During the pandemic Facebook continued to remain one of the leading platforms. With its two point seven billion daily users on its main platform alone.[23] With data traffic for messaging services, video and voice

[17] See Ruggie J. (7 April 2008), Human Rights Council, Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises, Protect, respect and remedy: a framework for business and human rights, UN Doc. A/HRC/8/5 and Guiding principles on business and human rights, implementing the United Nations "Protect, respect and remedy" framework, Annex to the Final Report of the Special Representative to the Human Rights Council, UN Doc. A/HRC/17/31 and adopted by the Human Rights Council (16 June 2011) by Resolution 17/4, Guidelines 17-21. See Benedek/Kettemann (2020), 85f.

[18] Especially after the adoption of the UN Guiding Principles on Business and Human Rights. See Radu Mares (ed.), The UN Guiding Principles on Business and Human Rights. Foundations and Implementation (Leiden: Nijhoff, 2011); and, for a comprehensive analysis, Wesley Cragg (ed.), Business and Human Rights (Cheltenham: Edward Elgar, 2012). For the international trade dimension relevant for aspects of ICTs, see Alistair M. Macleod, Human rights and international trade: normative underpinnings, in ibid., 179-196.

[19] Cf. Lindsey Cameron, Vincent Chetail, Privatizing War. Private Military and Security Companies under Public International Law (Cambridge: CUP, 2013), 288-382 (arguing that PMSCs can be bound both as companies and as the sum of their individual employees.). See also the body of scholarship cited in ibid., 269, note 22.

[20] Council of Europe, Committee of Ministers (4 April 2012), Recommendation CM/Rec(2012)3 on the protection of human rights with regard to search engines and Recommendation CM/Rec(2012)4 on the protection of human rights with regard to social networking services.

[21] Kettemann/Fertmann, 'Viral Information: How States and Platforms Deal with Covid-19-related Disinformation: an Exploratory Study of 18 Countries' (Hamburg: Verlag Hans-Bredow-Institut, 2021), GDHRNet Working Paper #1, 126.

[22] Kettemann et al., Healthy Conversations? Selected Trends in Covid-19-Related (Dis)Information Governance on Platforms, in: Kettemann/Fertmann (eds.), Viral Information: How States and Platforms Deal with Covid-19-related Disinformation: an Exploratory Study of 18 Countries (Hamburg: Verlag Hans-Bredow-Institut, 2021), GDHRNet Working Paper #1.

[23] John Clement, 'Facebook MAU Worldwide 2020' (Statista, 2020) <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/> accessed 3 December 2020.

calls throughout the time of the pandemic was an important space for online speech during the pandemic.[24] Before the pandemic, Facebook claimed not wanting to be an "arbiter of truth".[25] While this was never accurate, and Facebook has always influenced how online communication takes place on this platform, the reaction to COVID-19 was much stronger than any other single issue addressed by automated and human content moderation.

According to the report by 'Avaaz' Facebook projected three point eight billion pieces of content that were classified as misleading health content to its users[26]. While the amount of content on the platform has increased, its content moderation was more difficult during the pandemic.[27] Because of global lockdown constraints, Facebook had to rely even more on automated content moderation[28.] Facebook also changed the community standards and defined content related to anti-vaccine statements[29], or advertising claims for medical face masks, hand sanitizer, disinfectant wipes and COVID-19- test kits, as forbidden by its terms of service which also can be seen as a shift in the company's approach.[30]

In March 2020, Facebook introduced an 'Information Hub'[31] for most users to provide health information by trusted authorities like the 'Center for Disease Control and Prevention' or the 'World Health Organization' matched with content from hand-picked journalists, politicians or other selected content about the pandemic. Facebook makes also use of pop-ups as a user-interface-design decision to additionally remind users to wear facemasks or to provide further information about the pandemic. Another information-related action was the investment of 100 million dollars to support fact-checking and journalism on the Corona crisis.[32] The financial support by Facebook also included donations for relief efforts[33,] healthcare workers[34], small businesses[35] or supporting health crisis helplines.[36]

[24] Kiran Khan and others, 'The COVID-19 Infodemic: A Quantitative Analysis Through Facebook' (2020) 12 11.

[25] Tom McCarthy, 'Zuckerberg Says Facebook Won't Be "arbiters of Truth" after Trump Threat' The Guardian (28 May 2020) <https://www.theguardian.com/technology/2020/may/28/zuckerberg-facebook-police-online-speech-trump> accessed 3 December 2020.

[26] AVAAZ, 'Facebook's Algorithm: A Major Threat to Public Health' <https://secure.avaaz.org/campaign/en/facebook_threat_health/> accessed 3 December 2020.

[27] Facebook, 'Community Standards Enforcement Report, November 2020' (About Facebook, 19 November 2020) <https://about.fb.com/news/2020/11/community-standards-enforcement-report-nov-2020/> accessed 11 December 2020.

[28] 'Keeping People Safe and Informed About the Coronavirus - About Facebook' <https://about.fb.com/news/2020/10/coronavirus/> accessed 24 November 2020.

[29] Jin Kang-Xing, 'Supporting Public Health Experts' Vaccine Efforts' (About Facebook, 19 October 2020) <https://about.fb.com/news/2020/10/supporting-public-health-experts-vaccine-efforts/> accessed 3 December 2020.

[30] Facebook, 'Information about Ads about Social Issues, Elections or Politics and COVID-19' (Facebook Business Help Center, 2020) <https://www.facebook.com/business/help/213593616543953> accessed 15 January 2021 and Facebook, 'Banning Ads and Commerce Listings for Medical Face Masks' (6 March 2020) <https://about.fb.com/news/2020/12/coronavirus/>.

[31] Salvador Rodriguez, 'Facebook Is Encouraging Everybody to Take Social Distancing Seriously' CNBC (18 March 2020) <https://www.cnbc.com/2020/03/18/coronavirus-facebook-launches-information-center-at-top-of-news-feed.html> accessed 3 December 2020.

[32] Facebook, 'Investing $100 Million in the News Industry' (30 March 2020) <https://about.fb.com/news/2020/12/coronavirus/>.

[33] Facebook, 'Matching $20 Million in Donations to Support COVID-19 Relief Efforts' (13 March 2020) <https://about.fb.com/news/2020/12/coronavirus/>.

