Hertie School
Centre for Digital Governance

Harmful content on Twitter during COVID-19: Approaches and solutions

Zhihao Zhong, MPP 2022
Thilo Sander, MPP 2022

29 May 2022

The COVID-19 pandemic brought along an "infodemic" which has been amplified in the digital world. We intend to discuss the harmful content on Twitter during the COVID-19 pandemic and the approaches to solve the problem taken by different sides. We first lay out how disinformation and misinformation have become problems negatively affecting society. We then present solutions and roles by three organisational clusters, namely the platform Twitter itself, governing bodies in the EU and Germany, and civil society such, as the news media and NGOs. Firstly, Twitter's response can be found in their guidelines against harmful content during this pandemic. Secondly, we analyse Germany's Network Enforement Act and the EU's Digital Services Act (DSA). Thirdly, we present criticism from civil society such as "netzpolitik.org" and "HateAid". In conclusion, the paper finds it important to have all three clusters working together to form an effective system against harmful content online, which is in fact advocated by the EU's DSA.

#Twitter
HarmfulContent

1   Introduction

Since the outbreak of COVID-19, in addition to the remote work phenomenon, communication in many places has been diverted to the digital space. As a result, illegal and harmful content is more frequently observed on social media platforms, such as Twitter (Uyheng & Carley, 2021). Information on the internet has thus been discussed much more.

There are many problems that come with the internet, and already back in 1996 the European Union defined the category "Illegal and harmful content on the Internet", foreseeing the potential problems that would come with it (Commission of the European Communities, 1996). Terms such as disinformation, misinformation, and hate speech are being used much more often to address the issues arising alongside the development of this new people-connecting technology. In this research paper, we narrow down and focus on the problems and consequences of harmful content on the social media platform Twitter during the COVID-19 pandemic in Germany (i.e., 2020-2022).

Adapted from the definition by the Council of Europe, researchers put forward the concepts that fall into harmful content: "Misinformation refers to false information shared without the intention of causing harm. Disinformation refers to false information shared with the intent of causing harm. Mal-information deals with true information shared with the intent of causing harm via leaks and through harassment." (Goldzweig et al., 2018).

We first discuss the problems that have emerged or been amplified due to the pandemic, and then list the solutions or roles played respectively by the platform, the state, and civil society. Discussion of the roles these actors play regarding harmful content is included in each subsection, followed by the conclusion.

2   The problem

The combination of the internet and the "infodemic" has contributed to the great problem. Disinformation and misinformation during public health crises often become prevalent and grow into great concerns for citizens and policy makers alike (Biradar et al., 2022; Dredze et al., 2016). In recent years, researchers have confirmed an increase in the use of social media networks such as Twitter and Facebook by users to obtain information such as news (Ohme et al., 2021). This increases the severity when there is a misuse of the information on the internet. As the World Health Organization has coined the term "infodemic" for the times in pandemics when there was a lack of expertise and verification as well as empirical results in the early stages, COVID-19, without exception, has shown what the "infodemic" is like, regarding the spread of harmful contents. (Guess & Lyons, 2020; Okan et al., 2020). Further explanation on how the harmful content is given below in the case of Twitter use in Germany, including examples such as conspiracy theories, right-wing extremism, and how it becomes a problem.

Disinformation and misinformation such as conspiracy theories and rumours then become prevalent (Shahsavari et al., 2020). The problem becomes greater and more severe with more users surfing the internet. Posts about COVID-19 on Twitter had been increasing during the early outbreak around March 2020 as the infection and

death toll went up in 12 European countries, including Germany (Gencoglu & Gruber, 2020). Researchers studied the disinformation and misinformation on Twitter, and found that conspiracy theories were a great part of such information (Batzdorfer et al., 2021). During the pandemic people could have been obtaining an increasing number of anti-establishment information and coverage from the alternative critical viewpoint on COVID-19 on social media instead of traditional media sources (Ohme et al., 2021). This perhaps hindered the "prosocial behaviour".

How does this happen? People went to the internet for information about COVID-19, overlooked the credibility check and shared everything they saw on the internet out of panic (Biradar et al., 2022). Biradar et al. argue that such practice contributed to the amplification of the widespread misinformation on online platforms. Furthermore, there is an association between reliance on social media for information and the possibility of believing in disinformation and misinformation (El-Far Cardo et al., 2021).

