

The Paradoxes of Freedom of Expression on Social Media Platforms: Restrictions and Enforcement

How should free expression be limited on Social Media platforms, given the
paradoxes they expose?

Alec Maloney
長谷部 勇矢

July 2022

Ius Gentium Conimbrigae – Centro de Direitos Humanos (IGC/CDH)
Faculdade de Direito da Universidade de Coimbra (FDUC)
Universidade de Coimbra

The writing of this thesis was supervised by:

Professor Jónatas Eduardo Mendes Machado
Executive Director and Director of the Faculty of Law

Carla de Marcelino Gomes
Executive Director

Abstract

This thesis evaluates the societal impact of Social Media platforms since they became ubiquitous. While Social Media platforms can greatly benefit society and promote human rights, they also spawn a myriad of frightening consequences that result in horrific violations of human rights. While the benefits of Social media are undeniable, the technology spawns a great deal of problems for society that remain without a solution. First, Section 1 defines 'Social Media' and analyzes the role of Social Media platforms within society at large. The drastic communications paradigm shift caused by the widespread adoption of Social Media lay the foundations for sections 2 and 3. Section 2 investigates the role Social Media platforms have in the creation of information disorder. Finally Section 3 reviews some of the worst consequences of the communications paradigm shift, radicalization and the incitement of extremist violence.

Table of Contents

Introduction – 6

Section 1: A New Social Paradigm – 8

1.1: Social Media as a Communications Medium – 8

1.2: Social Media and the News – 10

1.3: Social Media and Politics – 13

1.4: Relevant Legal Restrictions of Speech on Social Media in the United States and European Union – 14

Section 1 Conclusion – 15

Section 2: Fake News – 17

2.1: A Brief History of Fake News – What Questions does it Raise? – 17

2.2: So What is ‘Fake News’? – 20

2.3: The Threats Disinformation Currently Poses to Human Rights and Regulatory Solutions – 26

2.3.1: A Brief Discussion of the Impact of Social Media on the Right to Non-Discrimination and the Right to Private Life – 27

2.3.2: A Comparative Examination of Legal Frameworks in the United States, China, and the European Union – 28

Section 2.3.3: Derogations of Human Rights in Emergency Situations – The Human Rights to Health and Life – COVID-19 and the Russian Invasion of Ukraine – 34

Section 2 Conclusion – 42

Section 3: Online Extremism and the Role of Social Media Platforms in the Radicalization of Individuals – 44

3.1: Definitions and Factors Contributing to Online Extremism and Radicalization – 44

3.1.1: QAnon and Other Conspiracy Theories – 47

3.2: The Role of Social Media Platforms in the Promotion of Online Extremism and Radicalization – 49

3.2.1: Radicalization Pipelines on Youtube – 51

3.2.2: Stochastic Terrorism and the ‘Chan’ Forums – 53

3.2.3: The Facilitation of Coordinated Extremism by Social Media Platforms – 55

Section 3 Conclusion – 57

Conclusion – 58

Bibliography – 61

Introduction

This thesis aimed to create a compelling analysis of the ways in which society has been shaped by Social Media platforms since their mass adoption. Social Media platforms have become ubiquitous facets of life despite the absolute failure of the businesses that own and operate the platforms as well as the governments that oversee them to ensure that this extremely powerful tool is used for the promotion rather than subjugation of human rights.

This thesis is split into three sections which are designed to illustrate the double-edged sword that is speech on Social Media platforms. Section 1 will contextualize the ways in which Social Media platforms changed the communications paradigm to lay a foundation for the rest of the thesis. Section 1 begins by defining the term ‘Social Media’ and exploring the scale and scope of its usage. Section 1 proceeds to discuss the democratization effect Social Media has had on media consumption and the subsequent interconnectedness of Social Media platforms and the news. Section 1 will continue by discussing the linkages between Social Media platforms and politics and the general nature of Social Media as a Habermasian public sphere. The section concludes with an evaluation of central legal frameworks in the United States and Europe, that illustrates both the history and potential future of regulations of online speech. The section will then examine the history of regulations concerning speech and usage on Social Media and potentially transformative developments of legal frameworks governing the use of and speech on Social Media platforms before concluding.

Section 2 will discuss the dynamics of the spread of fake news spreading on Social Media platforms, the subsequent risks posed to human rights, and legal recourses to disinformation in several contexts. Section 2 will discuss the term ‘fake news’ before focusing instead on disinformation and information disorder in Section 2.2, which will also explain some of the most common motives for the dissemination of disinformation. Section 2 will end by reviewing the potential threats posed to human rights by disinformation on Social Media platforms and regulations that purportedly protect against these threats, cementing disinformation and information disorder’s role as a root cause for downstream consequences of Social Media platforms.

Section 3 will investigate some of the worst consequences of disinformation and information disorder fomented by Social Media platforms, the systemic radicalization of users and the incitement of extremist acts. This section will explain the demographics and ideology behind modern extremism and explore popular online disinformation narratives such as the QAnon conspiracy theory and white supremacist ‘Great Replacement theory’ that have inspired bloody attacks against groups that are vulnerable on the basis of their race/ethnicity, gender/sexuality, national origin, or other factors. Section 3 will discuss the threats presented by individual radicalization and the organization of extremist groups. Section 3 will also discuss the central role played by Social Media platforms in radicalization. Section 3 will conclude with an investigation of the most harmful extremist acts in recent history.

This thesis was written using a qualitative methodology that focused on explaining the implications of speech on Social Media platforms through relevant, reasonably current illustrative examples. The sources used included primary sources such as laws and direct reporting as well as secondary sources, namely in the form of journal articles and reports.

Section 1: A New Social Paradigm

This section will contextualize the impact of Social Media on the broader landscape of communications and the human right to free expression. Section 1 will discuss some societal issues instigated or influenced by Social Media to lay the groundwork for later sections. Section 1 will begin by defining Social Media and contextualizing the scale and scope of its usage. The section will continue by discussing Social Media as the predominant mass communications medium and the paradigm shift caused by the rapid and prevalent replacement of traditional mass communications media by Social Media platforms. Section 1 will then illustrate the interconnectedness of Social Media platforms and the news by examining the role of Social Media as the primary distributor of news and the subsequent effects on the broader landscape of news. The section continues with a brief investigation of the general interactions between Social Media and politics and a theoretical discussion on the nature of Social Media as a space for expression. The section will then examine the history of regulations concerning speech and usage on Social Media and potentially transformative developments of legal frameworks governing the use of and speech on Social Media platforms before concluding.

1.1: Social Media as a Communications Medium

New technologies, especially Social Media platforms, have changed the rules of the game for social dynamics, communications, and free expression. Social Media is a term that describes a type of online platform wherein users can interact with others by creating and sharing content in a way that resembles a public setting. Major Social Media platforms include Facebook, YouTube, LinkedIn, Twitter, etc. Aichner et al. define Social Media as “an umbrella term that describes a variety of online platforms, including blogs, business networks, collaborative projects, enterprise social networks (SN), forums, microblogs, photo sharing, products review, social bookmarking, social gaming, SN, video sharing, and virtual worlds.”¹ Similar to previous instances of innovative technology reshaping the landscape of societal communications, the prominence of Social Media in society at large has skyrocketed. As with the printing press, radio, and television before it, Social Media has fundamentally changed most of the social dynamics we experience, from interpersonal relationships and entertainment to commerce and politics. A key differentiator between Social Media and its predecessors is that Social Media users engage with the platform in a two-way, participatory relationship rather than a one-way consumption relationship. The scale, scope, and participatory nature of Social Media empowers users to exercise their right to free speech by posting content, sharing existing content

¹ Aichner T and others, ‘Twenty-Five Years of Social Media: A Review of Social Media Applications and Definitions from 1994 to 2019’ (2021) 24 *Cyberpsychology, Behavior, and Social Networking* 215 <<https://www.liebertpub.com/doi/full/10.1089/cyber.2020.0134>> accessed 3 May 2022

from other users, and creating or joining communities for any interest. Social Media platforms are creating a brand new communications paradigm by offering instantaneous communication with huge scale, global reach, and highly personalized experiences for each user. No antecedent communications technology to Social Media was truly able to offer instantaneous communications accurately personalized for each user on a global level like Social Media has. Richard R. John argues that legal principles concerning free expression should be updated to reflect the reality created in part by these new technologies as Social Media has transcended the temporal and spatial boundaries of naturally constraining free expression on antecedent communications technologies.² We have incorporated new technologies like privately moderated Social Media platforms into the existing landscape of expression without reevaluating the norms of free expression to account for how it has shaped society. The double-edged sword of the Social Media communications paradigm is that it empowers speakers that increase nonconformity in ways that could promote or threaten the rights of others. The absolute scale and scope of Social Media is astounding: According to data reported by Kepios in late April of 2022, there are 4.65 billion Social Media users, which equates to 58.7 percent of the global population, who each use the internet for an average duration of six hours and twenty-three minutes daily.³ While each major iteration of the communications paradigm has offered individuals greater and greater capabilities to distribute content or information, Social Media has taken the ability to reach massive amounts of people to an entirely new, unprecedented level. In the past, if an individual was dissatisfied with their media options in areas such as news or entertainment, they could only create a competing newspaper or entertainment firm if they could overcome the barriers to entry such as production and distribution capabilities before they could distribute information on a large scale. In contrast, the barriers to entry to becoming a user on a Social Media platform are minuscule: Any person who can access the internet can become a Social Media user and create or consume content at little to no cost. The space can be quite profitable for some that make content to share on Social Media. Another technological innovation that has increased the importance of online communications, including Social Media, is the global increase in smartphone use. The same Kepios report found that of the 5.00 Billion total internet users across the globe in April 2022, 92.4 percent of users, about 4.62 Billion people, accessed the internet via a mobile device.⁴ While inequalities exist between countries studied, with Less Developed Countries generally worse off, the average global cost of a smartphone in July 2021 was USD 104, or 26 percent of an average monthly income.⁵ As of February 2022, 95 percent of people globally had access to some form of 3GPP cellular

² John RR, 'Freedom of Expression in the Digital Age: A Historian's Perspective' (2019) 4 Church, Communication and Culture 25 <<https://doi.org/10.1080/23753234.2019.1565918>> accessed 28 April 2022

³ Kemp S, 'Digital 2022: April Global Statshot Report' (Kepios 2022) <<https://datareportal.com/reports/digital-2022-april-global-statshot>> accessed 28 April 2022

⁴ Ibid.

⁵ 'How Expensive Is a Smartphone in Different Countries?' (Alliance for Affordable Internet 2021) <<https://a4ai.org/how-expensive-is-a-smartphone-in-different-countries/>> accessed 28 May 2022

coverage, and 86 percent had access to LTE coverage.⁶ The extreme ease of access to these platforms creates environments in which anyone with access to the internet can tell the world how they are feeling, contribute to ongoing discussions, insult strangers, or find new communities with extreme ease. Social Media platforms are vibrant, frenetic virtual spaces wherein users can choose to pursue their interests and join many diverse types of formal or informal communities. While Social Media would not have risen to its ubiquitous position in society without the minuscule barriers to entry to potential users, the obvious consequences of huge usership and low barriers to entry plague Social Media platforms: The ease of becoming a user and the subsequent ease of instantly broadcasting messages to huge numbers of people attracts malicious actors into the space that would use it to threaten or violate the rights of others. Social Media platforms have become so crucial for public expression that we are now experiencing the downstream impacts of the fundamental changes facilitated with or on these platforms.