[34] Facebook, 'Donating $25 Million to Support Healthcare Workers' (30 March 2020) <https://about.fb.com/news/2020/12/coronavirus/>.

[35] Facebook, 'Investing $100 Million in Small Businesses' (17 March 2020) <https://about.fb.com/news/2020/12/coronavirus/>.

[36] Facebook, 'Connecting People to Well-Being Tips and Resources' (9 April 2020) <https://about.fb.com/news/2020/12/coronavirus/>.

According to Kahn et al. 22,3 per cent of their investigated Facebook posts contained misinformation about COVID-19.[37] Facebook furthermore opened some data silos to the public and researchers[38] as part of the 'Data for Good' program.[39] To increase the use of this data Facebook had to further adapt its terms of service to the situation.[40] This data-support includes a COVID-19 map and dashboard with data about global symptom surveys, as well as information about datasets that mirror the movement range, or other mobility-related information of Facebooks users. This data can be used for research that e.g., takes a close look at the friendship-boundaries of Facebook users in two countries to predict the likelihood of the creation of coronavirus hotspots.[41]

Facebook had to send home content moderators on the 16th of March 2020.[42] This situation caused by the lockdown led to a high increase in artificial intelligence supported content moderation.[43] While the old moderation system was going through the amount of content chronologically, the use of a variety of algorithms (this includes machine learning approaches, filtering, ranking and sorting) now uses criteria[44] to sort through the content and prioritize it.[45] This change within the moderation system should help remove harmful content quicker than the chronological system did.

Nevertheless, Facebook remained a key platform for the spread of misinformation.[46] This claim is based on the high number of interactions related to the content in question compared to other platforms. A study also highlighted the connection between YouTube and Facebook, which are more strongly correlated through content shares than other platforms. The authors therefore come to the conclusion that misinformation is more likely to become viral if it is shared through Facebook.

Twitter

The company reports a total reach of its monetizable daily active users (mDAU) of 164 million in the first quarter of 2020, which is a growth of 23 per cent in comparison to the corresponding values in 2019.[47]

[37] Khan and others (n 2).

[38] Facebook, 'Data for Good: New Tools to Help Health Researchers Track and Combat COVID-19' (About Facebook, 6 April 2020) <https://about.fb.com/news/2020/04/data-for-good/> accessed 3 December 2020.

[39] Facebook, 'Our Work on COVID-19' (Facebook Data for Good) <https://dataforgood.fb.com/docs/covid19/> accessed 1 December 2020.

[40] Facebook, 'Protecting Privacy in Facebook Mobility Data during the COVID-19 Response' (Facebook Research, 3 June 2020) <https://research.fb.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response/> accessed 3 December 2020.

[41] Theresa Kuchler, Dominic Russel and Johannes Stroebel, 'The Geographic Spread of COVID-19 Correlates with the Structure of Social Networks as Measured by Facebook' [2020] arXiv:2004.03055 [physics, q-bio] 1.

[42] 'Keeping People Safe and Informed About the Coronavirus - About Facebook' (n 5).

[43] James Vincent, 'Facebook Is Now Using AI to Sort Content for Quicker Moderation' The Verge (13 November 2020) <https://www.theverge.com/2020/11/13/21562596/facebook-ai-moderation> accessed 15 January 2021.

[44] The criteria used are: virality, severity and how likely it is for the content to violate the Facebook Community Standards.

[45] Sílvia Majó-Vázquez and others, 'Volume and Patterns of Toxicity in Social Media Conversations during the Covid-19 Pandemic' 12.

[46] Aleksi Knuutila and others, 'Covid-Related Misinformation on Youtube' 7.

[47] Statista, 'Twitter Global MDAU 2020' (Statista) <https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide/> accessed 17 January 2021 and Hans Rosenberg, Shahbaz Syed and Salim Rezaie, 'The Twitter Pandemic: The Critical Role of Twitter in the Dissemination of Medical Information and Misinformation during the COVID-19 Pandemic' (2020) 22 Canadian Journal of Emergency Medicine 418.

While the traffic on the platform has risen in numbers the problems via moderation, misinformation and fake news became even more problematic for COVID-19 related content.[48] Twitter took several measures to overcome the challenges of the pandemic. It supported verified information sources and tried to make them easy to access[49] in order to protect the debate on its platform.[50] Twitter strengthened its organization-relationships and fostered public engagement on its platform.[51] Twitter also focussed on the research aspects as a fourth pillar of handling the pandemic.[52] Furthermore, Twitter decided to focus on the safety of partners and employees.[53] In order to provide valuable information to its users Twitter developed a COVID-19 tab in its 'Explore'[54] function. Here users have easy access to reliable sources and hand-picked page highlights from public health experts. Through the use of verified accounts misleading speech or misinformation should be tackled on the microblogging platform.[55]

Pulido et al. found out that during the pandemic misinformation increased in presence while it is retweeted less likely, compared to scientific content or evidence, which create more engagement within the online environment. [56] The COVID-19 search prompt is another design decision Twitter took in order to curb the spread of misinformation.[57] This search prompt should also correct misspellings within the search function and promote search results from credited sources like the 'World Health Organization' in relation to COVID-19.[58] The second cluster of actions against the pandemic amplified the need of clarifying statements about misleading information and how the company deals with it.[59]

Twitter published its three key questions which are taken into consideration for COVID-19 content removal decisions, an important element of justification governance. First, 'Is the content advancing a claim

[48] Anatoliy Gruzd and Philip Mai, 'Going Viral: How a Single Tweet Spawned a COVID-19 Conspiracy Theory on Twitter' (2020) <https://journals.sagepub.com/doi/full/10.1177/2053951720938405> accessed 24 November 2020 and Rosenberg, Syed and Rezaie (n 40).

[49] Twitter, 'Helping People Find Reliable Information: Staying Safe and Informed on Twitter' (18 May 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

[50] Twitter, 'Protecting the Public Conversation' (14 July 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

[51] Twitter, 'Partnering with Organizations and Public Engagement' (10 April 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

[52] Twitter, 'Empowering Research of COVID-19 on Twitter' (29 April 2020) and Twitter, 'Twitter Developer Labs' (2020) <https://developer.twitter.com/en/products/labs> accessed 17 January 2021<https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

[53] Jennifer Christie, 'Keeping Our Employees and Partners Safe during #coronavirus' (12 May 2020) <https://blog.twitter.com/en_us/topics/company/2020/keeping-our-employees-and-partners-safe-during-coronavirus.html> accessed 17 January 2021.