Fake accounts created by ill-intentioned users steal other people's identities to post false information on social media platforms, including Twitter (Biradar et al., 2022). Some of the disinformation contains partly true and partly made-up elements that connect with readers. The amount of disinformation and misinformation is rising at a rate that concerns researchers (Biradar et al., 2022). The mixed information lowers peoples' guards as it appears more trustworthy by presenting some pieces of factual information.

Over half of the population in Germany, according to Okan et al. (2020), have been found to possess literacy levels of "inadequate" or "problematic" in regards to health and well-being literature. Okan et al. also found that the youth population tends to feel confused about pandemic-related information partly because of their high exposure to diversified sources of information, which is in part attributable to social media use. Health literacy, different from media literacy, could be interacting with each other or complicating the process to obtain reliable information.

Besides knowledge, El-Far Cardo et al. (2021) found that in Germany political opinions also play a role in information obtention. Those leaning politically left who possess higher trust in health professionals and find information from reliable sources like public media tend to take the pandemic seriously (El-Far Cardo et al., 2021). The opposite applies to the right-leaning conservative who has lower trust in authorities and is prone to trusting misinformation (ibid). Another study finds that if a person has a robust perception of misinformation, it is more likely that they are incentivised to look for more information (Hameleers et al., 2020). Those who perceive a huge disinformation problem tend to avoid the news and are more reluctant to comply with the rules by authorities, and scholars thus suggest increasing public trust to counter this trend (Hameleers et al., 2020).

As news reports documented that conspiracy theories and misinformation spread at an extraordinary pace in Germany, the right-wing party "Alternative für Deutschland" (AfD), which represents a populist point of view, took advantage of the "infodemic" and generated content against institutions, discrediting government and officials (Leschzyk, 2021). The posts by AfD were at one point dominating among Tweets in Germany compared to other German parties (DemTech, 2017).

Another problem amplified by the pandemic is online hate speech targeted against certain minority communities, e.g. the Chinese community and East-Asian-presenting people (Budhwani & Sun, 2020; Uyheng & Carley, 2021).

3   The platform: Twitter

Twitter has faced many challenges moderating harmful content, has responded to public sentiment about mis and disinformation on its platform and has taken measures to combat the various phenomena. One of the most prominent practices that Twitter performed is the banning of former U.S. President Donald Trump, followed by other conservative lawmakers from the Republican Party (BallotPedia, 2022). This triggered more debates on the regulations and power of tech companies like Twitter. Twitter's early attempts to promote healthy debates and address issues on misinformation, as pressured to do so by the US Congress, tried to suspend accounts identified as bots and limit posting through multiple accounts at the same time (Goldzweig et al., 2018).

Regarding our specific research topic, Twitter has a separate page for its "COVID-19 misleading information policy", last updated in December 2021 (Twitter, 2021). The cases that trigger Tweet removal have been listed in detail, which includes disinformation, misinformation and illegal actions. Misleading information will be labelled. Twitter also clarifies what is allowed regarding COVID-19 related information. Punishment and its severity are also listed with permanent suspension being the most severe consequence. In addition, Twitter's Vice President introduced a community-driven approach called 'Birdwatch' to involve the users in reporting problematic Tweets (Coleman, 2021). The guidelines related to COVID-19 were written with clarity and examples. For instance, false claims include "COVID-19 is not a real disease" and "That COVID-19 vaccines are causing magnetic reactions in individuals who have been vaccinated" (Twitter, 2021). This is clarifying for content moderators. However, for disinformation outside the areas of COVID-19, Twitter has still not included specific regulations on its site.

The problem Twitter tried to address seemed to be very vague but the action was motivated by pressure from the US legislature, and the misinformation situation has not improved (Goldzweig et al., 2018). Moreover, it is not clear how decision-making is done, nor are the consequences when social media companies fail to take responsibility, and such lack of transparency and accountability needs to be considered (Goldzweig et al., 2018; Schaake & Vermeulen, 2016). Although the rules regarding COVID-19 are clearer, it is not possible to see exactly how many and who are being punished under what grounds, and it is certainly not able to tell whether the punishment falls into the case of the pandemic.