1.2: Social Media and the News

Social Media has become the primary source of news for many of its users, supplanting traditional communications mediums in many cases. A 2012 study from Gaskins and Jerit found that people were increasingly forgoing consumption of news on traditional media in favor of internet news consumption.⁷ This trend has intensified in the years since the study was published: Over two-thirds (67 percent) of Americans found their news through Social Media in 2021, down four percent from 2020.⁸ The unique features of Social Media have created a landscape for news consumption that barely resembles the interactions between any prior communication media. The free dissemination of public information, also known as the freedom of the press, is a central pillar of a society that guarantees human rights. Freedom of the press is a central tenet of the human right to free expression. The press often plays a crucial role in societal awareness, political accountability, and agenda-setting. Thinkers like John Stuart Mill criticized the pro-status quo agenda-setting power of journalism, which he posited would limit liberty by reinforcing the ‘despotism of custom,’ or reactionary and conservative positions.⁹ The lower barriers to entry to mass communication inherent to the internet and Social Media have given rise to many news outlets such as The Huffington Post, Breitbart News, and

⁶ O’Dea S, ‘Mobile Coverage by Technology Worldwide 2011-2027’ (Statista 2022)

<<https://www.statista.com/statistics/1016292/mobile-coverage-by-technology-worldwide/>> accessed 2 May 2022

⁷ Gaskins B and Jerit J, ‘Internet News: Is It a Replacement for Traditional Media Outlets?’ (2012) 17 *The International Journal of Press/Politics* 190 <<http://journals.sagepub.com/doi/10.1177/1940161211434640>> accessed 3 May 2022

⁸ Walker M and Matsa KE, ‘News Consumption Across Social Media in 2021’ (Pew Research Center 2021) <https://www.pewresearch.org/journalism/wp-content/uploads/sites/8/2021/09/PJ_2021.09.20_News-and-Social-Media_FINAL.pdf> accessed 28 April 2022

⁹ Mill JS, *On Liberty, Utilitarianism and Other Essays* (Mark Philp and Frederick Rosen eds, 2nd edn, Oxford University Press 2015) <<https://oxfordworldsclassics.com/view/10.1093/owc/9780199670802.001.0001/isbn-9780199670802>> accessed 4 May 2022

Buzzfeed, to name a few. Tewksbury and Rittenberg characterize this gold rush of news outlets that originated online as a democratization of public information: In contrast to the antecedent communications technologies such as newspapers, television, and radio, wherein firms or organizations that were often owned or operated by some of the richest and most powerful members of society disseminated the vast majority of public information, Social Media platforms allow for an unprecedented level of participatory involvement in the dissemination of public information from private citizens.¹⁰ This democratization effect has empowered a much larger and more diverse group of entrants into the field of disseminating public information. The diversity and relative size of the current media landscape necessitate that these outlets more closely resemble the userbase's true breadth of demand for content. In other words, the magnitudinal increase of outlets in the market to provide news to the globally expanded audience results in news outlets providing any perspectives and specializations to satisfy the population's interests. The sheer number of outlets able to provide coverage for sports, music, arts, sciences, and any other specific areas of interest bear this trend out. This phenomenon is fundamental to the dissemination of political information within the current paradigm. The ideological principles on offer from sources of political news coverage have expanded to provide offerings to cater to those with any political leanings. For example, American political news coverage on television offers a very basic slate of options among major outlets. The so-called 'Big Three' news outlets have ideologically distributed themselves to appeal to the mainstream American political spectrum broadly: Fox News promotes the ideals and political activities of the Republican Party, while the Cable News Network (CNN) offers a centrist, liberal, or center-left point of view that mainly supports the Democratic Party, and Microsoft/National Broadcasting Company (MSNBC) offers a more progressive, center-left bent. The new communications paradigm has completely disrupted this consolidated domination of the market for news in the United States and elsewhere. Television news consumption has fallen behind news consumption on mobile devices in the United States.¹¹ 97 American news companies received more than 10 million unique online visitors in Q4 2020, with an average unique visitor count of nearly 32.1 Million across these firms. The sheer number of firms, market size, and diverse desires of news consumers have created an environment that necessitates a higher degree of differentiation between firms to capture specific sections of the news market. This higher variability means that more news outlets offer unconventional, niche, or fringe perspectives. An example of a fringe news firm in the American market is InfoWars, a website owned and operated by notorious conspiracy theorist Alex Jones. Jones has constantly used InfoWars to peddle conspiracy

¹⁰ Public Information can also be understood as the News.

Tewksbury D and Rittenberg J, *News on the Internet: Information and Citizenship in the 21st Century* (Oxford University Press 2012)

<<http://www.oxfordscholarship.com/view/10.1093/acprof:osobl/9780195391961.001.0001/acprof-9780195391961>> accessed 29 April 2022

¹¹ Walker M, 'Americans Favor Mobile Devices over Desktops and Laptops for Getting News' (*Pew Research Center*)

<<https://www.pewresearch.org/fact-tank/2019/11/19/americans-favor-mobile-devices-over-desktops-and-laptops-for-getting-news/>> accessed 29 April 2022

theories, some of which have gone ‘viral’ (widely distributed) on Social Media platforms. A central feature of the spread of information on Social Media is the phenomenon of ‘virality’ of online content, where algorithms amplify the spread of content by including it into the personalized feeds of users when they interpret that the content is instigating engagement at a much higher rate than other content.¹² As mentioned previously, Social Media platforms are virtually all algorithmically personalizing every user's experience based on individual and collective usage trends to efficiently serve content and advertising to maximize user engagement and profits. While the owners of Social media companies closely guard the true nature of the exact functioning of these algorithms, the fact that the activity of users is tracked on Social Media platforms to enhance the ability of the platforms to personalize each user’s experience is certain.¹³ Levy found that a user's incidental exposure to various news sites on their Social Media feed significantly influences their news consumption. These outlets are less likely to be biased towards political perspectives contrary to the user's, increasing political polarization overall.¹⁴ The lack of contradictory narratives indicates that the user is in an online space known as an echo chamber, a phenomenon in which users only hear opinions of one type or general perspective. The polarization created by echo chambers often causes users in different echo chambers to exist in entirely different political realities.¹⁵ For example, many of those that participated in the January 6th, 2021 Capitol Riots in Washington, DC did so under the assumption that the results of the election were faked and could be reversed to preserve the presidency of Donald Trump despite a complete lack of tangible evidence of election interference and ability to overturn results. This example of a massive fringe movement that took extreme political action due to misinformation can be attributed in part to the inability of average news consumers to accurately determine the veracity of content presented as verified news on Social Media platforms.¹⁶ Social Media has fundamentally reshaped the global communications landscape, revolutionizing how individuals receive public information (aka. news). The importance of this dynamic cannot be overstated due to the importance of the free dissemination of public information (aka. freedom of the press) for the protection of human rights. The democratization of public information on Social Media platforms dilutes the news market formerly dominated by pro-custom (to borrow Mill’s terminology) perspectives with a

¹² The role of bots in the dissemination of disinformation, is very important and will be discussed in more detail in section 2. The role of the algorithms powering Social Media themselves in promoting disinformation will be discussed further in section 3.

¹³ Gruwell L, ‘Constructing Research, Constructing the Platform: Algorithms and the Rhetoricity of Social Media Research.’ (2018) 6 *Present Tense* 1 <<http://www.presenttensejournal.org/wp-content/uploads/2018/06/Gruwell.pdf>> accessed 30 April 2022

¹⁴ Levy R, ‘Social Media, News Consumption, and Polarization: Evidence from a Field Experiment’ (2021) 111 *American Economic Review* 831 <<https://www.aeaweb.org/articles?id=10.1257/aer.20191777>> accessed 29 April 2022

¹⁵ Fletcher R, Robertson CT and Nielsen RK, ‘How Many People Live in Politically Partisan Online News Echo Chambers in Different Countries?’ (2021) 1 *Journal of Quantitative Description: Digital Media* <<https://journalqd.org/article/view/2585>> accessed 30 April 2022

¹⁶ Klawier T, Prochazka F and Schweiger W, ‘Public Knowledge of Alternative Media in Times of Algorithmically Personalized News’ [2021] *New Media & Society* 146144482110210 <<http://journals.sagepub.com/doi/10.1177/14614448211021071>> accessed 30 April 2022

wider array of nonconformist perspectives that are more accurately distributed across the various demands for news content.

1.3: Social Media and Politics

The paradigm shift in communications caused by Social Media has fundamentally changed the political landscape. Jürgen Habermas defines his concept of a ‘public sphere’ as a social institution that all citizens are guaranteed access to in which something approaching public opinion can be formed through individuals conducting unrestricted conversations on any matter of interest to influence political action.¹⁷ In this 1964 article, he identifies newspapers, magazines, television, and radio as the media of the public sphere. It would be logical to interpret that since Social Media platforms generally seem to fit Habermas’ criteria and have increasingly supplanted the traditional media, Social Media could be considered a proper public sphere. Social Media platforms seem to more accurately match the requirements for a public sphere than traditional media because they allow users to publish, share, or interact with content rather than merely consume information. Park and de Zúñiga confirm this, finding that news consumption on mass communications platforms is associated with political knowledge.¹⁸ Furthermore, the widespread use of Social Media to distribute public information inherently makes Social Media platforms political actors as their algorithms and policies dictate the reach of said public information. Many scholars have shared their view that Social Media platforms are not proper Habermasian public spheres as many people refrain from using Social Media because they are afraid of social consequences for political speech.¹⁹ While workplace punishments for protected speech are seriously problematic, and the echo chamber effect significantly shapes each user’s experience, this conclusion is flawed. Their study conflates the social implications of speech with its legal or political consequences. There are admittedly very tangible social consequences to uncivil political behavior on Social Media platforms.²⁰ However, the ability to critically respond and give feedback to speech in and of itself is crucial for free expression, a far cry from a human rights violation such as censorship or unfair dismissal from work. There are many things that one could say that would not at all meet the criteria for legitimate prohibitions of speech that would provoke social consequences. Without the ability to exchange ideas and

¹⁷ Habermas J, Lennox S and Lennox F, ‘The Public Sphere: An Encyclopedia Article (1964)’ [1974] *New German Critique* 49 <<https://www.jstor.org/stable/487737?origin=crossref>> accessed 1 May 2022

¹⁸ Park CS and Gil de Zúñiga H, ‘Learning about Politics from Mass Media and Social Media: Moderating Roles of Press Freedom and Public Service Broadcasting in 11 Countries’ (2021) 33 *International Journal of Public Opinion Research* 315 <<https://academic.oup.com/ijpor/article/33/2/315/6031067>> accessed 30 May 2022

¹⁹ A) Kruse LM, Norris DR and Flinchum JR, ‘Social Media as a Public Sphere? Politics on Social Media’ (2018) 59 *The Sociological Quarterly* 62 <<https://doi.org/10.1080/00380253.2017.1383143>> accessed 1 May 2022

B) Reilly P, *Digital Contention in a Divided Society Social Media, Parades and Protests in Northern Ireland*. (2021)
 C) Storsul T, ‘Deliberation or Self-Presentation?: *Young People, Politics and Social Media*’ (2014) 35 *Nordicom Review* 17 <<https://www.sciendo.com/article/10.2478/nor-2014-0012>> accessed 1 May 2022

²⁰ Goyanes M, Borah P and Gil de Zúñiga H, ‘Social Media Filtering and Democracy: Effects of Social Media News Use and Uncivil Political Discussions on Social Media Unfriending’ (2021) 120 *Computers in Human Behavior* 106759 <<https://linkinghub.elsevier.com/retrieve/pii/S0747563221000819>> accessed 1 May 2022

subsequently receive feedback, the utility of free expression would be severely diminished. Social Media platforms offer environments incredibly close to the public sphere Habermas envisioned as one of its many varied features and critical societal roles. Habermas envisioned the public sphere as having an antagonistic relationship with the state. In this sense, Social Media platforms and users have scored some notable wins over state power if these states respect all aspects of free expression. The most prominent example of this is the ban of Donald Trump from Twitter on January 8th, 2021, for the role he played in the January 6th, 2021 Capitol Riot. Regardless of the morality or legality of Donald Trump's conduct in the situation in question, it is astounding that Twitter was able to completely ban the sitting President of the United States from its platform through an opaque and unilateral process with no contest from the state at all.²¹ Another salient and associated but positive example is the way that the Black Lives Matter movement was able to quickly scale, attracting global attention and solidarity in large part due to Social Media attention.²² The fact that, while flawed due to private moderation, Social Media can accurately resemble a public sphere in some capacities makes it a space with the potential to spawn meaningful political action by amplifying a diverse array of perspectives.

1.4: Relevant Legal Restrictions of Speech on Social Media in the United States and European Union

As the primary forum for expression within the new communications paradigm, the interactions between politics and Social Media are extremely consequential to the wider political landscape. Therefore, this subsection will review the laws governing expression on Social Media. Historically, there has been a significant difference in the broadcasting regulations and general restrictions on speech in Europe and the United States. The American model of broadcast regulation was shaped by corporate managers who believed that the only way to stop the monopolization of radio was to have a competitive market of private commercial broadcasters funded by advertisements. On the other hand, the European model is based on the British Broadcasting Company, which operated as an independent public service according to John Reith's model.²³ The most important American law protecting the human right to free expression on the internet is Section 230 of the Communications Decency Act (47 U.S.C. § 230), which declares that "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." This provision protects any online service or provider from legal responsibility for what others say and do, except in certain instances of criminal and intellectual property claims. Interestingly, the original intent of the Communications Decency Act was to restrict online speech: The act was the American legislature's first real attempt at regulating online pornographic material to protect

²¹ This issue will be discussed in more detail in later sections.

²² Mundt M, Ross K and Burnett CM, 'Scaling Social Movements Through Social Media: The Case of Black Lives Matter' (2018) 4 *Social Media + Society* 205630511880791
<<http://journals.sagepub.com/doi/10.1177/2056305118807911>> accessed 30 May 2022

²³ *Ibid* 2.

children. The United States Supreme Court struck down the anti-indecency provisions of the Act restricting free speech in the landmark 1997 decision *Reno v. American Civil Liberties Union* (521 U.S. 844) because overly broad and extensive provisions unnecessarily suppressed speech, violating the first amendment of the United States Constitution. However, provisions of the legislation that were judged not to violate the first amendment were preserved. Section 230 of the Communications Decency Act is a law that is somewhat unique to the United States despite similar levels of online activity in other nations. This law is why many online services like Social Media firms are based in the United States. Section 230 has established a system in which Social Media platforms, which often transcend national boundaries, can essentially self-regulate the content on their platforms, restricting speech in a supra-legal manner as they see fit, as in the example above of Donald Trump's Twitter ban. In an attempt to regain regulatory control over expression on Social Media, the European Union passed a law known as the Digital Services Act on April 23rd, 2022. The most significant objectives, according to the press release from the European Parliament on the matter, are listed:²⁴

- To protect the human rights to free expression and privacy on Social Media
- To allow users to better understand how content is recommended to them by making the algorithms of Social Media platforms transparent
- To combat the spread of disinformation on Social Media
- To ameliorate the societal harms of Social Media by requiring administrators to implement new mechanisms of self-moderation on their platform to remove illegal content that violates the rights of others such as hate speech, promotion of terrorism, and other speech restricted by countries within the European Union

While it is too soon to evaluate the impacts of the law, the potential for this law to change the regulatory landscape of speech on Social Media, at least in the European Union, cannot be understated. The developments stemming from this clash of contrasting regulatory frameworks and the overall scope of free speech between the United States and the European Union will be interesting to observe as the true impacts of the law become clearer over time.