[54] Twitter, 'Coronavirus: Staying Safe and Informed on COVID-19 Tab in Explore' (18 May 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

[55] Twitter, 'COVID-19 Account Verification' (20 March 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

[56] Cristina M Pulido and others, 'COVID-19 Infodemic: More Retweets for Science-Based Information on Coronavirus than for False Information' (2020) 35 International Sociology 377.

[57] Twitter, 'Global Expansion of the COVID-19 Search Prompt' (4 March 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

[58] World Health Organization, 'World Health Organization (WHO) (@WHO) / Twitter page' (Twitter, 2020) <https://twitter.com/WHO> accessed 17 January 2021.

[59] Twitter, 'Broadening Our Guidance on Unverified Claims' (22 April 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021 and Twitter, 'Clarifying How We Assess Misleading Information' (14 July 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

of fact regarding COVID-19?' Second, 'Is the claim demonstrably false or misleading?' The third question risen by Twitter is: 'Would belief in this information, as presented, lead to harm?'

The first question demands the existence of more than an opinion and rather seeks for content that covers some degree of factual truth. The expression has to have the power to influence the behaviour of other users on the platform in order to fulfil the criteria Twitter has set. The second question analyses the degree of truth of the statement or otherwise it will classify the Tweet as misleading.[60] The Tweet either contains already falsified information[61] or the claim could confuse users through the process of visibility and sharing pattern.[62] The third question tries to minimize the harm that misinformation could cause through its platform. Twitter explicitly names content that could increase the likelihood of exposure to the virus or information that could lead to capacity bottlenecks within the public health system. When a Tweet meets all three of the forementioned questions and criteria Twitter grants itself the right to block or remove the content in question.

On 11 May 2020 Twitter updated its 'Terms of Service' for the placement of warning labels on Tweets that come with a reduced visibility for others.[63] Twitter's ads policy had to be renewed in order to meet the COVID-19 needs on the platform. The update restricted content that could cause panic, and content that could influence prices or the advertising of products that might be short in stock like face masks or hand sanitizers. Twitter also widened its understanding of harm on its platform.[64] Now the term also addresses speech that directly challenges the guidance from authoritative sources that contain public health information.

The first layer of the moderation process of Twitter is automated and Twitter's systems questioned one and a half million accounts that were under suspicion of amplifying COVID-19 discussion through spamming or other manipulative behaviours. Tasks related to judgement of the content itself had to be changed due to the pandemic. Twitter clarified its use of automated systems on the 16th of March 2020.[65] Twitter reported the automated surfacing of the uploaded content on its platform through the help of data trained on previous moderation decisions taken by its human moderation team. While misleading or false claims around COVID-19 often demand for additional context, the human moderation team of Twitter will take review decisions 'by hand'.[66] Twitter also informs its users of longer waiting periods for content

[60] An example given by Twitter includes statements like: „The National Guard just announced that no more shipments of food will be arriving for two months — run to the grocery store ASAP and buy everything" or "5G causes coronavirus — go destroy the cell towers in your neighbourhood!".

[61] This process of falsification is supported by subject-matter experts.

[62] Twitter gives the following examples: „Whether the content of the Tweet, including media, has been significantly altered, manipulated, doctored, or fabricated; Whether claims are presented improperly or out of context; Whether claims shared in a Tweet are widely accepted by experts to be inaccurate or false."

[63] Yoel Roth and Nick Pickles, 'Updating Our Approach to Misleading Information' (11 May 2020) <https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html> accessed 17 January 2021 and "Tweets that are labelled under this expanded guidance will have reduced visibility across the service. Reducing the visibility of Tweets means that we will not amplify the Tweets on a number of surfaces across Twitter. However, anyone following the account will still be able to see the Tweet and Retweet„.

[64] Twitter, 'Broadening Our Definition of "Harm"' (1 April 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

[65] Twitter, 'An Update on Our Content Moderation Work' (23 March 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

[66] Twitter, 'Coronavirus' (n 54).

moderation, while also giving the user a right to appeal.[67] Furthermore, Twitter announced to change its hierarchy of the global 'content severity triage system'. It now prioritizes content that might be classified as a rule violation, because this contravention is attributed as the highest risk by the platform to cause harm to its users.[68] The company also reported to have implemented a daily assurance check of its moderation system.[69] On 3 March 2020, Twitter also reminded its users of the 'zero-tolerance approach' the platform has towards manipulation.[70]

The third category of measures include the Twitter questions and answers that supported public engagement and promoted actions like 'Clapping for our healthcare heroes'[71] or '#AsktheGov'[72] where elected leaders were able to answer questions of Twitter users. Twitter announced a global software solution hackathon to fight the pandemic.[73]The company also donated one million dollars to the 'Committee to Protect Journalists' and the 'International Women's Media Foundation'.

As a further response to the crisis, Twitter tried to keep the public conversation alive while also using valuable information about the pandemic through the user data. In order to do that, Twitter created 'Twitter Developer Labs'[74] to grant access of real-time data to developers and researchers. Open research data is used for projects that take a closer look at trends and COVID-19 related discriminatory conversation.[75] There are other examples for valuable insight through Twitter's data to determine the amount or magnitude of misinformation.[76]

[67] Twitter, 'Appeal an Account Suspension or Locked Account' (Help Center, 2020) <https://help.twitter.com/forms/general> accessed 17 January 2021.

[68] @Vijaya and Matt Derella, 'An Update on Our Continuity Strategy during COVID-19' (16 March 2020) <https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html> accessed 17 January 2021.

[69] ibid.

[70] Twitter, 'Our Zero-Tolerance Approach to Platform Manipulation' (4 March 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

[71] Twitter, 'Clapping for Our Healthcare Heroes' (7 April 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

[72] Twitter, '#AsktheGov & #AsktheMayor Twitter Q&As' (2 April 2020) <https://blog.twitter.com/en_us/topics/company/2020/covid-19.html> accessed 17 January 2021.