In Twitter's most recent effort to address the war in Ukraine in 2022, Twitter added a "Crisis Misinformation Policy" for armed conflict to deal with misleading information, for which the hardest "7-day time-outs" can be applied (Twitter, 2022). Both COVID-19 and armed conflict-related guidelines are retrospective regulations proposed by Twitter itself and users do not play a part in the rulemaking on this platform filled with user-generated content. Nevertheless, the upside is that such self-regulation practice is very fast-paced and grants companies the autonomy to act promptly and flexibly. Whether the regulations can be fully implemented requires a separate assessment.

Inconsistency of practices is possible under Twitter's self-regulation. Millionaire Elon Musk revealed the attempt to purchase Twitter and reverse the ban on Trump (Dang, 2022). This is a reminder that social media platforms are after all private companies that can make decisions which they see fit; possibly to maximise the company stakeholders' financial benefits or to allow responsible moderators to decide arbitrarily. Content moderators are not discussed or made transparent by Twitter on its website. Content moderation is often outsourced to companies, allowing these moderators the right to deal with content including removal (Roberts, 2017). Nonetheless, the qualification of these persons is often not open to users for oversight, nor is the fact that moderating harmful content could cause psychological problems, which could further worsen the sustainability of content moderation. The lack of transparency is additionally upheld by the non-disclosure agreement signed by these moderators (Roberts, 2017).

Little effort has been found regarding social media literacy education offered by Twitter, except for the partnership with UNESCO on media and information literacy, but the related content was not linked to the media release article (Costello, 2019). Can Twitter contribute more on media literacy? Some may say that Twitter offering COVID-19 guidelines explaining banned behaviours and content is a form of media literacy education, but whether education is an objective of these guidelines remains unclear.

4   Germany & the EU

Social media companies in Germany must follow strict regulation, and the European Union (EU) is proposing a broader and more comprehensive regulation which is still in progress. The Twitter user population in Germany alone has reached 7.75 million in January 2022, roughly one tenth of its population (DataReportal, 2022). Twitter has to comply with the German regulation and publishes reports regularly for its users in Germany, but this practice is not yet performed in the broader EU context. This section will discuss solutions adopted and being proposed by Germany and the EU, in which the latter will serve as a comparison to the German approach.

German policy makers have been trying to increase the speed and efficiency of response mechanisms when it comes to harmful content such as hate speech and disinformation. Germany adopted the Network Enforcement Act (Netzwerkdurchsetzungsgesetz, in short NetzDG) in October 2017, paving the way for users to submit complaints about online content which then require action within 7 days (Bundesministerium der Justiz, 2022). A bi-annual report of how platforms deal with the content is also required. The NetzDG is among the very few legislations in the world on digital platforms and targets illegal content including some forms of hate speech and disinformation (Bayer et al., 2021). As legislation, the NetzDG does not clearly stipulate what constitutes illegal content, and thus less clear but potentially unlawful content would be given a maximum of 7 days to be acted upon rather than the 24 hours for clearly illegal content.

Through such regulation, companies accumulate experience and could optimise their operations to abide by the law. There is also criticism that the law contains the danger of over-blocking, indicating a tendency to moderate the content under request, which companies tend to do to reduce legal risk. Without giving clear explanations to affected users, this can lead to potentially harming the freedom of expression (Bayer

et al., 2021). The critique is partly validated by the finding that platforms rely on algorithmic content moderation to efficiently delete improper content all the while the challenge to recognise hateful content and disinformation remains too huge for what technology can actually achieve today (Heldt, 2019).

Positive results have also been observed. After implementing the NetzDG, Raphaela Andres and Olga Slivko, through studying followers of the right-wing AfD party, found that the harmful effects of hate speech were dampened, especially when companies were encouraged to deal with hateful content in a prompt manner (Andres & Slivko, 2021).

Even with the NetzDG legislation in Germany, the day-to-day practice of content moderation still relies heavily on the tech companies. This seems to align with the approach taken by the EU. The EU has been proposing the Digital Services Act (DSA) package containing the Digital Services Act and the Digital Markets Act that define the rules of digital services and delegate platforms as the "gatekeepers" of content (European Commission, 2022; European Parliament, 2022). The DSA policy process involves multiple stakeholders including states, social media companies, experts, and civil society (de Jesús Butler, 2008; European Commission, 2022; European Parliament, 2022). The DSA as a process-based approach to platform regulation places more emphasis on accountability and transparency by establishing an independent administration of oversight and opening access for researchers to study (Stockmann, 2022).