1: Conclusion

The rapid and widespread adoption of new technologies such as mobile phones, wireless data, and Social Media platforms has caused a radical paradigm shift in the landscape of mass communications. As may be expected of such a revolutionary technological innovation, the exponentially expanded capabilities of individuals to broadcast their expression to others on

²⁴ 'Digital Services Act: Agreement for a Transparent and Safe Online Environment' (*European Parliament*, 23 April 2022)
<<https://www.europarl.europa.eu/news/en/press-room/20220412IPR27111/digital-services-act-agreement-for-a-transparent-and-safe-online-environment>> accessed 1 May 2022

these new online platforms is affecting many changes offline in society at large. The massive global user base and ease for users to post, interact with, and share information on these platforms has grown, democratized, and diversified the pool of sources disseminating public information, also known as news. The huge demand for content and low barriers to entry dilutes the overall pool of media sources, disrupting the dominance of traditional pro-custom outlets with outlets willing and able to fill any viable niche in the news market. Social Media platforms have become a modern, more representative version of the traditional communications media of Habermas' conception of the public sphere, fundamentally intertwining Social Media platforms and politics. The two-way, participatory exchange Social Media offers users allows them to become political actors instead of mere participants in political actions. However, the private management of these public fora for expression creates a situation in which expression is privately policed and promoted based on the terms of service and algorithms set by the administrators of Social Media platforms. Regardless of the specificities of the situation, Twitter unilaterally banned Donald Trump from their platform permanently while he was the sitting President of the United States without any significant resistance. As most prominent Social Media firms are based in the United States, Section 230 of the Communications Decency Act has given these firms the power to fully self-regulate the content on their platforms, creating a situation wherein American laws governed Social Media platforms globally. In an attempt to regain regulatory control over Social Media platforms, the European Union will imminently pass the Digital Services Act. The act is a key development to track, as it could completely change the regulatory landscape of expression on Social Media, at least in the European Union.

Section 2: Fake News

In this Section, we will discuss the dynamics of fake news spreading on Social Media platforms, the subsequent risks posed by fake news on Human Rights, and limitations to speech aimed at eliminating these harmful effects. A basic working definition of fake news is false information misrepresented as factual information. The Section will add nuance to this definition and present a history of fake news. This thesis will move away from using the term ‘fake news’ after Section 2.2, which will also define the more apt terms ‘disinformation’ and ‘information disorder,’ whose specificity is better suited to explain the phenomenon. Section 2.2 will also describe some of the main motives behind the dissemination of disinformation on Social Media platforms. The final subsection will demonstrate the potential threats posed to human rights by disinformation on Social Media platforms and by regulations that purportedly protect against these threats.

2.1: A Brief History of ‘Fake News’ – What Questions does it Raise?

The misrepresentation of false information as true information, more commonly known as Fake News, is not unique to Social Media or a new phenomenon at all.²⁵ We are all accustomed to hearing outlandish or untrue claims, some more believable or consequential than others: A famous urban legend claims that humans eat eight spiders per year in their sleep, but Snopes.com recently admitted to completely fabricating this legend as a commentary on the gullibility of people believing logically implausible information that they read online.²⁶ Politics is no exception: False information that is disseminated for political ends can be far more dangerous and hateful than other types of disinformation. For example, propaganda created during World War II by any party to the conflict is often filled with hateful lies and stereotypes about the party’s enemies to boost their war efforts.²⁷ The history of the term ‘fake news’ is inextricable from the events of the 2016 United States Presidential Election and the subsequent presidency of Donald Trump. Data from Google Trends shows that search interest for the term was relatively insignificant until the American Presidential Election in November 2016.²⁸ While the term was first used by Hillary Clinton, Donald Trump’s main opponent in 2016, Trump was the person who took the prominence of this phenomenon to victory in 2016 as he rode a near-constant media spotlight on him caused in part by his brash, outlandish, and often false proclamations that often infamously utilized what came to be dubbed ‘alternative facts,’ or white lies that put him in a good light or furthered his political goals and narratives. Notorious early

²⁵ Section 2.2 will spend some time nuancing and reframing the concept of Fake News. The term ‘fake news’ will only be used up until that point. The working definition this thesis will use until that point is false information misrepresented as true information.

²⁶ ‘Lisa Birgit Holst’ (*Snopes.com*) <<https://www.snopes.com/lisa-birgit-holst/>> accessed 15 May 2022.

²⁷ Pavia JR and Brack N, ‘Racism in Japanese and U.S. Wartime Propaganda’ (1994) 56 *The Historian* 671 <<https://www.jstor.org/stable/24449072>>

²⁸ This data from Google Trends can be found at this link: <https://trends.google.com/trends/explore?date=all&q=fake%20news>

fake news stories from this period mostly sought to promote the presidential candidacy of Donald Trump and denigrate his opponents:²⁹. Three days before the 2016 United States Presidential Election, a now-defunct fake news outlet called the Denver Guardian posted a story on their website that made the rounds on Social Media titled “FBI Agent Suspected In Hillary Email Leaks Found Dead In Apparent Murder-Suicide.”³⁰ Another notorious fake news narrative, now known commonly as ‘Pizzagate,’ alleged that leaked emails from Hillary Clinton’s campaign staff contained encoded messages indicating that there was a child sex trafficking enterprise of some sort being operated out of the basement of a local Washington, DC family-friendly pizzeria, Comet Ping Pong. The fact of the matter is that there have been numerous instances in recent history in which highly credible allegations of child molestation, facilitated by sophisticated sex trafficking organizations, have been made against politicians and other high-profile members of society.³¹ That being said, the belief that Comet Ping Pong was the base for another such organization was very misguided. The shop does not even have a basement, let alone moonlight as an underground facility for sex trafficking. Another favorite practice of Donald Trump was to level information or outlets that did not favor him as ‘fake news.’ In this way, the term fake news is not solely used to describe verifiably false information but also as an ideological weapon to advance favorable narratives without any factual basis. Habgood-Coote explains that this imprecise use of language, based on subjective perceptions, can legitimate hateful or anti-democratic propaganda.³² This contest over the nature of the truth has led to a widespread lack of trust in government and between peers in society. van Prooijen et al. found that there is a widespread distrust of political institutions, which in turn leads people to believe conspiracy theories about these institutions.³³ Distrust of institutions is especially prevalent in the case of the United States. The widespread self-reported feeling of dissatisfaction with the way things are going with the state of the country shows a bleak picture of public trust in government.³⁴ While the 2016 United States Presidential Election saw fake news gain

²⁹ Guess A, Nagler J and Tucker J, ‘Less than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook’ (2019) 5 *Science Advances* eaau4586 <<https://www.science.org/doi/10.1126/sciadv.aau4586>> accessed 12 May 2022

³⁰ Lubbers E, ‘There Is No Such Thing as the Denver Guardian, despite That Facebook Post You Saw’ (The Denver Post, 6 November 2016) <<https://www.denverpost.com/2016/11/05/there-is-no-such-thing-as-the-denver-guardian/>> accessed 22 May 2022

³¹ A) Bryant, Nicholas A. *The Franklin Scandal: A Story of Powerbrokers, Child Abuse and Betrayal*. First edition (revised for softcover), Trine Day, 2012.

B) Araoz, Jennifer. “Opinion | Jeffrey Epstein Raped Me When I Was 15.” *The New York Times*, 14 Aug. 2019. *NYTimes.com*, <https://www.nytimes.com/2019/08/14/opinion/jeffrey-epstein-jennifer-araoz.html>.

C) Winter, Dan Mangan, Tom. “Ghislaine Maxwell Trial: Sex Crimes Accuser Testifies Jeffrey Epstein Took Her to Meet Trump When She Was 14.” *CNBC*, 1 Dec. 2021, <https://www.cbc.com/2021/12/01/ghislaine-maxwell-trial-jeffrey-epstein-took-teen-girl-to-meet-trump-accuser-say-s.html>.

³² Habgood-Coote J, ‘Stop Talking about Fake News!’ (2019) 62 *Inquiry* 1033 <<https://www.tandfonline.com/doi/full/10.1080/0020174X.2018.1508363>> accessed 29 June 2022

³³ van Prooijen J-W, Spadaro G and Wang H, ‘Suspicion of Institutions: How Distrust and Conspiracy Theories Deteriorate Social Relationships’ (2022) 43 *Current Opinion in Psychology* 65 <<https://www.sciencedirect.com/science/article/pii/S2352250X21000828>> accessed 20 May 2022

³⁴ ‘Satisfaction With the United States’ (Gallup 2022) <<https://news.gallup.com/poll/1669/General-Mood-Country.aspx>> accessed 24 May 2022

widespread prominence, the United States Capitol Riot in Washington, DC on January 6th, 2020 is evidence of the influential power of fake news, especially when fake news is spread from some of the world's most famous or followed people. The fact that many political institutions and the individuals that make up those bodies often do illegal, unethical, or outright corrupt things to achieve their goals only exacerbates this problem, as it can act as a sort of confirmation bias for those who already believe conspiracy theories. In the case of the Washington, DC Riots on January 6th, many of those who rioted believed that the election had been unfairly stolen from Donald Trump, despite a complete dearth of evidence to prove that claim. However, this is another instance like the Pizzagate saga in which mistrust of institutions from the past informs future events: The Brooks Brothers' Riot of the 2000 United States Presidential Election was an instance wherein election results were unfairly influenced after the fact by partisan operatives. While Guess et al. found that Americans self-identifying as conservative were more likely to share fake news articles in 2016, fake news is not just a problem that is facing those who support or are affiliated with Donald Trump, especially after the conclusion of the 2016 American Presidential Election.³⁵ Guess et al. found that older individuals are much more likely to share fake news than younger generations who have incorporated Social Media into their lives from a younger age. The high-profile allegations that Donald Trump's campaign colluded with Russia were exposed as fabrication with the indictment of a significant source in the Steele Dossier (the original BuzzFeed report that broke the 'news' on the allegations against Trump) for making false statements to the United States Federal Bureau of Investigations.³⁶ Fake news is not solely an American problem, either. A survey of people in all 27 European Union Member States found that 70 percent fully agreed or tended to agree that they often came across news and information that misrepresented reality or was false.³⁷ An article from Horbyk et al. expresses grave concern over regimes that meet the criteria for Davide Banis' idea of a fictiocracy, a political regime that disregards the distinction between fact and fiction. Horbyk et al. identify disturbing trends in the Philippines, Ukraine, Russia, and Brazil.³⁸ While it is difficult to measure the volume or frequency of actual fake news stories, one survey found that 85 percent of respondents in Brazil were concerned about fake news in general.³⁹ Hungary is a notable example within the European Union, and there is currently a developing situation that has seen Australian Prime Minister Scott Morrison's close ties to a prominent QAnon conspiracy theorist

³⁵ Ibid 28

³⁶ The indictment for making false statements to the FBI of Igor Y. Danchenko, a critical source for the Steele Dossier, can be found here: <https://www.justice.gov/sco/press-release/file/1446386/download> 'Trump-Russia Steele Dossier Analyst Charged with Lying to FBI' *BBC News* (5 November 2021) <<https://www.bbc.com/news/world-us-canada-59168626>> accessed 26 May 2022

³⁷ Watson A, 'Fake News Consumption Frequency in the EU by Country 2022' (*Statista*, 12 May 2022) <<https://www.statista.com/statistics/1076701/fake-news-frequency-europe/>> accessed 7 June 2022

³⁸ A) Horbyk R and others, 'Fake News as Meta-Mimesis: Imitative Genres and Storytelling in the Philippines, Brazil, Russia and Ukraine' <<https://aaltodoc.aalto.fi:443/handle/123456789/111278>> accessed 17 May 2022

³⁹ McCarthy N, 'Where Concern Is Highest About Fake News On The Internet [Infographic]' (*Forbes*) <<https://www.forbes.com/sites/niallmccarthy/2019/06/12/where-concern-is-highest-about-fake-news-on-the-internet-infographic/>> accessed 7 June 2022

potentially influence a speech he gave.⁴⁰ Fake news is widespread online in many different contexts across the globe, especially after the 2016 American Presidential Election. Still, the actual definition and nature of fake news is a vague and often contested concept in discourse. The increased focus on fake news and the problems it causes in the content of the widespread adoption of Social Media and other new communications technologies has raised many philosophical questions. As mentioned before, the label of fake news is frequently instrumentalized to discredit information for personal, political, or financial reasons. The overuse of the term fake news points to a fundamental disconnect between different segments of society on the actual nature of the truth in many situations. Much of this could come from differences in background, upbringing, education, etc. However, the general distrust many individuals have for institutions reported by van Prooijen et al. and Gallup poll results undoubtedly harms social cohesion and the ability to perform collective action by peddling distorted versions of reality.