[73] DEVPOST, 'COVID-19 Global Hackathon 1.0' (COVID-19 Global Hackathon 1.0, 2020) <https://covid-global-hackathon.devpost.com/> accessed 17 January 2021.

[74] Twitter, 'Twitter Developer Labs' (n 47).

[75] Maria Renee Jimenez-Sotomayor, Carolina Gomez-Moreno and Enrique Soto-Perez-de-Celis, 'Coronavirus, Ageism, and Twitter: An Evaluation of Tweets about Older Adults and COVID-19' (2020) 68 Journal of the American Geriatrics Society 1661.

[76] Matthew D Kearney, Shawn C Chiang and Philip M Massey, 'The Twitter Origins and Evolution of the COVID-19 "Plandemic" Conspiracy Theory' (2020) 1 Harvard Kennedy School Misinformation Review; Ramez Kouzy and others, 'Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter' (2020) 12 Cureus; Anna Kruspe and others, 'Cross-Language Sentiment Analysis of European Twitter Messages during the COVID-19 Pandemic', Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020 (Association for Computational Linguistics 2020) <https://www.aclweb.org/anthology/2020.nlpcovid19-acl.14> accessed 24 November 2020; Richard J Medford and others, 'An "Infodemic": Leveraging High-Volume Twitter Data to Understand Early Public Sentiment for the Coronavirus Disease 2019 Outbreak' (2020) 7 Open Forum Infectious Diseases <https://academic.oup.com/ofid/article/7/7/ofaa258/5865318> accessed 24 November 2020; Akif Mustafa, Subham Mohanta and Shalem Balla, 'Public Reaction to COVID-19 on Twitter: A Thematic Analysis' [2020] EPRA International Journal of Multidisciplinary Research (IJMR) 2455.

[76] Gautam Kishore Shahi, Anne Dirkson and TA Majchrzak, 'An Exploratory Study of COVID-19 Misinformation on Twitter' [2020] ArXiv; Gautam Kishore Shahi, Anne Dirkson and TA Majchrzak, 'An Exploratory Study of COVID-19 Misinformation on Twitter' [2020] ArXiv; Gautam Kishore Shahi, Anne Dirkson and TA Majchrzak, 'An Exploratory Study of COVID-19 Misinformation on Twitter' [2020] ArXiv and Karishma Sharma and others, 'COVID-19 on Social Media: Analyzing Misinformation in Twitter Conversations' [2020] arXiv:2003.12309 [cs].

Kouzy found that 24.8 per cent of tweets contained misinformation, while not only the tweet is of interest but also its author. Kouzy found that the rate of misinformation increased to 33.8 per cent when the author was an informal individual or posted within a group account setting. This finding is also mirrored within the usage of verified accounts. Where 31.0 per cent of the unverified accounts were classified as misinformation, while only 12.6 per cent of verified accounts contained misinformation. The company also focussed on parameters like site reliability in the pandemic due to an increase in service demand.[77] Metrics can provide a valuable insight into numbers and statistics or in this case of sentiment analysis. According to Ordun et al.[78] the information related to COVID-19 was about 50 Minutes faster retweeted compared to other Chinese networks.

Kruspe,[79] Mustafa et al.[80] and Proharel[81] used Twitter data to employee a sentiment analysis of the tweets to find out more about people's moods. But not only ordinary Twitter users are under investigation – Rufai and Bunce analysed tweets from leaders of G7 countries where the majority of tweets were classified as 'informative' content (82.8 per cent) by the researchers while the G7 leaders also used their twitter accounts to boost the moral of their citizens (nine point four per cent) of Tweets. [82]

Twitter reported to have taken into account several measures to support its employee's safety through mandatory[83] work from home whenever possible, while also assuring contractual fulfilment in cases where home office solutions are not possible.[84] In order to smoothen the change in working conditions the company also provided reimbursement toward home office related costs and additional resources for parents in the form of financial help for COVID-19 related additional day-care expenses.

YouTube

YouTube has a current user base of two billion that consumes one billion hours of content daily.[85] YouTube had some prior knowledge and experience for how to deal with pandemics.[86] Misleading information amounts to a fourth of classified COVID-19 related misleading content, which reached up to 62 million users around the globe.[87]

[77] @Vijaya and Derella (n 66).

[78] Catherine Ordun, S Purushotham and Edward Raff, 'Exploratory Analysis of Covid-19 Tweets Using Topic Modeling, UMAP, and DiGraphs' [2020] ArXiv 1.

[79] Kruspe and others (n 76).

[80] Mustafa, Mohanta and Balla (n 78).

[81] Bishwo Prakash Pokharel, 'Twitter Sentiment Analysis During Covid-19 Outbreak in Nepal' (Social Science Research Network 2020) SSRN Scholarly Paper ID 3624719 <https://papers.ssrn.com/abstract=3624719> accessed 24 November 2020.

[82] Sohaib Rufai and Catey Bunce, 'World Leaders' Usage of Twitter in Response to the COVID-19 Pandemic: A Content Analysis' (2020) 42 Journal of public health (Oxford, England) 1.

[83] Twitter reported on the Updated April 1, 2020 to Send home content moderators

[84] For contractors and hourly workers who are not able to perform their responsibilities from home, Twitter will continue to pay their labor costs to cover standard working hours while Twitter's work-from-home guidance and/or travel restrictions related to their assigned office are in effect. March 11, 2020

[85] YouTube, 'YouTube in Numbers' (2020) <https://www.youtube.com/intl/en-GB/about/press/> accessed 14 December 2020.

[86] Kaustubh Bora and others, 'Are Internet Videos Useful Sources of Information during Global Public Health Emergencies? A Case Study of YouTube Videos during the 2015–16 Zika Virus Pandemic' (2018) 112 Pathogens and Global Health 320.

[87] Heidi Li and others, 'YouTube as a Source of Information on COVID-19: A Pandemic of Misinformation?' (2020) 5 BMJ Global Health 1.