Co-regulation could help platforms to create and adopt universal guidelines that fulfil the DSA objectives and are also practical for platforms per se, that eliminate the downsides of self-regulation. Since tech companies will be the frontline implementation and enforcement forces in digital space, they have been included in the consultation during the law-making process and are termed "gatekeepers" in the DSA proposal. In a way, this is sensible because social media platforms prioritise their community guidelines (Heldt, 2019), and having a universal guideline will help clarify responsibilities for not just the platforms but also internet users and law makers. Including tech companies in the legislation process allows them to lobby in their own interest and have the ability to affect the legislation, but the transparent democratic process in the EU is designed to and hopefully can maximise public interest while being pragmatic for tech companies. The legislation has been proposed to enter into effect in 2024.

5   Civil society: NGOs and media

Civil society like non-governmental organisations (NGOs) and especially the media plays a very important role ineffectively regulating and controlling social media platforms. They oftentimes play the role of the watchdog that unveils inappropriate reactions to any kind of mis- and disinformation. They hold social media platforms accountable and reveal loopholes that need to be closed up, either by the social media platforms themselves or through legislation from public authorities.

In Germany there is an active civil society that tries to hold the social media platforms and especially Twitter accountable for their actions. In addition to the large, high-quality newspapers in Germany, two organisations are especially important as watchdogs. The first is the news website "netzpolitik.org" that publishes articles

mainly in the field of digital politics and freedom rights in the 21st century. It is funded almost exclusively by donations of individuals and thus is independent from publishers in the traditional sense. The other organisation is an NGO that supports victims of online hate crimes called "HateAid". It offers consulting for victims of hate crimes and supports them in case of potential costs for court trials.

Civil society has criticised Twitter in three different ways, at least in Germany. The first criticism is that the guidelines which Twitter uses to execute their content moderation is not transparent and not publicly available (Stell, 2021). A court case from six NGOs in France against Twitter with a similar complaint led to the ruling that Twitter must publish their guidelines (UEJF - Union des Etudiants Juifs de France, 2021). Without any transparency about the principles and guidelines Twitter aims to establish on their platform, it is not possible to see whether they adhere to their own standards.

The second criticism arose as Twitter complied with the court ruling. After Twitter published certain elements of their content moderation guidelines, civil society stepped in to monitor whether Twitter had been adhering to its own guidelines. This was apparently not always the case. Max Otte, the candidate from the German far-right party AfD for the federal presidency, the highest German office, repeatedly used paid advertisements on Twitter to spread his messages to a wider audience. According to Twitter's own guidelines political advertisement is banned on the platform. This advertisement was possible as Twitter did not properly oversee the advertisements of Max Otte (Dachwitz, 2022). To this date, there have been no consequences known to the public for Max Otte's political advertising on Twitter, and Otte's Twitter profile is still available.

The third criticism has demanded improvements in both Twitter's own guidelines regarding harmful content and legislation by public authorities, like the German federal government or the European Union. For example, the online magazine "The Conversation" showed that Russia uses a loophole on Twitter to spread disinformation after the Russian Television channels Sputnik and RT were shut down internationally (Thompson & Graham, 2022). Official government accounts have more freedom regarding the content they can create and share on Twitter and Russia uses its official accounts (e.g. of their embassies) to coordinate the dissemination of disinformation by retweeting the same sources and articles. These accounts exploit a loophole in Twitters guidelines to continue with disinformation campaigns. The German news website "netzpolitik.org" demanded a closure of that loophole (Pallaske, 2022). Another criticism is that Twitter is not really banning disinformation from their platform: If some content in one country is found to be illegal or not compatible with local laws, it is often only geo-blocked. Geo-blocking means that the content in question is only blocked in the country where it is illegal, but still visible for people located in other countries and accessing the content from there. For example, content created in Germany that questions the Holocaust is blocked in Germany, but potentially still available in Austria and Switzerland (Laufer, 2021).