2.2: So What is ‘Fake News’?

While the misrepresentation of the veracity of information through baseless allegations, rumors, conspiracy theories, and plain old white lies has existed since the beginning of human communication, there has been a newfound concern over the societal impacts of ‘Fake News.’ The term ‘Fake News’ has become a central focal point in societal discourse in approximately the last half-decade due to the increasing virality and subsequent prominence of false information misrepresented as accurate information on Social Media and eventually entering political and social discussions. However, the term itself has been plagued by overly aggressive usage that has blurred and diluted the actual meaning of the term. Many, most notoriously former President of the United States Donald Trump, frequently brand information that does not correspond to their worldview or contradicts their positions as such. The newfound ease of access to a wide, democratized variety of information outlets has exposed readers' inability to discern the veracity and potentially biased presentation of information. A report written by Wardle and Derakhshan and published by the Council of Europe echoed Habgood-Coote, characterizing the term ‘fake news’ as a vague and ineffective descriptor of the phenomenon of the misrepresentation of the integrity of information online. They propose that this phenomenon should instead be understood as an ‘information disorder’ wherein people are constantly bombarded with an incomprehensible deluge of information with widely variable levels of honesty that can instantaneously travel from any corner of the globe. Wardle and Derakhshan propose useful categorizations for the three types of harmful or false information

⁴⁰ Karp P, ‘Scott Morrison Says He Doesn’t Back QAnon and Refuses to “Cancel” Supporter Friend’ *The Guardian* (21 June 2021)

<<https://www.theguardian.com/australia-news/2021/jun/21/scott-morrison-says-he-doesnt-back-qanon-and-refuses-to-cancel-supporter-friend>> accessed 1 July 2022

that is distributed in public spheres like Social Media or traditional media to create this phenomenon:⁴¹

- **Disinformation** is when false information is knowingly shared to cause harm
- **Misinformation** is when incorrect information is shared, but no harm is meant
- **Malinformation** is when genuine information is shared to cause harm, often by moving information designed to stay private into the public sphere

They also emphasize the importance of the elements of information disorder. The elements of information disorder are the agent and their motivations behind the production and distribution of the instance of information disorder, the format and characteristics of the content, and the possible interpretations of said content by readers. Generally, the concept presented by Wardle and Derakhshan adds a much-needed clarification of important distinctions often overlooked in discussions of fake news. Wardle and Derakhshan suggest a much better way to contextualize the phenomenon of the massive spread of false information than the term fake news, a term whose meaning has been lost due to increasingly vague and politicized usage. Before the age of information disorder brought on by the communications paradigm shift toward Social Media, the news cycles that people experienced gave people a very different overall experience with the media. On top of the participatory relationship users have with content on Social Media, first of all, the pure amount of content on Social Media platforms creates a frenetic, moment-to-moment news cycle.

In contrast, the traditional media that preceded Social Media usually had twelve to twenty-four-hour news cycles. Horbyk et al. also describe a philosophical shift in the news set in motion by the Ronald Reagan administration in the United States, which blurred the line between information and narrative.⁴² These blurred lines are exacerbated by the fact that the democratization of news outlets creates many different, competing, and often contradictory narratives. This democratization effect has made a cultural arrangement in which there is no longer a unifying narrative binding individuals together into a society. This is not to say that this

⁴¹ From this point forward, the term ‘disinformation’ will be used instead of the term ‘fake news.’ The definition of disinformation presented by Wardle and Derakhshan will be the working definition for disinformation for the rest of this thesis.

Wardle and Derakhshan also present useful typology of disinformation and misinformation that adds further nuance to this definition: Satire/parody, which does not intend to harm but could confuse some readers; Misleading content, which misleads readers with an intentionally misleading framing or bias; Imposter content, or content that attempts to impersonate trusted sources; Fabricated content, which is completely false, designed to harm, and usually novel; Manipulated content, when genuine information or imagery is manipulated to deceive readers; False content, which clouds the truth with false contextualization; And finally false connection content, which uses false or sensational headlines with content that do not support the claims made in the headline, visuals, or captions.

Wardle C and Hossein D, ‘Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making’ (*Council of Europe Publishing*) <<https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>>. Accessed 26 May 2022.

⁴² Ibid 37A.

type of unification is always a good thing, as there are countless examples of nationalistic, discriminatory, or anti-democratic narratives uniting a populace. For example, the claims that Iraq had manufactured weapons of mass destruction that sought to justify the American-led invasion in the early aughts (before the communications paradigm shifted) had no factual basis. Still, many of the largest television and print news media outlets heavily pushed them to advance the political and economic interests of those in positions of capital ownership and power with enormous consequences. Despite the lack of societal unity, these narratives still exist in public discourse. There are myriad examples of adherents to narratives that create far-reaching communities around these narratives. Different examples of these communities often become insular echo chambers that believe easily debunkable stories that confirm their chosen narratives and discredit information that contradicts their narrative no matter how convincing the evidence may be. QAnon conspiracy theories and their far-fetched, unsupported, and easily debunkable 'Q Proofs,' all of which will be discussed in greater detail during section 3, perfectly illustrate this trend. The tremendous amount of competition within the news/media space due to the democratization effects of Social Media also necessitate that outlets and narratives have to one-up each other to escalate the perceived stakes of their narratives continually, or they will fall behind others in a sensationalism race to the bottom.

QAnon describes a 'shadow war' between the forces of 'good and evil,' the good being Donald Trump and the QAnon believers and the bad being a 'cabal' of international powerbrokers that commit pedophilia and massacre children on an unfathomably large scale. After the right to abortion was struck down in the United States by the Supreme Court, prominent leaders in the Democratic Party sent fundraising emails soliciting donations from supporters with extremely alarmist language. Understandably, this attracted widespread criticism for failing to uphold abortion rights despite being the party ostensibly in power, then profiting from the catastrophe.⁴³ A common narrative pushed by Hungarian Prime Minister Viktor Orban and Fox News host Tucker Carlson, among other white supremacists, is the 'Great Replacement Theory,' which makes fascist-style claims that White people in the western world are facing a genocidal crisis due to the influx of nonwhites through immigration. Despite the obvious logical shortcomings of this xenophobic narrative, multiple mass shooters (Utoya 2011, Pittsburgh 2018, El Paso 2019, Buffalo 2022, to name some high-profile examples)* have cited this narrative as a

⁴³ Gravel Institute, 'Nancy Pelosi just sent out her first email fundraising off the decision overturning Roe v. Wade.' (Twitter, 24 June 2022) <<https://twitter.com/GravelInstitute/status/1540347369111392256/photo/1>> accessed 13 July 2022

'Democrats Pull in \$80 Million Following Supreme Court Ruling on Roe' (*PBS NewsHour*, 2 July 2022) <<https://www.pbs.org/newshour/politics/democrats-pull-in-80-million-following-supreme-court-ruling-on-roe>> accessed 14 July 2022

motive for the atrocities they committed.⁴⁴ This trend will be discussed in more detail in Section 3. Information disorder ultimately erodes individuals' trust in each other on a societal level, as people are left questioning: “Does my neighbor believe in the same reality as I do?”

The most problematic type of false information is disinformation: Rapp and Salovich find that users often experience confusion about the nature of the truth, doubt of accurate understandings, and reliance on falsehoods when exposed to false information.⁴⁵ While all of these types of false or harmful instances of information disorder could be highly problematic when shared publicly, attempts to regulate mis- or mal-information could prove overzealous and generally more harmful than helpful to the protection of human rights. For example, overzealous punishment of false or harmful information that is not disinformation could discourage whistleblowing on illegal or unethical activities, an example of mal-information. Suppose a news broadcaster unintentionally publishes misinformation by committing a few minor errors while reporting a story. In that case, it seems sensible that a retraction or correction from the broadcaster is preferable to legal intervention with news reporting. Disinformation is inherently dangerous because the average individual is categorically awful at determining the truthfulness of news, not due to intellectual shortcomings but a deep dependence on heuristics instead of critical analysis to interpret the information they come across. The examples of disinformation presented in Section 2.1 (Pizzagate, January 6th riots in Washington, DC) help us distill some of the key ingredients of successful disinformation narratives: They capitalize on the cynicism of readers cultivated by a deep distrust of institutions by distributing a story that is just outrageous enough to attract attention without immediate skepticism and mixing true statements with false statements to increase the believability of the story. Vosoughi et al. found that the allure of novel information is also powerful: Readers want to feel like they are making a discovery and they are ‘in the know.’⁴⁶ Horbyk et al. describe the current wave of disinformation as a cocktail of journalism, fiction, and satire that creates a hollow imitation of normal news manufactured with the specific purpose of subverting news

⁴⁴ A) Ling J, ‘How 4chan’s Toxic Culture Helped Radicalize Buffalo Shooting Suspect’ *The Guardian* (18 May 2022) <<https://www.theguardian.com/us-news/2022/may/18/4chan-radicalize-buffalo-shooting-white-supremacy>> accessed 14 July 2022

B) Gabbatt A, ‘Fox News Suddenly Goes Quiet on “Great Replacement” Theory after Buffalo Shooting’ *The Guardian* (17 May 2022) <<https://www.theguardian.com/us-news/2022/may/17/buffalo-shooting-fox-news-tucker-carlson-great-replacement-theory>> accessed 14 July 2022

* The names of these shooters will not be included if possible, following guidance from scholars and activists on reducing the contagion effect as much as possible. This topic will be discussed in Section 3.2.2. More information on this topic can be found at this source:

Lankford A and Madfis E, ‘Don’t Name Them, Don’t Show Them, But Report Everything Else: A Pragmatic Proposal for Denying Mass Killers the Attention They Seek and Deterring Future Offenders’ (2018) 62 *American Behavioral Scientist* 260 <<http://journals.sagepub.com/doi/10.1177/0002764217730854>> accessed 14 July 2022

⁴⁵ Rapp DN and Salovich NA, ‘Can’t We Just Disregard Fake News? The Consequences of Exposure to Inaccurate Information’ (2018) 5 *Policy Insights from the Behavioral and Brain Sciences* 232 <<http://journals.sagepub.com/doi/10.1177/2372732218785193>> accessed 1 July 2022

⁴⁶ Vosoughi S, Roy D and Aral S, ‘The Spread of True and False News Online’ (2018) 359 *Science* 1146 <<https://www.science.org/doi/10.1126/science.aap9559>> accessed 2 July 2022

media as a whole.⁴⁷ This model of disinformation is highly successful: Despite 84 percent of Americans saying they were at least somewhat confident in their ability to detect fake news in 2016, Moravec et al. concluded that only 17 percent of Social Media users could identify fake news more than half of the time. Users depend on confirmation bias to shape their perceptions so intensely to the point that they believe disinformation despite inoculation techniques that indicate that the content contains false information.⁴⁸ While partisanship certainly influenced some users to believe disinformation they encountered online, it was less influential than other heuristics users employed to determine the veracity of information. Pennycook and Rand found that the most likely reason a user would inaccurately determine the veracity of news is due to a lack of information and a failure to reason around that information gap due to relying on the illusory truth effect and the third-person effect. The third-person effect is a heuristic that describes the phenomenon where ‘credible’ people, especially cultural elites or those otherwise attracting large Social Media followings significant on Social Media, have not only a much larger reach but potential for influence over those who trust them. The third-person effect is especially significant on Social Media due to the ability of any user to share endorsements of specific content or worldviews. Information shared by someone like Donald Trump, who had 88.7 million Twitter followers before his ban in January 2020, was much more likely to spread widely and subsequently be believed by his supporters. However, Pennycook and Rand also found that most people do not necessarily share content they care about on Social Media and that only a fraction of online content draws enough attention to be shared on Social Media.⁴⁹ This finding supports some earlier points made by this thesis that most people are reluctant to present themselves as overly political on Social Media platforms for fear that this type of posting may be seen as obnoxious by peers.⁵⁰ People do not usually share disinformation narratives: Chengcheng Shao et al. found that Social Media accounts that share fake news are significantly more likely to be used by bots than human users. Shao et al. also found that bot accounts are critical in the earlier stages of spreading disinformation content, focusing on targeting large, famous, and influential users to amplify the reach of the content through the third-person effect.

⁴⁷ Ibid 37A.

⁴⁸ Psychological inoculation is a concept that posits that users would be less likely to believe disinformation if they are forewarned about potential disinformation and engage with diluted versions of those narratives.*

Moravec P, Minas R and Dennis AR, ‘Fake News on Social Media: People Believe What They Want to Believe When It Makes No Sense at All’ (Social Science Research Network 2018) SSRN Scholarly Paper ID 3269541 <<https://papers.ssrn.com/abstract=3269541>> accessed 12 May 2022

* A more recent study showed that the effectiveness of psychological inoculation in helping users identify disinformation may be increasing.