YouTube uses a search algorithm coupled with a recommendation system that makes use of 'collaborative filtering' in order to individually sort content according to user preferences.[88] Research in user behaviour sheds light on the importance of the ranking order of YouTube's search results. Gudivada et al. found out, that users usually only consider the top 20 search results for consumption, therefore the algorithmic recommendation of YouTube is responsible for approximately 70 per cent of content consumed by users on their platform.[89] Furthermore, Li et al. claim that during the c-Corona-crisis the content of credited sources on the platform are under-represented compared to other content creators.[90]

YouTube used several measures to curb the spread of Corona-related disinformation its platform. YouTube implemented the following key strategies: authoritative voices, providing helpful information, boosting remote learning, removing misinformation, reducing the spread of borderline content through the creation of a COVID-19 'Medical Misinformation Policy', while also providing infrastructure to its users to stay connected.[91]

With YouTube's efforts for making authoritative voices more visual, the company displayed information panels of health organisations connected to search results related to COVID-19 queries. According to YouTube, this promoted content had around 100 billion views.[92] COVID-19 related content also has high engagement, while content that also is politicized raises on average around 9000 comments for a video and factual content gained 3000 comments on average.[93] Furthermore, the consumption of news (compared to the numbers of the previous year) on the platform soared up to 75 per cent.[94] Marchal et al. found out that four-fifths of channels on YouTube sharing information are professional news agencies.[95] Nevertheless, content containing misinformation reached high volumes of shares on social media platforms and add up to the sum of shares of the five biggest English media and news sites.[96]

The company also increased the visibility of non-profit organisation and governments through free ad inventory. Another change in the user interface is the news shelf for COVID-19 related information to highlight news from authoritative sources and health agencies[97] while also building a fact-checker network that can place warning labels on content that also reduces the visibility of the video.[98]

[88] James Davidson and others, 'The YouTube Video Recommendation System' (2010).

[89] VN Gudivada, D Rao and J Paris, 'Understanding Search-Engine Optimization' (2015) 48 Computer 43.

[90] Li and others (n 91); Nahema Marchal, Hubert Au and Philip N Howard, 'Coronavirus News and Information on YouTube': 5 and Nahema Marchal and Hubert Au, '"Coronavirus EXPLAINED": YouTube, COVID-19, and the Socio-Technical Mediation of Expertise' (2020) 6 Social Media + Society 2056305120948158, 19.

[91] YouTube, 'Youtube Response During Coronavirus - How YouTube Works' (Youtube Response During Coronavirus - How YouTube Works, 2020) <https://www.youtube.com/howyoutubeworks/our-commitments/covid-response/> accessed 12 December 2020.

[92] ibid.

[93] Marchal, Au and Howard (n 95).

[94] Casey Newton, 'How YouTube's Moderators Are Keeping up with Changing Guidance around COVID-19' The Verge (29 April 2020) 19 <https://www.theverge.com/interface/2020/4/29/21239928/youtube-fact-check-neal-mohan-interview-misinformation-covid-19> accessed 12 December 2020.

[95] Marchal, Au and Howard (n 95).

[96] Knuutila and others (n 35).

[97] YouTube, 'Youtube Response During Coronavirus - How YouTube Works' (n 97) and Newton (n 100) 19.

[98] Ibid and Knuutila and others (n 35).

On 13 July 2020, YouTube first launched a feature called 'Depression and Anxiety Information Panels'[99] that uses information and guidelines provided by the 'Centre for Disease Control' (CDC).[100] One of the latest changes to YouTube's information channels on the 17 November now also corners content about misinformation on vaccines for COVID-19.[101] The platform started in 2019 to limit its recommendation for borderline content.[102] Borderline content makes up for around one per cent of the content on YouTube and describes cases that almost meet the criteria of deletion according to the 'Community Guidelines'.[103] Furthermore, YouTube promotes content for fundraising through a specific tag and a donation button.[104]

According to YouTube over almost eight million videos were removed by the platform between July and September this year.[105] The platform now exercises more intensive oversight over, and strives to limit the reach of, content that contains medical misinformation or discredits authoritative health authority's guidance in one of the following categories: treatment,[106] prevention,[107] Corona diagnostics[108] and/or transmission.[109]

YouTube, in contrast to Facebook, monetises COVID-19 related content.[110] This is a change in the platform's monetarisation approach that prohibited the utilisation of sensitive events it followed only month before.[111] On 16 March 2020, the company announced that it will use more automated content

[99] YouTube, 'Health Information Panels' (2020) <https://support.google.com/youtube/answer/9795167> accessed 14 December 2020.

[100] YouTube, 'Update to COVID-19 Information Panels' (11 June 2020) <https://support.google.com/youtube/answer/9777243?hl=en-GB> accessed 17 January 2021.

[101] YouTube, 'Update to COVID-19 Information Panels' (17 November 2020) <https://support.google.com/youtube/answer/9777243?hl=en-GB>.

[102] YouTube, 'Continuing Our Work to Improve Recommendations on YouTube' (blog.youtube, 2019) <https://blog.youtube/news-and-events/continuing-our-work-to-improve/> accessed 14 December 2020.

[103] YouTube, 'YouTube Community Guidelines & Policies - How YouTube Works' (YouTube Community Guidelines & Policies - How YouTube Works, 2020) <https://www.youtube.com/howyoutubeworks/policies/community-guidelines/> accessed 14 December 2020.

[104] Sarah Perez, 'YouTube Launches a Suite of Fundraising Tools' (TechCrunch, 2018) <https://social.techcrunch.com/2018/08/30/youtube-launches-a-suite-of-fundraising-tools/> accessed 14 December 2020 and YouTube, 'Youtube Response During Coronavirus - How YouTube Works' (n 97).

[105] 'YouTube Community Guidelines Enforcement – Google Transparency Report' <https://transparencyreport.google.com/youtube-policy/removals?hl=en> accessed 24 November 2020.

[106] YouTube gives the following examples: „Content that encourages the use of home remedies in place of medical treatment such as consulting a doctor or going to the hospital, Content that encourages the use of prayer or rituals in place of medical treatment, Content that claims that there's a guaranteed cure for COVID-19, Claims about COVID-19 vaccinations that contradict expert consensus from local health authorities or WHO, Content that claims that any currently-available medicine prevents you from getting the coronavirus" or "Other content that discourages people from consulting a medical professional or seeking medical advice".