The regulation of social media platforms and of Twitter in particular benefits immensely from the work of civil society. For example, the NGO HateAid pushed a court case where a meme with a false quote of German politician Renate Künast circulated for years on Facebook. The court finally ruled that Facebook must delete the content and all its related content with the same false quote (HateAid, 2022). This is undoubtedly a remarkable case as this will have influence on the interpretation of the

newly passed Digital Services Act and Digital Markets Act. In another HateAid supported court case, Twitter must pay a fine to a victim of sexual insults on Twitter for the first time (HateAid, 2021).

6 Conclusion

This research paper analysed the approach of three different organisational clusters in dealing with harmful content on Twitter from 2020 till present, in the midst of the COVID-19 pandemic. The three organisational clusters are the legislative bodies of Germany and the EU, civil society organisations and Twitter itself, a social media platform. All three actors are concerned about the development and containment of harmful content on Twitter and all three are actively engaged in the debate on how to best deal with this situation, not only on Twitter but on social media platforms in general.

Although all clusters are active in the debate, it must be stated that all of them can improve their effectiveness and impact regarding the combat against and the prevention of harmful content. First, Twitter is rather not transparent regarding its own guidelines on content moderation and harmful content, and especially about the application of its own guidelines on content posted on its platform. Researchers and civil society have only limited access to data about Twitter's content moderation activities. Second, the legislative bodies can and must improve their prescriptions on how to deal with harmful content on social media platforms regularly. Germany was among the first to pass substantial legislation regarding harmful content online, but this was also criticised by civil society groups due to the fear of over-blocking and censorship. The European DSA and DMA will substantially influence the state of content moderation for all social media platforms. It will likely emerge that not all different aspects are regulated in the best possible way, and some cases might be subject to court cases, meaning that legislation needs to be updated periodically in order to be fine-tuned with respect to effectiveness against harmful content. Third, civil society groups are already very active in the combat against harmful content, but their influence is still limited. A greater diversity of civil society groups that cover all different aspects of harmful content online is desirable. Additionally, civil society groups need to target a larger audience in the middle of society - nowadays the debate is often limited to a very well-informed bubble with limited impact on the overall society.

Finally, it is important to mention that all three organisational clusters (Twitter, legislative bodies, and civil society) are crucial for an effective system against harmful content. If any of these three or any of the interconnections are not working well, harmful content is likely to spread in the future as it does presently.

Bibliography

Andres, R., & Slivko, O. (2021). Combating Online Hate Speech: The Impact of
Legislation on Twitter. *SSRN Electronic Journal*.
https://doi.org/10.2139/ssrn.4013662

BallotPedia. (2022, March 15). *Elected officials suspended or banned from social media
platforms—Ballotpedia*.
https://ballotpedia.org/Elected_officials_suspended_or_banned_from_social
_media_platforms

Batzdorfer, V., Steinmetz, H., Biella, M., & Alizadeh, M. (2021). Conspiracy theories
on Twitter: Emerging motifs and temporal dynamics during the COVID-19
pandemic. *International Journal of Data Science and Analytics*.
https://doi.org/10.1007/s41060-021-00298-6

Bayer, J., Holznagel, B., Korpisaari, P., & Woods, L. (Eds.). (2021). *Perspectives on
Platform Regulation: Concepts and Models of Social Media Governance |
Across the Globe*. Nomos Verlagsgesellschaft mbH & Co. KG.
https://doi.org/10.5771/9783748929789

Biradar, S., Saumya, S., & Chauhan, A. (2022). Combating the infodemic: COVID-19
induced fake news recognition in social media networks. *Complex &
Intelligent Systems*. https://doi.org/10.1007/s40747-022-00672-2

Budhwani, H., & Sun, R. (2020). Creating COVID-19 Stigma by Referencing the Novel
Coronavirus as the "Chinese virus" on Twitter: Quantitative Analysis of Social
Media Data. *Journal of Medical Internet Research*, *22*(5), e19301.
https://doi.org/10.2196/19301

Bundesministerium der Justiz. (2022, May 25). *BMJ | Netzwerkdurchsetzungsgesetz*.
https://www.bmj.de/DE/Themen/FokusThemen/NetzDG/NetzDG_node.html