Lewandowsky S and van der Linden S, ‘Countering Misinformation and Fake News Through Inoculation and Prebunking’ (2021) 32 *European Review of Social Psychology* 348

<<https://www.tandfonline.com/doi/full/10.1080/10463283.2021.1876983>> accessed 1 July 2022

⁴⁹ The Illusory Truth Effect is a phenomenon whereby prior exposure to a disinformation narrative increases later belief in the same claim, regardless of the implausibility of the claim itself. The Authors cite the example that in World War II, rumors that were more familiar or widespread were more likely to be believed. The Authors also find that the Illusory Truth Effect often trumps political partisanship.

Pennycook G and Rand DG, ‘The Psychology of Fake News’ (2021) 25 *Trends in Cognitive Sciences* 388 <<https://www.sciencedirect.com/science/article/pii/S1364661321000516>> accessed 12 May 2022

⁵⁰ Ibid 19.

Furthermore, bots heavily support successful instances or sources of disinformation content, which contributes to the spread of existing fake news narratives to an ever-increasing audience of Social Media users.⁵¹ Vosoughi et al. found that disinformation content tended to spread faster and further than legitimate news on Twitter.⁵² Once fake news is shared by non-bot users, though, it tends to be for partisan reasons: Osmundsen et al. find that the sharing of fake news usually shares the same motivations that drive other partisan behavior.⁵³ Partisanship is a major factor in the sharing of disinformation because of the prevalence of echo chambers and siloed spaces online as users self-determine the content and users they interact with.

So, what are the common motives behind disinformation creation, distribution, and promotion on Social Media platforms? The framework of information disorder presented by Wardle and Derakhshan considers the motive for publishing the story in question to be as crucial as the veracity of the information posted to its classification within the framework. Some creators of disinformation are motivated by material financial gain: Hughes and Waismel-Manor found that a large portion of the disinformation shared on Facebook that specifically targeted American conservatives in the 2016 American Presidential Election came from a town called Veles in Macedonia. These creators were motivated solely by financial factors, as some were quoted by the BBC explicitly remarking that they did not care about the results of the 2016 American Presidential Election and that they would generate much more income than they would by working in other ways. The BBC also reported that these creators of disinformation in Veles often paid Facebook to promote their fake news stories in their algorithmic recommendation systems.⁵⁴ Hughes and Waismel-Manor also corroborated that the common motive for creating and spreading disinformation in Veles was purely for profit. There aren't any obvious or poorly-hidden ties to Russian state-backed bodies or other malicious third parties. They found that the disinformation coming from Veles was essentially already present in the American political discourse: Most of the content produced in Veles was cut-and-pasted stories and conspiracy theories from American radical right-wing websites. Once these stories were taken from their original place of posting, they were repackaged to appear as if they were published by a trustworthy news website (such as the example of the Denver Guardian from Section 2.1). Once repackaged, these stories were distributed to a broader consumer base that does not usually frequent these extremist sites yet had a large amount of demand for pro-Trump, anti-Clinton content leading up to the 2016 election, generating massive sums of money through

⁵¹ This issue will be discussed in more detail in later sections.

Shao C and others, 'The Spread of Low-Credibility Content by Social Bots' (2018) 9 *Nature Communications* 4787 <<http://arxiv.org/abs/1707.07592>> accessed 3 June 2022

⁵² *Ibid* 41.

⁵³ Osmundsen M and others, 'Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter' (2021) 115 *American Political Science Review* 999 <https://www.cambridge.org/core/product/identifier/S0003055421000290/type/journal_article> accessed 1 July 2022

⁵⁴ Kirby EJ, 'The City Getting Rich from Fake News' *BBC News* (5 December 2016) <<https://www.bbc.com/news/magazine-38168281>> accessed 23 May 2022

advertisement revenue for the producers of disinformation in Veles.⁵⁵ However, research shows significant evidence of non-financial motives for creating fake news.

A literature review of fake news research, mainly from the field of journalism, found that one of the most commonly established motives among studies of disinformation was agenda-setting.⁵⁶ This can be a much more insidious form of disinformation than we see from disinformation with the goal of short-term financial gain. In some cases, instances of malinformation could also be instrumentalized for agenda-setting purposes. Examples of situations that produce this kind of disinformation include half-truths or lies of omission, data manipulation, sponsored content or native advertisements, deep fakes, astroturfing, and motivated reasoning. These situations are challenging to detect and counteract as these situations manipulate rather than fabricate the truth. This means that unlike disinformation, such as the type of content produced in Veles in 2016, which is usually relatively easy to spot once vetted and debunk, this type of information requires individuals to assess the nuanced motivations of the source of information critically. On a theoretical level, partial control over the political agenda or perceived reality of one or many individuals is a potent tool. Park and de Zúñiga find that the news media has a significant role in creating a political learning environment.⁵⁷ The media has substantial agenda-setting power, which can manipulate segments of the populace into radical actions like the January 6th, 2020 riots in Washington, DC. The agenda-setting power of Social Media platforms will be discussed in detail in Section 3.2. This leads us to the implications fake news has on human rights. The fine lines between legitimate expression that must be guaranteed and spreading disinformation have not yet been authoritatively defined, creating a situation where an appropriate balance between protecting freedom of expression and the rights threatened by disinformation must be achieved.

2.3: The Threats Disinformation Currently Poses to Human Rights and Regulatory Solutions

The rampant spread of disinformation on Social Media platforms mainly threatens individuals' rights to health, life, non-discrimination, private life, and a free and fair election. This subsection will begin by briefly touching on the rights to non-discrimination and private life. This subsection will continue by discussing the rights to free and fair elections, health, and life by first comparing national and regional approaches to policing discrimination on Social Media platforms before discussing derogations to human rights in emergencies. This subsection also begins to explain some downstream effects of online disinformation.

⁵⁵ Hughes HC and Waismel-Manor I, 'The Macedonian Fake News Industry and the 2016 US Election' (2021) 54 PS: Political Science & Politics 19 <<https://www.cambridge.org/core/journals/ps-political-science-and-politics/article/macedonian-fake-news-industry-and-the-2016-us-election/79F67A4F23148D230F120A3BD7E3384F>> accessed 7 June 2022

⁵⁶ Abu Arqoub O and others, 'Mapping the Scholarship of Fake News Research: A Systematic Review' (2022) 16 Journalism Practice 56 <<https://doi.org/10.1080/17512786.2020.1805791>> accessed 14 May 2022

⁵⁷ Ibid 18.

2.3.1: A Brief Discussion of the Impact of Social Media on the Right to Non-Discrimination and the Right to Private Life

Firstly, a quick note on the right to non-discrimination: Some examples of disinformation with heavy consequences for the right to non-discrimination, like the white supremacist ‘Great Replacement theory,’ have been mentioned. Of course, these have tremendous and disproportionate implications for marginalized groups who are often deliberately targeted by those who have been radicalized by malicious disinformation online. These risks will be discussed in more detail in Section 3.

Secondly, on the right to private life: Some specific instances of Fake News could be said to violate the right to private life, which guarantees individuals protection from attacks on their honor and reputation. Existing defamation legislation seems to adequately protect against violations of the right to private life. For example, the aforementioned Alex Jones was sued by Chobani (a large American food company known to employ refugees) for disseminating xenophobic disinformation about the company and its owner. The case was resolved after Jones issued a statement during a subsequent broadcast unreservedly retracting the comments, apologizing to the offended parties, and deleting any content that contained those narratives.⁵⁸ While this may seem like a light punishment for actions that could have caused Chobani to suffer significant financial damages, the practice of retracting and correcting disinformation is undoubtedly a suitable recourse for attacks on honor or reputation that violate the right to private life that has been a long-standing principle of journalistic ethics. Another example of this type of recourse through retraction could be seen in the aftermath of the 2020 American Presidential Election. Conservative American news outlets like Fox News and Newsmax discredit the integrity of voting machines supplied by Dominion Voting Systems and Smartmatic. Both Fox News and Newsmax seemed to issue retractions to their claims against these firms implicitly, but Dominion Voting Systems’ 1.6 Billion USD defamation lawsuit against Fox News is still pending.⁵⁹ Cyberbullying is also a problem that has suitable avenues for recourse through anti-harassment legislation. While disinformation that attacks the reputation of individuals or groups in a defamatory way can be harmful, there is generally a legal consensus on the issue, with defamation and anti-harassment laws and suits common in many legal systems.

⁵⁸ Hauser C, ‘Alex Jones Retracts Chobani Claims to Resolve Lawsuit’ *The New York Times* (17 May 2017) <<https://www.nytimes.com/2017/05/17/us/alex-jones-chobani-lawsuit.html>> accessed 1 July 2022

⁵⁹ A) McEvoy J, ‘Fox News, Newsmax Walk Back Election Fraud Claims After Voting Machine Manufacturer Threatens Legal Action’ (*Forbes*) <<https://www.forbes.com/sites/jemimamcevoy/2020/12/21/fox-news-newsmax-walk-back-election-fraud-claims-after-voting-machine-manufacturer-threatens-legal-action/>> accessed 1 July 2022

B) Stempel J, ‘Fox News Parent Must Face Defamation Lawsuit over Election Coverage’ *Reuters* (22 June 2022) <<https://www.reuters.com/world/us/fox-news-parent-must-face-defamation-lawsuit-over-election-coverage-2022-06-21/>> accessed 1 July 2022

2.3.2: A Comparative Examination of Legal Frameworks in the United States, China, and the European Union

This subsection investigates the differences in the regulations broadly governing the dissemination of disinformation in the United States, China, and the European Union to analyze their impacts on the balance of the human rights to free expression and free and fair elections. These illustrative examples have been chosen for two reasons: Each model has immense geopolitical importance, and an overall comparison of these three legal approaches showcases the breadth of the spectrum of legal anti-disinformation approaches. As a short preview, the United States excessively defends the right to freedom of expression, China undermines both human rights through high degrees of censorship, and the European Union seems to be looking to establish a middle path that strikes the appropriate balance through desperately-needed reforms.

The societal role Social Media platforms have developed as an accessible global public sphere for political speech, organizing, and action creates spaces that can often become highly political. Article 21 of the Universal Declaration of Human Rights enshrines the right of subjects of a state to govern directly or freely choose representatives. This includes the right to a free press and the right to informed choices based on quality information.⁶⁰ While this can be utilized to protect and promote human rights, the general distrust of institutions and information disorder users experience on Social Media platforms can easily blur the lines between fact, half-truth, and fiction. The feeling of confusion this creates for users harms democracy. Donald Trump's unfounded claims of election fraud that were widely circulated on Twitter and other platforms in the 2020 US presidential election hurt the legitimacy of those election results to this day and led to instances of armed intimidation outside of local vote-counting facilities.⁶¹ These risks have sparked a debate on the necessity for legislation to regulate disinformation that could potentially affect political processes by creating information disorder.

In the United States, legislation has been proposed but seems to have stalled, leaving the principle enshrined in Section 230 of the Communications Decency Act that Social Media platforms are not responsible for content posted by users in place. Generally, legislative regulations on disinformation are a priority of the Democratic Party that fails to attract any support from the Republican Party. Senator Amy Klobuchar, a Democrat from Minnesota, proposed a bill known as the Honest Ads Act in late 2017 to increase transparency and accountability for paid political advertising on Social Media platforms. This bill was eventually included in the For the People Act, also known as H.R.1, a bill with a much larger scope that

⁶⁰ United Nations Human Rights Office of the High Commissioner, *Human Rights and Elections – A Handbook on International Human Rights Standards on Elections* (2021) <<https://www.ohchr.org/sites/default/files/Documents/Publications/Human-Rights-and-Elections.pdf>> accessed 27 June 2022

⁶¹ 'Increasingly Normal: Guns Seen Outside Vote-Counting Centres' <<https://www.aljazeera.com/news/2020/11/7/increasingly-normal-guns-seen-outside-vote-counting-centres>> accessed 1 July 2022

aimed to remove barriers to voting access, end unfair redistricting (also known as gerrymandering), and eliminate avenues for campaign finance corruption. However, this bill failed to pass in the Senate, leaving the United States with no legislation to protect Social Media users from disinformation. During the 2020 United States Presidential Election, Joe Biden criticized Facebook after Donald Trump's campaign ran advertisements that contained unsubstantiated claims about Biden's ties to Ukraine, vowing publicly that the platform would cease to be a conduit for disinformation and political distortion. Other high-ranking members of the Democratic Party, such as Elizabeth Warren, campaigned on their own proposals for laws that would penalize intentional distribution of disinformation as a civil or criminal offense. Aside from the political difficulties of enacting legislation aiming to regulate online disinformation, these general prospects for these laws are bleak from the perspective of American constitutional law. The American Supreme Court has repeatedly upheld the First Amendment of the Constitution of the United States, which prohibits legislation "abridging the freedom of speech." The Court has protected the right of the press to discuss governmental affairs in any circumstance in *Mills v. Alabama*, 384 U.S. 214 (1966), the right to receive political information without interference in *Lamont v. Postmaster General*, 381 U.S. 301 (1965), and the right to anonymity when distributing political materials or content in *McIntyre v. Ohio Elections Commission*, 514 U.S. 334 (1995). Most significantly, the courts found in *United States v. Alvarez*, 567 U.S. 709 (2012) that falseness alone was not grounds to restrict expression and expressed the unlikelihood that speech could ever be restricted based on its content in *United States v. Playboy Entertainment Group, Inc.*, 529 U.S. 803 (2000). Generally, the United States prioritizes the protection of free expression over the protection of rights that could be harmed by said free expression. The United States does not even define hate speech, protecting many instances of it as rightful expression, and only outlaws direct incitements of violence.⁶² As mentioned in Section 1, the United States legal system has been the legal framework that Social Media platforms abide by. This caused the American prioritization of free expression to dominate the diverse needs for regulating disinformation in other parts of the world. It also disincentivizes strict moderation of the content users post to a firm's platform, as they will not be held accountable for the post, and moderating the platform could be unprofitable or too small of a return on investment. Subsequently, some states and international collectives of states have taken measures to exert control over online information flows.