[107] YouTube gives the following examples: „Claims that there is a guaranteed prevention method for COVID-19, Claims that an approved COVID-19 vaccine will cause death, infertility, or contraction of other infectious diseases, Claims that an approved COVID-19 vaccine will contain substances that are not on the vaccine ingredient list, such as fetal tissue, Claims that an approved COVID-19 vaccine will contain substances or devices meant to track or identify those who've received it, Claims that an approved COVID-19 vaccine will alter a person's genetic makeup, Claims that any vaccine causes contraction of COVID-19, Claims that a specific population will be required (by any entity except for a government) to take part in vaccine trials or receive the vaccine first".

[108] YouTube gives the following example: "Content that promotes diagnostic methods that contradict local health authorities or WHO".

[109] YouTube gives the following examples: „Content that claims that COVID-19 is not caused by a viral infectionContent that claims COVID-19 is not contagious, Content that claims that COVID-19 cannot spread in certain climates or geographies, Content that claims that any group or individual has immunity to the virus or cannot transmit the virus, Content that disputes the efficacy of local health authorities' or WHO's guidance on physical distancing or self-isolation measures to reduce transmission of COVID-19"; see also YouTube, 'COVID-19 Medical Misinformation Policy - YouTube Help' (n 114).

[110] YouTube, 'Monetising COVID-19-Related Content' (2 April 2020) <https://support.google.com/youtube/answer/9777243?hl=en-GB> see also Marchal, Au and Howard (n 95).

[111] Sarah Perez, 'YouTube Warns of Increased Video Removals during COVID-19 Crisis' <https://techcrunch.com/2020/03/16/youtube-warns-of-increased-video-removals-during-covid-19-crisis/> accessed 12 December 2020.

moderation and informed its platform users about the fact that more false positives and false negatives will be visible.[112] According to their enforcement report, YouTube removed 99 per cent of comments[113] through automated filtering.[114] Furthermore, YouTube defines exceptions for removal in cases of educational, documentary, scientific or artistic settings. The platform grants itself the power to remove content that violates a provision of its 'Community Guidelines', where YouTube also informs the uploading-user of the content removal per mail. Users, that violate the company's rules for the first time will only be warned, while YouTube will strike against the user's channel for further violations. When a user has reached three strikes YouTube will delete the channel.[115]

According to Priyanka et al. users are a central player in the creation or sustainment of misinformation. The authors argue that independent user content, which accounts for 11 per cent of total video content, is seven times less likely useful information about COVID-19 compared to academic institution content.[116]

The platform is a popular host for remote learning. YouTube launched 'Learn@Home', an extension to its 'Learning Hub' and is supported by several educational content creators and services like e.g., 'Khan Academy'.[117]

The removal of content on the platform is one way to target misinformation, but here the technological eco-system is more entwined than expected. In the deletion process of a video, YouTube had a longer removing time of several hours that was also viral on Facebook and Twitter.[118] According to Knuutila et al. YouTube needed 41 days to remove misleading videos that gained 149,825 views on average according to their sample.[119] As mentioned in section a.), the authors describe that the audience for misleading content of COVID-19 on YouTube is closely correlated[120] to (and on a large scale caused by) Facebook shares.

This entwined ecosystem was studied by Cinelli et al. for several platforms including YouTube.[121] The authors discovered that users have a specific timing pattern for content consumption. Furthermore, 'mainstream social media' only grants a small fraction of interaction to questionable content.[122] The questionable content on the platform can reach different degrees of visibility. In order to compare the platform's approach, the authors used the coefficient of relative amplification.[123] According to their

[112] Ibid; see also YouTube, 'Actions to Reduce the Need for People to Come into Our Offices' (Google, 16 March 2020) <https://blog.google/inside-google/company-announcements/update-extended-workforce-covid-19/> accessed 10 December 2020.

[113] The total amount of comments removed between July and September this year add up to 1,140,278,887 comments on the platform.

[114] 'YouTube Community Guidelines Enforcement – Google Transparency Report' (n 115).

[115] YouTube, 'Community Guidelines Strike Basics' (2020) <https://support.google.com/youtube/answer/2802032> accessed 14 December 2020.

[116] Priyanka Khatri and others, 'YouTube as Source of Information on 2019 Novel Coronavirus Outbreak: A Cross Sectional Study of English and Mandarin Content' (2020) 35 Travel Medicine and Infectious Disease 1.

[117] Khan Academy, 'Khan Academy' (2020) <https://www.youtube.com/user/khanacademy> accessed 14 December 2020.

[118] Statt (n 53).

[119] Knuutila and others (n 35).

[120] With a positive correlation of 0,7 for the variables „views on YouTube" and „Shares on Facebook" .

[121] The authors investigated: Twitter, YouTube, Gab, Reddit.

[122] M Cinelli and others, 'The COVID-19 Social Media Infodemic' [2020] Scientific reports 10.

[123] The coefficient of amplification is a metric to capture the amplification on a platform for the fraction of average engagement for unreliable posts to reliable posts.

findings, YouTube amplifies unreliable content less compared to reliable content with a ratio of four out of ten.

Telegram

Telegram is a Russian instant messaging service and was founded in 2013 by Pavel Durov.[124] Pavel Durov also founded the Russian social network 'VKontakte' which can be seen as a *pendant* to Facebook.[125] The service has more than 200 million[126] active users. Germany, Austria and Switzerland together account for eight million users on a daily basis.[127] The service's popularity can be explained through the one-to-many messaging option which also provides for the creation of groups reaching up to 200.000 members. Messages send within those groups can only be seen if searched for or appear within the group for every user.[128] A user can stay anonymous while posting to other users. Telegram therefore can create a wide reach for an individual user, while the user's personality can be hidden. Furthermore, the platform, in contrast to Facebook or Twitter, does not use a recommendation system nor an algorithmic timeline.[129]

The service is available within the EU or the United Kingdom for users that are 16 according to the company's terms of service.[130] Telegrams terms of service are very brief. A user has to avoid practices that: 'Use our service to send spam or scam users, promote violence on publicly viewable Telegram channels, bots, etc. or post illegal pornographic content on publicly viewable Telegram channels, bots, etc.'.[131]

Through this open formulation of the online behaviour of users, Telegram grants its online population an ample understanding of free speech. Telegram therefore is an *El Dorado* for extremist groups like the Islamic state[132] or the far right.[133] Nevertheless, Telegram announced cooperation with the EUROPOL to counter terrorist propaganda online.[134] Because of the *laissez-faire* approach the company has towards content moderation and fake news, it poses a serious threat for COVID-19 misinformation.[135]

[124] Anna Baydakova, 'Telegram CEO Donates 10 BTC to Pandemic Relief Effort' (CoinDesk, 28 May 2020) <https://www.coindesk.com/telegram-ceo-donates-10-btc-to-pandemic-relief-effort> accessed 12 December 2020.