Coleman, K. (2021, January 25). *Introducing Birdwatch, a community-based approach
to misinformation*.
https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-
a-community-based-approach-to-misinformation

Commission of the European Communities. (1996). *Illegal and harmful content on the Internet*. https://eur-
lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:1996:0487:FIN:en:PDF

Costello, R. (2019, October 24). *Twitter builds partnership with UNESCO on media and information literacy*.
https://blog.twitter.com/en_us/topics/company/2019/twitter-launches-new-media-literacy-handbook-for-schools

Dachwitz, I. (2022, March 1). *Violation of Twitter's ban on political ads: No consequences for far-right presidential candidate*. netzpolitik.org.
https://netzpolitik.org/2022/election-advertising-violation-of-twitters-ban-on-political-ads-no-consequences-for-far-right-presidential-candidate/

Dang, S. (2022, May 11). Musk says he would reverse Twitter ban on Donald Trump | Reuters. *Reuters*. https://www.reuters.com/world/us/musk-says-he-would-reverse-twitter-ban-donald-trump-2022-05-10/

DataReportal. (2022). *Digital 2022: Global digital overview* (p. 195). DataReportal.
https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

de Jesús Butler, I. (2008). Non-governmental Organisation Participation in the EU Law-making Process: The Example of Social Non-governmental Organisations at the Commission, Parliament and Council. *European Law Journal*, *14*(5), 558–582. https://doi.org/10.1111/j.1468-0386.2008.00430.x

DemTech. (2017, September 22). *DemTech | Rightwing populist AfD dominates German Twitter, new study shows*. https://demtech.oii.ox.ac.uk/rightwing-populist-afd-dominates-german-twitter-new-study-shows/

Dredze, M., Broniatowski, D. A., & Hilyard, K. M. (2016). Zika vaccine misconceptions: A social media analysis. *Vaccine*, *34*(30), 3441–3442. https://doi.org/10.1016/j.vaccine.2016.05.008

El-Far Cardo, A., Kraus, T., & Kaifie, A. (2021). Factors That Shape People's Attitudes towards the COVID-19 Pandemic in Germany—The Influence of MEDIA,

Politics and Personal Characteristics. *International Journal of Environmental Research and Public Health*, *18*(15), 7772. https://doi.org/10.3390/ijerph18157772

European Commission. (2022, May 25). *The Digital Services Act package | Shaping Europe's digital future*. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

European Parliament. (2022). *Digital Services Act ***I - Amendments adopted by the European Parliament on 20 January 2022 on the proposal for a regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC*. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:1996:0487:FIN:en:PDF.

Gencoglu, O., & Gruber, M. (2020). Causal Modeling of Twitter Activity during COVID-19. *Computation*, *8*(4), 85. https://doi.org/10.3390/computation8040085

Goldzweig, R., Wachinger, M., Stockmann, D., & Römmele, A. (2018). *Beyond Regulation: Approaching the challenges of the new media environment*. 25. https://www.dahrendorf-forum.eu/publications/beyond-regulation-approaching-the-challenges-of-the-new-media-environment/

Guess, A. M., & Lyons, B. A. (2020). *Social Media and Democracy: The State of the Field, Prospects for Reform* (N. Persily & J. A. Tucker, Eds.; 1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108890960

Hameleers, M., van der Meer, T. G. L. A., & Brosius, A. (2020). Feeling "disinformed" lowers compliance with COVID-19 guidelines: Evidence from the US, UK, Netherlands and Germany. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-023

HateAid. (2021, December 10). *HateAid sets a precedent in Germany: Twitter must pay 6,000 euros financial compensation for not removing sexist insults to a user*. https://hateaid.org/wp-content/uploads/2021/12/Press-Release-Twitter-Case-HateAid.pdf

HateAid. (2022, April 8). *Sensation bei Grundsatzurteil in HateAid-Prozess gegen Facebook: Plattform muss illegale Inhalte endlich konsequent löschen*. https://hateaid.org/wp-content/uploads/2022/04/HateAid-Pressemitteilung-Grundsatzprozess-Kuenast-Facebook.pdf

Heldt, A. (2019). Reading between the lines and the numbers: An analysis of the first NetzDG reports. *Internet Policy Review*, *8*(2), Article 2. https://doi.org/10.14763/2019.2.1398