While the American emphasis on free expression lives on one side of the spectrum, the Chinese focus on maintaining societal order lives on the other end of the spectrum. Social Media platforms in China are prohibited from distributing news content that is not state-sponsored, and the distribution of 'rumors' (the Chinese Government's preferred term for disinformation) carries heavy criminal penalties.⁶³ Many of the largest tech companies and

⁶² 'Hate Speech and Hate Crime' (*Advocacy, Legislation & Issues*, 12 December 2017) <<https://www.ala.org/advocacy/intfreedom/hate>> accessed 2 July 2022

⁶³ Repnikova M, 'China's Lessons for Fighting Fake News' (*Foreign Policy*) <<https://foreignpolicy.com/2018/09/06/chinas-lessons-for-fighting-fake-news/>> accessed 1 July 2022

Social Media platforms are not easily accessible in China, leading to the creation of specific government-approved platforms for the Chinese market. Weibo, a Chinese microblogging platform that has been compared to Twitter, had 582 million monthly active users and 252 million daily active users as of the first quarter of 2022.⁶⁴ China notoriously stringently monitors and very frequently censors content that does not support the regime's preferred narratives and advance their goals.⁶⁵ The regime's monitoring and censorship efforts are also highly forward-thinking, with an emphasis on the censorship of multimedia over plain text posts.⁶⁶ Most cases of censorship on Chinese Social Media platforms describe a complete white-out of search results (blank search result pages) for topics that, considering their prominence in societal discourse, should at least show some results.⁶⁷ The strict censorship on these platforms exemplifies the clear violations of the right to free and fair elections and free expression. The press is severely curtailed by this law and other authoritarian measures that target expression and the press in China. These measures also do nothing to promote the truth and reduce harms caused by disinformation; instead, the distribution of public information, the ability to determine the societal perception of reality, and agenda-setting power have all been centralized within the regime. The Chinese Communist Party micromanages public information, content, and narratives to counteract these democratization effects. The censorship of online expression in China is so extensive and exacting to the point that pieces of online content are frequently censored despite being considered to be completely innocuous and suitable for children of any age in most other contexts: Another example of centralized control of Social Media platforms in China is the trend of censoring animated characters, most notoriously Winnie the Pooh due to the emergence of a viral meme comparing Chinese President Xi Jinping to the beloved cartoon toy bear. Many of these examples of banned animated content were censored for reasons such as taking market share away from domestically produced animated content in the case of *SpongeBob SquarePants* or promoting nihilism and a 'slacker' lifestyle in the case of *Peppa Pig* and Netflix's *Bojack Horseman*.⁶⁸ This is a case in which the scales have been tipped too far in

⁶⁴ Thomala LL, 'Weibo Corporation: DAUs 2022' (Statista)

<<https://www.statista.com/statistics/1058070/china-sina-weibo-dau/>> accessed 13 July 2022

⁶⁵ Meadows CZ, Tang L and Zou W, 'Managing Government Legitimacy during the COVID-19 Pandemic in China: A Semantic Network Analysis of State-Run Media Sina Weibo Posts' (2022) 15 *Chinese Journal of Communication* 156 <<https://doi.org/10.1080/17544750.2021.2016876>> accessed 13 July 2022

⁶⁶ Liu J and Zhao J, 'More than Plain Text: Censorship Deletion in the Chinese Social Media' (2021) 72 *Journal of the Association for Information Science and Technology* 18 <<https://onlinelibrary.wiley.com/doi/10.1002/asi.24390>> accessed 11 July 2022

⁶⁷ Yip M, 'Netizens Debate Bo Xilai despite Censorship' *BBC News* (11 April 2012)

<<https://www.bbc.com/news/world-asia-china-17673673>> accessed 11 July 2022

⁶⁸ Memes are online content usually in the form of an image. Memes are usually meant for comedic purposes, but can also be used for hateful reasons. 4chan, a message board notorious for complete user anonymity and extremely hateful content, is one of the early adopters of memes as community-building tools. Online political action vs. real world political action → simulacrum that can spill into reality

A) James L, 'Seven Cartoon Characters Peppa Pig Joined on China's Banned List' (South China Morning Post, 3 May 2018)

<<https://www.scmp.com/culture/film-tv/article/2144561/after-peppa-pigs-ban-china-here-are-eight-other-cartoons-and>> accessed 5 July 2022

favor of restricting speech and controlling media, justified by a fabricated threat of societal harm. In stark contrast to the legal framework for disinformation in the United States, which is intentionally vague about classifying speech to ensure it is protected to the highest degree possible, the Chinese system's use of highly ambiguous parameters for illegal content is instrumentalized to provide the highest degree of control over expression possible. The judiciary in China itself is not independent, showing that there are no consistent principles of law regulating disinformation to uphold, but rather censorship is committed without the possibility of legal checks on the regime's power.⁶⁹ Ultimately, this authoritarian treatment violates the right to free expression of all of the Chinese government's subjects.

The European Union's Digital Services Act attempts to find the regulatory balance somewhere between the American and Chinese philosophies that will best protect both the freedom of expression and rights that could be violated by online disinformation. In the European Union, the Digital Services Act attempts to eliminate the risk of violations by increasing the control the public can wield over Social Media platforms. Before the Digital Services Act, some European Union member states had laws regulating disinformation. These laws mostly attempted to combat the risks posed by disinformation, but they were deeply flawed. This was due to the difficulty of defining disinformation and the parameters for false declarations that could actually be outlawed or potentially overzealous regulation that criminalized disinformation. Fathaigh et al. review the different regulations on disinformation across the eleven European Union Member States that have enacted such legislation.⁷⁰ For reference, the current policy definition of disinformation in the European Union is set out by the European Commission, whose 2018 Communication on tackling online disinformation states that

B) Haas B, 'China Bans Winnie the Pooh Film after Comparisons to President Xi' *The Guardian* (7 August 2018) <<https://www.theguardian.com/world/2018/aug/07/china-bans-winnie-the-pooh-film-to-stop-comparisons-to-president-xi>> accessed 6 July 2022

⁶⁹ Zhu M, 'Jury Still out on China's Legal Reform and Judicial Independence' (*South China Morning Post*, 3 April 2021) <<https://www.scmp.com/economy/article/3128422/jury-still-out-chinas-legal-reform-and-judicial-independence>> accessed 14 July 2022

⁷⁰ According to the authors, all of the Member States' definitions find common ground in ruling that disinformation is "(a) false information, (b) disseminated with a specific intention (malicious or bad faith) (c) and causes certain harms." However, definitions diverged on the overall scope, addressee, and legal sanctions, as well as specific harms that were mentioned and the culpability of the creator and/or the disseminator of disinformation. Some of these regulations prescribe lengthy prison sentences, ranging from 30 days to five years. Lithuania is the sole European Union Member State with a law explicitly defining and outlawing disinformation. Their definition aligns with Wardle and Derakhshan's, but is limited to harm done to a specific individual. Of the other ten Member States, the authors create three groups of member states in question based on the context of the regulation in question: "[Member states with] false news or false information laws contained in criminal legislation and non-criminal legislation; false news laws enacted during the Covid-19 pandemic; and laws targeting false news and false information during elections." According to the authors, the ten member states other than Lithuania that have laws regulating disinformation are: Austria, Croatia, Cyprus, the Czech Republic, France, Greece, Hungary, Malta, Romania, and Slovakia.

Fathaigh R, Helberger N and Appelman N, 'The Perils of Legally Defining Disinformation' (2021) 10 *Internet Policy Review* <<https://policyreview.info/articles/analysis/perils-legally-defining-disinformation>> accessed 4 July 2022

disinformation is “verifiably false or misleading information that is created, presented and disseminated for economic gain or to deceive the public intentionally, and may cause public harm.”⁷¹ Within the European Union, the Digital Services Act is a much-needed replacement for the preceding flagship legal regulation, the Directive on Electronic Commerce. This directive was issued in the year 2000 and was mainly geared towards basic rules for online service providers and commercial communications. The European Union has attempted to address disinformation online since the issuance of the Directive on Electronic Commerce in the form of the Code of Practice on Disinformation. Still, the impact of these methods was negligible as these guidelines were not legally binding and therefore created huge implementation gaps. Notably, the Directive includes the Internal Market Clause, which establishes the principle of the application of the law of the country an internet service provider is in.⁷² The Digital Services Act, as mentioned before, allows the European Union to regulate the activities of Social Media platforms within its borders, regardless of where the company is based. EU DisinfoLab, an independent organization combating the harmful effects of disinformation in the European Union, called the adoption of the Digital Services Act the opening of the ‘era of accountability’ for online service providers.⁷³ Specific provisions of the Digital Services Act designed to protect individuals from the harmful effects of disinformation include establishing consistent mechanisms for the removal of illegal content, a higher emphasis on user reporting, the right of recipients of services to seek compensation for damages, increased transparency for the algorithms that these platforms depend on, mandatory risk assessments, the collaboration of online service providers with experts to reduce risks, and equal punishment for media outlets that spread disinformation. The act does not intend for any governmental body to take over moderation of Social Media platforms fully but to change the current paradigm where Social Media platforms self-regulate into one of co-regulation. Despite the private owners of Social media platforms retaining some control over expression on their platforms, this establishment of co-regulation is extremely impactful when put in conversation with the newfound transparency of the algorithms that moderate Social Media platforms. This allows the European Union to identify harmful functions of the algorithm; for example, the processes termed ‘pipelines’ that gradually radicalize users can be identified and eliminated through co-regulation without any loss of expertise in operating or amending the algorithm.⁷⁴ An important distinction is the one

⁷¹ This definition can be found in *Section 2.1: Scope of Online Disinformation* of the Communication. Within this definition, public harm is understood as “threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens' health, the environment or security.” The Communication also states that this definition of disinformation excludes “reporting errors, satire and parody, or clearly identified partisan news and commentary.”

⁷² ‘Tackling Online Disinformation: A European Approach’ (European Commission, 26 April 2018) <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236>>

⁷³ ‘E-Commerce Directive | Shaping Europe’s Digital Future’ <<https://digital-strategy.ec.europa.eu/en/policies/e-commerce-directive>> accessed 6 July 2022

⁷⁴ ‘Tackling Disinformation Online: The Digital Services Act Opens the Era of Accountability’ (*EU DisinfoLab*) <<https://www.disinfo.eu/advocacy/tackling-disinformation-online-the-digital-services-act-opens-the-era-of-accountability/>> accessed 10 July 2022

⁷⁴ The ‘pipelines’ to extremism Social Media platforms often create will be the topic of Section 3.

made between harmful content, like disinformation, and illegal content, such as child pornography, hate speech, or incitements to violence. The main purpose of the Digital Services Act is to reduce the amount of illegal content. This distinction is crucial to note as, similarly to the situation in the United States; disinformation is not generally illegal in the European Union. For example, many examples of satire could be interpreted as disinformation. The aforementioned Snopes.com claim that humans eat spiders in their sleep on a semi-regular basis is an example of protected expression despite its absolute lack of basis in fact. The main way in which the Digital Services Act counteracts disinformation is by issuing a reinforced update on the Codes of Practice on Disinformation.⁷⁵ The impact of the Digital Services Act on the European Union's law and policy is perhaps most important in that it lays the groundwork for further measures to reduce the harmful effects of disinformation. Another main stated goal of the Digital Services Act is to produce a sort of legal harmony across all European Union member states on policing disinformation. Interestingly, Fathaigh et al. observe that the Digital Services Act seems to target procedural harmony rather than holistic legal harmony by enforcing the relevant national regulations on illegal content, which may or may not include disinformation.⁷⁶ It also signals that the European Union is hesitant, at this point, to take a stance on the way disinformation should be legally defined and punished at the regional level. The emphasis on risk assessments and data collection in collaboration with experts suggests that the European Union has yet to commit to a specific way to legally prevent disinformation without violating the right to free expression. The Digital Services Act continues the European trend of vaguely defining disinformation. This is not a crucial oversight, as criminal prosecution for disseminating disinformation is a difficult, convoluted, and potentially dangerous path to protect individuals from its harmful effects. Instead, the emphasis on non-regulatory solutions is quite forward-thinking and could have much greater societal benefits overall while avoiding potential consequences for the overall state of human rights through regulation that goes too far.