[125] Katsiaryna Baran and Wolfgang Stock, 'Facebook Has Been Smacked Down. The Russian Special Way of SNSs: Vkontakte as a Case Study' (2015).

[126] Manish Singh, 'Telegram, Nearing 500 Million Users, to Begin Monetizing the App' (TechCrunch, 23 December 2020) <https://social.techcrunch.com/2020/12/23/telegram-to-launch-an-ad-platform-as-it-approaches-500-million-users/> accessed 8 January 2021.

[127] BR, 'Hildmann, Naidoo & Co.: Warum Verschwörungsfans Telegram nutzen' (BR24, 8 May 2020) <https://www.br.de/nachrichten/netzwelt/hildmann-naidoo-and-co-warum-verschwoerungsfans-telegram-nutzen,RyOCmN4> accessed 12 December 2020.

[128] Aleksi Knuutila and others, 'Junk News Distribution on Telegram: The Visibility of English-Language News Sources on Public Telegram Channels' 1.

[129] ibid.

[130] Telegram, 'Terms of Service' (Telegram) <https://telegram.org/tos> accessed 12 December 2020.

[131] ibid.

[132] Ahmad Shehabat, Teodor Mitew and Yehia Alzoubi, 'Encrypted Jihad: Investigating the Role of Telegram App in Lone Wolf Attacks in the West' (2017) 10 Journal of Strategic Security 1; Ahmet Yayla and Anne Speckhard, 'Telegram: The Mighty Application That ISIS Loves' [2017] International Center for the Study of Violent Extremism (ICSVE) 10.

[133] Alexandre Bovet and Peter Grindrod, 'The Activity of the Far Right on Telegram v2.11' (2020) 11, researchgate.net.

[134] EUROPOL, 'Europol and Telegram Take on Terrorist Propaganda Online' (Europol, 2019) <https://www.europol.europa.eu/newsroom/news/europol-and-telegram-take-terrorist-propaganda-online> accessed 8 January 2021.

[135] Knuutila and others (n 140).

Telegram has a much less strict approach to governing COVID-19 information than other major platforms. Yet, Pavel Durov started to promote verified channels[136] on his platform.[137] Those channels can be verified if an active official channel, bot or a public group is concerned and another platform (Twitter, Facebook, Instagram or YouTube) already has verified a similar account.[138] If the user has no verified account on any of those platforms, an undisputed page on Wikipedia that is in accordance with its ‘Notability Guidelines’[139] also is accepted by Telegram. Ordinary user accounts cannot be verified. These are reserved for ‘big and active official channels and bots’.[140] Therefore, Telegram expands their cooperation with worldwide[141] health ministries[142]. Telegram also allowed for notification of users by verified channels do address COVID-19[143].

Hui Xian Ng and Loke Jia were researching group behaviour and misinformation on Telegram in relation to the COVID-19.[144] Most activity could be measured at midday or between eight to ten pm. According to them zero point zero five per cent of overall content could be classified as misinformation. The corresponding answers to misinformation on the platform express scepticism to overall zero point four per cent. The authors found that activity within the group increased, when governments announces were made. Whereas the soar in confirmed COVID-19 cases did not influence the activity level upon the platform as much. Hui Xian Ng and Loke Jia also found, that the sentiment of the user’s content could be labelled within a rather negative spectrum which correlates to governmental communication.

Conclusion and Outlook

Private Ordering of COVID-19-related Content

During the pandemic all of the platforms mentioned above took some measures related to COVID-19 while the amount of action differs. Telegram is based on a very broad understanding of free speech. Its one-to-one and one-to-few communication channels are rightly protected by law, but the groups and other one-to-many communication facilities leave room for largely unregulated online speech which can turn

[136] Telegram, ‘Telegram Channels’ (Telegram, 29 January 2018) <https://telegram.org/tour/channels> accessed 8 January 2021.

[137] E Hacking News, ‘Pavel Durov: The World Will Not Be the Same after the COVID-19 Pandemic’ (E Hacking News - Latest Hacker News and IT Security News) <https://www.ehackingnews.com/2020/04/pavel-durov-world-will-not-be-same.html> accessed 12 December 2020.

[138] Telegram, ‘Page Verification Guidelines’ (Telegram) <https://telegram.org/verify?setln=en> accessed 8 January 2021.

[139] Wikipedia, ‘Notability’, Wikipedia (2020) <https://en.wikipedia.org/w/index.php?title=Wikipedia:Notability&oldid=995288718> accessed 8 January 2021.

[140] Telegram, ‘Page Verification Guidelines’ (n 152).

[141] Ministerio de Salud Pública de Cuba, ‘Canal Oficial Del Ministerio de Salud Pública de La República de Cuba Para Ofrecer Información Sobre La #COVID19.’ (Telegram) <https://t.me/MINSAPCuba> accessed 8 January 2021; Ministry of Georgia, ‘StopCoV.Ge GE’ (Telegram) <https://t.me/StopCoVge> accessed 8 January 2021; German Federal Ministry of Health, ‘Corona-Infokanal Des Bundesministeriums Für Gesundheit’ (Telegram) <https://t.me/Corona_Infokanal_BMG> accessed 8 January 2021; Government of India, ‘MyGov Corona Newsdesk’ (Telegram) <https://t.me/MyGovCoronaNewsdesk> accessed 8 January 2021; Italy Ministry of Health, ‘Ministero Della Salute’ (Telegram) <https://t.me/MinisteroSalute> accessed 8 January 2021 Italy Ministry of Health, ‘Ministero Della Salute’ (Telegram) <https://t.me/MinisteroSalute> accessed 8 January 2021; Russian Ministry of Health, ‘СТОПКОРОНАВИРУС.РФ’ (Telegram) <https://t.me/stopcoronavirusrussia> accessed 8 January 2021.