Laufer, D. (2021, January 27). *Holocaust-Verharmlosung: Wie Twitter dabei scheiterte, bei Antisemitismus durchzugreifen*. netzpolitik.org. https://netzpolitik.org/2021/holocaust-verharmlosung-wie-twitter-dabei-scheiterte-bei-antisemitismus-durchzugreifen/

Leschzyk, D. K. (2021). Infodemic in Germany and Brazil: How the AfD and Jair Bolsonaro are Sowing Distrust During the Corona Pandemic. *Zeitschrift Für Literaturwissenschaft Und Linguistik*, *51*(3), 477–503. https://doi.org/10.1007/s41244-021-00210-6

Ohme, J., Hameleers, M., Brosius, A., & Van der Meer, T. (2021). Attenuating the crisis: The relationship between media use, prosocial political participation, and holding misinformation beliefs during the COVID-19 pandemic. *Journal of Elections, Public Opinion and Parties*, *31*(sup1), 285–298. https://doi.org/10.1080/17457289.2021.1924735

Okan, O., Bollweg, T. M., Berens, E.-M., Hurrelmann, K., Bauer, U., & Schaeffer, D. (2020). Coronavirus-Related Health Literacy: A Cross-Sectional Study in Adults during the COVID-19 Infodemic in Germany. *International Journal of Environmental Research and Public Health*, *17*(15), 5503. https://doi.org/10.3390/ijerph17155503

Pallaske, O. (2022, March 18). *Falschinformationen: Wie Russland die Desinformationsregeln von Twitter umgeht*. netzpolitik.org. https://netzpolitik.org/2022/falschinformationen-wie-russland-die-desinformationsregeln-von-twitter-umgeht/

Roberts, S. T. (2017). UCLA Previously Published Works. In *Content moderation*.
https://escholarship.org/uc/item/7371c1hf.

Schaake, M., & Vermeulen, M. (2016). Towards a values-based European foreign
policy to cybersecurity. *Journal of Cyber Policy*, *1*(1), 75–84.
https://doi.org/10.1080/23738871.2016.1157617

Shahsavari, S., Holur, P., Wang, T., Tangherlini, T. R., & Roychowdhury, V. (2020).
Conspiracy in the time of corona: Automatic detection of emerging COVID-
19 conspiracy theories in social media and the news. *Journal of
Computational Social Science*, *3*(2), 279–317. https://doi.org/10.1007/s42001-
020-00086-5

Stell, D. (2021, July 7). *Frankreich: „Twitter muss endlich Verantwortung übernehmen"*.
netzpolitik.org. https://netzpolitik.org/2021/frankreich-twitter-muss-endlich-
verantwortung-uebernehmen/

Stockmann, D. (2022). Tech companies and the public interest: The role of the state
in governing social media platforms. *Information, Communication & Society*,
1–15. https://doi.org/10.1080/1369118X.2022.2032796

Thompson, J. D., & Graham, T. (2022, March 15). *Russian government accounts are
using a Twitter loophole to spread disinformation*. The Conversation.
http://theconversation.com/russian-government-accounts-are-using-a-
twitter-loophole-to-spread-disinformation-178001

Twitter. (2021, December). *COVID-19 misleading information policy*.
https://help.twitter.com/en/rules-and-policies/medical-misinformation-
policy

Twitter. (2022, May). *Crisis misinformation policy*. https://help.twitter.com/en/rules-
and-policies/crisis-misinformation

UEJF - Union des Etudiants Juifs de France. (2021, July 6). Communiqué de presse:
Twitter est condamné par la justice française à divulguer l'ensemble de ses
moyens de modération. *UEJF - Union des Etudiants Juifs de France*.

//uejf.org/2021/07/06/communique-de-presse-twitter-est-condamne-par-la-

justice-francaise-a-divulguer-lensemble-de-ses-moyens-de-moderation/

Uyheng, J., & Carley, K. M. (2021). Characterizing network dynamics of online hate

communities around the COVID-19 pandemic. *Applied Network Science*, *6*(1),

20. https://doi.org/10.1007/s41109-021-00362-x

Policy Brief, 29 May 2022