One example of nonregulatory solutions discussed in the European Union is media literacy training for Social Media users. Media literacy initiatives are extremely promising: While legal prosecution for the dissemination of disinformation is usually an *a posteriori* way to protect the population, media literacy is a strategy that aims to protect the population *a priori* that can also foster resiliency to disinformation, especially among those most initially vulnerable to disinformation.⁷⁷ The inoculation strategy of combating disinformation discussed in the previous subsection could be seen as similar to media literacy training as they are both forms of pre-emptive intervention against the persuasive power of disinformation.

Munn L, 'Alt-Right Pipeline: Individual Journeys to Extremism Online' [2019] First Monday <<https://firstmonday.org/ojs/index.php/fm/article/view/10108>> accessed 18 July 2022

⁷⁵ 'Disinformation: Commission Welcomes the New Stronger and More Comprehensive Code of Practice on Disinformation' *European Commission* (Brussels, 16 June 2022) <https://ec.europa.eu/commission/presscorner/detail/en/IP_22_3664> accessed 14 July 2022

⁷⁶ Ibid 68.

⁷⁷ Moore RC and Hancock JT, 'A Digital Media Literacy Intervention for Older Adults Improves Resilience to Fake News' (2022) 12 *Scientific Reports* 6008 <<https://www.nature.com/articles/s41598-022-08437-0>> accessed 13 July 2022

While inoculation strategies were shown to have mixed results in the previous subsection, media literacy initiatives could also help to overcome the confirmation bias that often thwarts inoculation attempts. The training for a Social Media user to be more resilient against disinformation is crucial as it provides individuals with skills that will protect them against many if not most instances of disinformation, while strategies like inoculation and debunking address one false narrative at a time. This type of training could have the double benefits of de-escalating online political discussions and fostering healthy political engagement while helping to avoid the legal headaches and potential human rights violations that overzealous regulations may spawn. Media literacy education, inoculation, and other similar, extralegal measures attempt to address the root causes of the widespread information disorder Social Media users face, especially when used in concert to give Social Media users the skills and tools necessary to avoid falling into the traps of disinformation. Another critical avenue in which to address the root causes of disinformation is to strengthen faith in public institutions, the lack of which is another contributing factor to information disorder. Hitlin and Shutava indicate that the main contributing factors to public trust in government institutions is the actual competence in terms of quality of governance as well as effectively communicating with the public.⁷⁸ Governments should also change corrupt, conspiratorial, or dishonest practices (such as the dissemination of lies about Weapons of Mass Destruction by the United States in the early aughts) that could add to the information disorder many users encounter on Social Media. The holistic approach of the Digital Services Act to Disinformation regulation is promising and could soon prove to strike an appropriate balance.

The Digital Services Act seems to signal that the European Union no longer considers Social Media platforms as products but as partially public spaces that should be held somewhat accountable to the public. However, it should not be mistaken as a silver bullet that will stop the harmful effects of disinformation in the European Union but instead as a much-needed legislative update that will lay the groundwork for future regulations and ensure procedural harmony across the Union.

2.3.3: Derogations of Human Rights in Emergency Situations – The Human Rights to Health and Life – COVID-19 and the Russian Invasion of Ukraine

This section will address the threats posed to the Human Rights of Life and Health by disinformation on Social Media.

The modern Social Media ecosystem raises a pressing question through disinformation that exacerbates or obfuscates problems caused by emergencies such as the ongoing Russian Hybrid War against Ukraine and the COVID-19 pandemic that threaten human rights on a massive scale at the regional and global levels. As the stakes are so high in these

⁷⁸ Hitlin P and Shutava N, ‘Trust in Government: A Close Look at Public Perceptions of the Federal Government and Its Employees’ (Partnership for Public Service and Freedman Consulting, LLC 2022) <<https://ourpublicservice.org/wp-content/uploads/2022/03/Trust-in-Government.pdf>> accessed 15 July 2022

situations, the question is whether or not the balance of protecting free expression and other rights should correspondingly shift. It is a general principle enshrined in foundational international human rights documents like the Universal Declaration of Human Rights (Article 29.2) that states may temporarily limit or derogate human rights to protect other rights in war or other critical emergencies threatening society depending on their national regulations. For example, if there is a natural disaster imminent, it is undoubtedly appropriate to temporarily derogate the rights to private life and a good standard of living in an evacuation situation. The COVID-19 pandemic and the Russian invasion of Ukraine fit into this criteria.

This principle has been applied across the globe to stop the spread of the COVID-19 virus. People worldwide suffer human rights violations to lower overall transmission and protect the most vulnerable during the COVID-19 pandemic. Due to not only the threat to self but also the unavoidable threats the COVID-19 virus poses to the health and lives of others due to the virus's ability to rapidly spread through the air, public health measures such as mandatory stay-at-home lockdowns, travel restrictions, mask mandates, etc. all fall into this type of necessary actions to safeguard human rights that also unfortunately temporarily restrict other rights. Therefore, in the case of COVID-19, restrictions aimed at protecting the lives and health of all in society and the proper functioning of medical services are necessary to respond to the pandemic. However, these measures faced quite a bit of resistance in different contexts across the globe. Historically, disinformation about health and medicine is rampant on Social Media platforms: Przemyslaw M. Waszak et al. found that in 2018, before the pandemic, there was a great deal of disinformation in online health news. Their results showed that 40% of the top links relating to common diseases contained false information.⁷⁹ This problem has only deepened since the COVID-19 pandemic began. The WHO has described the specific instances of information disorder relating to the pandemic as an 'Infodemic,' and Naeem and Bhatti found that large swaths of Social Media users were exposed to disinformation related to the pandemic.⁸⁰ Prominent disinformation narratives relating to the COVID-19 pandemic advocated against public health measures such as stay-in-place lockdowns and mask mandates by fomenting anti-vaccine sentiments. The Center for Countering Digital Hate, a nonprofit operating in the United States and the United Kingdom, shockingly found that up to 65% of disinformation related to the COVID-19 vaccines distributed on Facebook, Instagram, and Twitter was shared by just twelve extremely influential 'influencers,' or users with a vast reach on Social Media platforms. Despite the volume of disinformation they spread, and the repeated violations of the terms of service of the platforms they use, nine of these influencers have access to all three

⁷⁹ Waszak PM, Kasprzycka-Waszak W and Kubanek A, 'The Spread of Medical Fake News in Social Media – The Pilot Quantitative Study' (2018) 7 Health Policy and Technology 115
<<https://www.sciencedirect.com/science/article/pii/S2211883718300881>> accessed 16 July 2022

⁸⁰ A) Bradd S, 'Infodemic' (*World Health Organization*, n.d.) <<https://www.who.int/health-topics/infodemic>> accessed 18 July 2022

B) Naeem SB and Bhatti R, 'The Covid-19 "Infodemic": A New Front for Information Professionals' (2020) 37 Health Information and Libraries Journal 233 <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7323420/>> accessed 18 July 2022

platforms, with a total following of 59.2 million users at the time the report was published. The report criticizes Facebook, Twitter, and Instagram for their continued role in enabling these influencers and recommends deplatforming these individuals and organizations related to them, disallowing ‘secret’ groups within the platform that echo these sentiments, more transparent and better-enforced moderation standards, inoculation, and further data collection.⁸¹ Generally, these campaigns seem to be motivated for purely selfish reasons, namely building online audiences to monetize. Furthermore, many populist leaders such as Donald Trump and Jair Bolsonaro politicized the pandemic and pushed narratives that worked against many of the anti-pandemic and pro-vaccine public health measures established during the crisis. Interestingly, Jair Bolsonaro and Donald Trump disseminated disinformation and other rhetoric that undermined public health measures in very similar capacities: The main goal seemed to be to deflect blame for the severity of the crisis and to redefine the crisis from a public health emergency to an unwanted constraint on the economy.⁸²

In the case of the United States, this politicization amplified existing structural factors to cause higher death rates related to the COVID-19 virus in counties with a majority of their vote share going to the Republican Party. Donald Trump had a powerful and dangerous influential effect over significant segments of the population, pushing unverifiable and false narratives such as the narrative that ivermectin, a horse deworming medication, hydroxychloroquine, a malaria drug, and a diluted bleach solution could all provide cures or treatments for COVID-19.⁸³ This led to a report published by the United States Center for Disease Control, which found that an alarmingly large percentage of respondents acted on Donald Trump’s unfounded and uninformed speculation, with 39% of respondents intentionally engaging in ‘high-risk practices’ that the Center for Disease Control finds are harmful. 4% of people even drank or gargled diluted bleach solutions.⁸⁴ The researcher found that this was due to “voting behavior act[ing] as a proxy for compliance with and support for public health

⁸¹ According to the report, the individuals in the Disinformation Dozen are: Joseph Mercola, Robert F. Kennedy, Jr., Ty and Charlene Bollinger (Operating accounts together), Sherri Tenpenny, Rizza Islam, Rashid Buttar, Erin Elizabeth, Sayer Ji, Kelly Brogan, Christiane Northrup, Ben Tapper, and Kevin Jenkins.

Ahmed I, ‘The Disinformation Dozen: Why Platforms Must Act On Twelve Leading Online Anti-Vaxxers’ (Center for Countering Digital Hate 2021)
<<https://counterhate.com/wp-content/uploads/2022/05/210324-The-Disinformation-Dozen.pdf>> accessed 18 July 2022

⁸² Béland D and others, ‘Trump, Bolsonaro, and the Framing of the COVID-19 Crisis: How Political Institutions Shaped Presidential Strategies’ (2021) 184 *World Affairs* 413
<<http://journals.sagepub.com/doi/10.1177/00438200211053314>> accessed 16 July 2022

⁸³ A) McGraw M and Sam S, ‘It’s Been Exactly One Year since Trump Suggested Injecting Bleach. We’ve Never Been the Same.’ (*Politico*, 23 April 2021)

<<https://www.politico.com/news/2021/04/23/trump-bleach-one-year-484399>> accessed 18 July 2022

B) Bump P, ‘Analysis | Ivermectin Is the Signature Example of Politics Trumping Health’ *Washington Post* (31 March 2022)

<<https://www.washingtonpost.com/politics/2022/03/31/ivermectin-is-signature-example-politics-trumping-health/>> accessed 18 July 2022

⁸⁴ Radhika Gharpure and others, ‘Knowledge and Practices Regarding Safe Household Cleaning and Disinfection for COVID-19 Prevention — United States, May 2020’ (United States Center for Disease Control 2020)

<<https://www.cdc.gov/mmwr/volumes/69/wr/pdfs/mm6923e2-H.pdf>>

measures, vaccine uptake, and the likelihood of engaging in riskier behaviors (for example, unmasked social events and in-person dining) that could affect disease spread and mortality." Crucially, the researcher found that this relationship between policy and vulnerability created a spatial effect rather than an individual one, meaning that the likelihood of death is not related to the actual political leanings of the individual but to the spatial context of public health which the individual is subject to.⁸⁵ These results were replicated very similarly in Brazil, while analysis of death rates in Hungary saw spatial effects at play as well, disproportionately impacting communities that had the largest share of Roma people in the population. While the quality of public health enforcement in a spatial area is a concern, overly extensive regulation in the name of public health is also a problem: While the pandemic was politicized in Hungary, the way in which Viktor Orban capitalized on the politicization, passing measures that severely crippled the right to free and fair elections and freedom of expression, as centralized control over the media was strengthened for an indefinite period. The International Press Institute reported that a measure was passed that criminalizes the dissemination of ‘misinformation’ determined by central authorities to undermine anti-pandemic public health measures with up to five years of imprisonment. Orban has built a sizable media apparatus that promotes his policy agenda and preferred narratives, and Journalists that are critical of Viktor Orbán and his Fidesz party have frequently faced fines or imprisonment.⁸⁶ The International Press Institute also reported that 17 other nations capitalized on the need for derogations to human rights to combat the COVID-19 pandemic: While most of these regulations are genuinely meant to benefit public health initiatives by addressing disinformation that undermines anti-pandemic and pro-vaccine public health measures, there is significant potential for these censorship tools to be used against legitimate journalism.⁸⁷ These are clear examples in which the defense of the rights to life and health can be used to justify excessively oppressive legislation that exceeds the extent of the derogation mandated by the principle of proportionality. However, the fact that the COVID-19 virus spatially transmits harm means that some derogation of human rights to protect the rights of others is proportional.