[142] Telegram, ‘Coronavirus Info Telegram’ (Telegram, 26 March 2020) </s/corona?before=23> accessed 4 January 2021 and Leong Dymples, ‘Responding to COVID-19 with Telegram’ (East Asia Forum, 1 May 2020) <https://www.eastasiaforum.org/2020/05/01/responding-to-covid-19-with-telegram/> accessed 12 December 2020.

[143] Telegram, ‘Coronavirus News and Verified Channels’ (Telegram, 2020) <https://telegram.org/blog/coronavirus> accessed 4 January 2021.

[144] Lynnette Hui Xian Ng and Yuan Loke Jia, ‘Is This Pofma? Analysing Public Opinion and Misinformation in a COVID-19 Telegram Group Chat’ (2020).

problematic.[145] This gap between Telegram and the other platforms grew when measures and moderation on other social networks or messaging services became stricter. Facebook, Twitter and YouTube all have taken a selection of different means to tackle COVID-19.

The 'Organisation for Economic Co-operation and Development' (OECD) provides four recommendations to handle the pandemic: first ‚supporting a multiplicity of independent fact-checking organisations'; second, 'ensuring human moderators are in place to complement technological solutions'; third, 'voluntarily issuing transparency reports about COVID-19 disinformation'; fourth, 'improving users' media, digital and health literacy skills'.[146]

The first recommendation was *in nuce* supported by Facebook, Twitter and YouTube. The second recommendation was only partly deployed through the platforms and was not implemented when lockdowns were in place. The third recommendation was of special importance, because only with increased transparency the phenomenon of misinformation can be studied properly and tackled across platforms. The fourth recommendation is also partly employed by Facebook, Twitter and YouTube.

The European Commission also provided recommendations to digital companies.[147] It stressed the visibility of trusted content by authoritative sources, the awareness of users for content that is displayed to them, the detection of harmful content and the reduced advertising for disinformation.[148] Platforms largely incorporated the recommendations.

Misinformation can only be tackled effectively if measures are taken coherently upon platforms. With a general increase in users and views this year the platforms have a severe duty to prevent users from harm through their offered services. This increase in numbers also will lead to a gain in profit for most of the platforms. Content moderation is at the core of company's service and has changed for Facebook, Twitter and YouTube. The working conditions for moderators at Facebook are problematic especially during the pandemic. Most had to work from home or were unable to work. That is why the usage of automated systems for content moderation soared for Facebook, Twitter and YouTube. Automated systems have drawbacks compared to human content moderation and could foster the spread of misinformation online. On average 25 per cent of content relating to COVID-19 could be classified as misleading on all platforms.[149] This amount further increased up to 31 per cent when the users stayed anonymous.[150]

The recommendation algorithms employed by the platform act as 'digital curators' on platforms and are responsible for most of the content consumed by users.[151] Because the business model platforms employee user's views and reaction to content is an important key performance indicator, misleading content with

[145] Kettemann/Fertmann, 'Viral Information: How States and Platforms Deal with Covid-19-related Disinformation: an Exploratory Study of 18 Countries' (Hamburg: Verlag Hans-Bredow-Institut, 2021), GDHRNet Working Paper #1, 126.

[146] OECD, 'Combatting COVID-19 Disinformation on Online Platforms' (OECD, 3 July 2020) <https://www.oecd.org/coronavirus/policy-responses/combatting-covid-19-disinformation-on-online-platforms-d854ec48/> accessed 15 January 2021.

[147] European Commission, 'Disinformation: EU Assesses the Code of Practice' (European Commission - European Commission, 10 September 2020) <https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1568> accessed 15 January 2021.

[148] ibid.

[149] Kouzy and others (n 75).

[150] ibid.

[151] Gudivada, Rao and Paris (n 93).

high engagement and visibility can increase the company's profit.[152] This relation between profit and polarizing content can also explain why YouTube is monetising COVID-19 content after it has banned it only a month before. Trusted sources are still under-represented and should be promoted even more on the platforms. It is important to give authoritative sources and trusted healthcare content a loud voice in the pandemic to keep misinformation at bay.

Outlook

Platforms are here to stay. Their communicative role is likely to remain influential and to even to grow, especially in developing states. Private ordering, that is the application of private norms in online spaces through which they are constituted as normative orders, will continue to be a useful concept to understand platform behaviour. States and platforms both have different duties and responsibilities vis-à-vis freedom of expression. As we have shown, private ordering has its limits: Public law is necessary in order to control public values. Privately constructed normative orders often lack a socially responsible finality. Even carefully constructed quasi-judicial entities, meant to increase legitimacy of platform law, suffer from flaws.

A basic problem of content moderation cannot be solved by even the most cleverly crafted law. It is this: While the primary responsibility for safeguarding individual spheres of freedom and social cohesion rests with states, it is platforms that have the primary power (in the sense of effective impact) to realize and influence rights and thereby cohesion. They set the rules, they design the automated tools, they delete and flag. Platforms have started to do better in terms of protecting rights, but they are still far off - in normative terms - when it comes to ensuring social cohesion.

Currently, all major platforms follow the approach of leaving as much "voice" online as possible (though overblocking happens), deleting only dangerous postings (e.g., death threats) and adding counter-statements (e.g. warnings) to problematic speech (e.g. disinformation). Covid-19 has gradually changed this, as we have seen above. For the first time, a cross-platform phenomenon became visible: the recognition that mostly lawful speech could be highly corrosive of societal values (like public health) and that platforms needed to use all tools in their normative arsenal, automatic filtering, downranking, deleting, counterinformation, flagging, to support efforts to fight Corona. If it worked overall rather well for fighting Corona, the one questions which remains is this: What about protecting other societal values against less-well designed threats? Here both more rights-conscious and more authoritarian futures are possible and continued engagement in critical platform research is essential.

[152] Svenja Boberg and others, 'Pandemic Populism: Facebook Pages of Alternative News Media and the Corona Crisis -- A Computational Content Analysis' [2020] arXiv:2004.02566 [cs] 21, 11.