The situation of disinformation in Ukraine shares many similar characteristics to aspects of the disinformation related to the COVID-19 pandemic: From a bird’s eye view, disinformation harms the human rights of Ukrainians, and derogations to human rights may be necessary to protect the human rights of those potentially in the way of harm in the conflict. However, once contextual details and nuances are considered, there are significant differences between the disinformation surrounding the illegal invasion of Ukraine from the disinformation

⁸⁵ Osman Segar N, ‘As Hospitals Restrict Visitors, What Constitutes A “Good Death”?’ (2022) 41 Health Affairs <<https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2021.01857>> accessed 16 July 2022

⁸⁶ ‘Hungary: Press Freedom Threatened as Orbán Handed New Powers’ *International Press Institute* (30 March 2020) <<https://ipi.media/hungary-press-freedom-threatened-as-orban-handed-new-powers/>> accessed 18 July 2022

⁸⁷ Wiseman Few J, ‘Rush to Pass “Fake News” Laws during Covid-19 Intensifying Global Media Freedom Challenges’ *International Press Institute* (3 October 2020) <<https://ipi.media/rush-to-pass-fake-news-laws-during-covid-19-intensifying-global-media-freedom-challenges/>> accessed 18 July 2022

relating to the COVID-19 pandemic. To properly contextualize the nature of disinformation in Ukraine, we must first briefly discuss the concept of hybrid warfare: Frank Hoffman of the Potomac Institute for Policy Studies proposed the idea of hybrid warfare in December 2007. Hoffman argues that combinational military threats employing some mixture of conventional, irregular, political, and terrorist tactics (if not all at once), will become much more prevalent.⁸⁸ However, since the publication of Hoffman's report, the term itself has suffered definition problems similar to the problems facing the term 'fake news': The combination of the aforementioned types of warfare is as old as warfare itself. Hybrid warfare can perhaps best be described in simple terms as the targeted use of all available tools a combatant has available to them, including the online dissemination of information and narratives about the conflict.⁸⁹ The ongoing conflict between Russia and Ukraine is undoubtedly a hybrid war in this manner, and the information battleground has emerged as a crucial front of the illegal invasion of Ukraine.

About seven and a half months before the invasion began, Vladimir Putin published an essay online in Russian, Ukrainian, and English that used baseless nationalist narratives that mythologized an exaggerated if not mostly fabricated shared history between the two nations.⁹⁰ Presciently, journalist Anne Applebaum characterized this essay as a call to arms a few weeks after it was published.⁹¹ Since the invasion began, the Russian government has systematically dismantled freedom of expression and independent information sources. Twitter, Facebook, and Instagram (used by nearly 40 million Russians) have all been blocked in Russia.⁹² These platforms have been targeted specifically for their role in the dissemination of information in Russia, according to an online statement from Roskomnadzor.⁹³ Independent Russian media outlets have been systematically shut down, resulting in a mass exodus of journalists who fear the exceedingly oppressive regime.⁹⁴ These efforts have affected the Russian public as well, with

⁸⁸ Frank Hoffman, 'Conflict in the 21st Century: The Rise of Hybrid Wars' (Potomac Institute for Policy Studies 2007) <<https://potomac institute.org/reports/19-reports/1163-conflict-in-the-21st-century-the-rise-of-hybrid-wars>> accessed 19 July 2022

⁸⁹ This thesis will focus exclusively on the impacts of the competing narratives that can be found on Social Media relating to the illegal invasion of Ukraine by Russian military forces.

Michael Kofman and Matthew Rojansky, 'A Closer Look at Russia's "Hybrid War"' (Wilson Center and Kennan Institute 2015) 7 <<https://www.files.ethz.ch/isn/190090/5-KENNAN%20CABLE-ROJANSKY%20KOFMAN.pdf>> accessed 19 July 2022

⁹⁰ A) Putin V, 'Article by Vladimir Putin "On the Historical Unity of Russians and Ukrainians"' (*President of Russia*, 12 July 2021) <<http://en.kremlin.ru/events/president/news/66181>> accessed 22 July 2022

B) Kartsev D, 'The Heirs of Ancient Rus Vladimir Putin Runs Roughshod over Ukrainian History in New Essay' (*Meduza*) <<https://meduza.io/en/feature/2021/07/12/the-heirs-of-ancient-rus>> accessed 22 July 2022

⁹¹ Applebaum A, 'Tweet from 3:38 PM, Jul 31, 2021' (*Twitter*) <<https://twitter.com/anneapplebaum/status/1421480489748254723>> accessed 22 July 2022

⁹² Vinokour M, 'Russia's Media Is Now Totally in Putin's Hands' (*Foreign Policy*) <<https://foreignpolicy.com/2022/04/05/russia-media-independence-putin/>> accessed 22 July 2022

⁹³ 'On Limiting Access to the Social Network Instagram' (*Federal Service for Supervision of Communications, Information Technology, and Mass Media*, 11 March 2022) <<https://rkn.gov.ru/news/rsoc/news74176.htm>> accessed 25 July 2022

⁹⁴ Sauer P and Michaelson R, "'It Was Game over": Russian Journalists Flee to Istanbul after Putin's Shutdown' *The Guardian* (18 March 2022)

videos of forced checks of private communications such as text messages of ordinary people on the streets in Moscow by police wearing some sort of riot gear being shared on telegram by journalists.⁹⁵ Traditional media infrastructure in Ukraine has been illegally attacked, as five non-combatant residents of Kyiv were killed in a bombing of the main television tower in the Ukrainian capital.⁹⁶ These are intentional efforts by the Russian authorities to restrict the flow of information about the situation on the ground in conflict zones in Ukraine to the rest of the world, near or far.

Some of the main false narratives pushed as part of the Russian invasion effort include the aforementioned claims to a shared history between Russia and Ukraine and allegations of genocide, ‘Nazification’ in Ukraine, the use of chemical weapons, and other atrocities being committed against people in Donbas and ethnic Russians living in Ukraine.⁹⁷ These efforts seem to mainly be aimed at controlling domestic narratives to maintain a semblance of public support for the invasion. This is a relatively complex task, partly due to the large number of Russians that have family residing in Ukraine. Ukrainian outlet Meduza describes the influential power of the pro-Kremlin, pro-invasion narratives that dominate the media in the occupied Luhansk and Donetsk oblasti: Unfortunately, the Meduza article reports that many Ukrainians with Russian relatives find that those relatives believe the false narratives peddled in Russia despite first-hand accounts or photos and videos taken during attacks from family. The article describes a relationship that has many characteristics of an abusive relationship, where influenced individuals forgive harmful Russian actions due to their constant minimization and justification by all of the available media outlets.⁹⁸ This generally supports the findings of Erlich and Garner, who find that Ukrainians with pro-Russian views are significantly more likely to believe false pro-Kremlin, pro-invasion narratives.⁹⁹ While Russia has significant manipulative control over domestic media and media in occupied regions of Ukraine, there are still instances of resistance and dissent against these narratives. Kizilova and Norris found that while approximately 60% of Russians support the invasion, these results may reflect “state

<<https://www.theguardian.com/world/2022/mar/18/dozhd-tv-staff-russian-channel-in-exile-istanbul>> accessed 22 July 2022

⁹⁵ Vasilyeva A, *Kommersant correspondent Anya Vasilyeva reports from Detsky Mir on Lubyanka Street*: (2022) <<https://t.me/Crexcrexcrex/3059>>

⁹⁶ ‘Ukraine: Kyiv TV Tower Struck by Russian Missile’ BBC News <<https://www.bbc.com/news/av/world-europe-60579847>> accessed 19 July 2022

⁹⁷ A) ‘Disinformation About Russia’s Invasion of Ukraine - Debunking Seven Myths Spread by Russia | EEAS Website’ <https://www.eeas.europa.eu/delegations/china/disinformation-about-russias-invasion-ukraine-debunking-seven-myths-spread-russia_en?s=166> accessed 22 July 2022

B) Mirsky A, ‘Dying to Kill The Russian Neo-Nazis Fighting Vladimir Putin’s War to “Denazify” Ukraine’ (*Meduza*) <<https://meduza.io/en/feature/2022/07/15/dying-to-kill>> accessed 22 July 2022

⁹⁸ Svetlova N, ‘“The TV Is Winning” Many Ukrainians Now Share a Common Experience: Their Relatives in Russia Refuse to Believe Their Accounts of the War.’ (*Meduza*) <<https://meduza.io/en/feature/2022/03/14/the-tv-is-winning>> accessed 22 July 2022

⁹⁹ Erlich A and Garner C, ‘Is Pro-Kremlin Disinformation Effective? Evidence from Ukraine’ [2021] *The International Journal of Press/Politics* 194016122110452 <<http://journals.sagepub.com/doi/10.1177/19401612211045221>> accessed 22 July 2022

censorship; self-censorship and response bias; the existence of protests even in the authoritarian Russian context; and the fact that some of the early polls were asking about a hypothetical invasion that many Russians might not have given much thought.”¹⁰⁰ Radio Free Europe reported that nearly five thousand people were unfairly imprisoned for protesting peacefully in Russia in the first two weeks after the invasion.¹⁰¹ While disinformation was constantly used by both the Soviet Union and United States during the Cold War, Russia took to their modern disinformation tactics of spreading disinformation on Social Media platforms and other alternative media in the aftermath of their 2008 invasion of Georgia. Their failures to disseminate pro-Kremlin narratives into mainstream media sources in this situation caused a pivot towards distributing their disinformation narratives on alternative information sources, including Social Media.¹⁰² Paul and Matthews described the Russian disinformation dissemination strategy as a ‘firehose of falsehood’ wherein the narratives that are pushed are far less important than ultimately fomenting extreme confusion and/or distrust of institutions in the targeted locale.¹⁰³ Outside of the disinformation campaigns in Russia and occupied portions of Ukraine, Europe, the United States, and the rest of the nations in the BRICS are emerging as the main foreign targets for Russian disinformation.¹⁰⁴ In Bulgaria, a traditional Russian ally that has subsequently become a NATO member state, David Klepper of the Associated Press reported that Russian authorities paid ‘influential’ Bulgarians 2000 Euros per month to post false pro-Kremlin, pro-invasion narratives on their Social Media accounts. In Poland, a hacker group aligned with the Kremlin spread false claims that Polish criminal gangs were harvesting the organs of Ukrainian refugees to stoke tensions and distrust between the two nations.¹⁰⁵ On the global level, the Twitter hashtag #IStandWithPutin, which trended for some amount of time, is an instance of a pro-Kremlin disinformation narrative that was astroturfed by automated accounts

¹⁰⁰ Kizilova K and Norris P, ‘Assessing Russian Public Opinion on the Ukraine War’ (2022) 281 *Russian Analytical Digest*

<<https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/RAD281.pdf>> accessed 26 July 2022

¹⁰¹ RFE/RL’s Kazakh Service, ‘Thousands Detained At Anti-War Protests Across Russia’

RadioFreeEurope/RadioLiberty (6 March 2022)

<<https://www.rferl.org/a/russia-1000-protesters-arrested-ukraine-invasion/31738786.html>>

¹⁰² Yablokov I, ‘Russian Disinformation Finds Fertile Ground in the West’ (2022) 6 *Nature Human Behaviour* 766

<<https://www.nature.com/articles/s41562-022-01399-3>> accessed 23 July 2022

¹⁰³ The four key characteristics of this ‘firehose’ model are: (1) the high volume, multi channeled distribution of disinformation that is (2) rapidly, repetitively, and constantly disseminated and (3) lacks commitment to objective reality or (4) consistency.

Paul C and Matthews M, ‘The Russian “Firehose of Falsehood” Propaganda Model Why It Might Work and Options to Counter It’ (RAND Corporation 2016) <<https://www.rand.org/pubs/perspectives/PE198.html>> accessed 23 July 2022

¹⁰⁴ BRICS is an acronym denoting five of the most prominent emerging national economies: Brazil, Russia, India, China, and South Africa.

Ibid 100.

¹⁰⁵ Klepper D, ‘Russia’s Information War Expands through Eastern Europe’ (*AP News*, 15 July 2022)

<<https://apnews.com/article/russia-ukraine-nato-bulgaria-misinformation-eastern-europe-0e87db7fe9263a465d6cf40d3287efe>> accessed 22 July 2022

(‘bots’) that were later removed from Twitter.¹⁰⁶ Russian disinformation narratives pushed in foreign markets try to justify Russia’s illegal actions in Ukraine by referencing past illegal military actions committed by the United States. This is a rhetorical technique known as ‘whataboutism,’ where the direct criticism of unethical or illegal actions is deflected by citing instances wherein others did similarly unethical or illegal acts.¹⁰⁷ Whataboutism can be appealing to those who wish to deflect blame, but under any scrutiny, the narratives do not hold up. While the instances cited in these narratives, such as the false narratives that Saddam Hussein had weapons of mass destruction to justify the invasion of Iraq in 2003, deserve condemnation, the Russian invasion of Ukraine is similarly illegal and should therefore be similarly condemned. Another tactic of Russian online subversion is trolling: Trolling usually sees users instigate anger and argument on divisive issues. In 2016, pro-Russian trolls on Twitter would attempt to instigate these divisive online spats in the United States by posting on hotly contested topics such as gun control, Black Lives Matter, and immigration.¹⁰⁸ While disinformation relating to COVID-19 is distributed for selfish reasons, such as personal gain or deflecting political responsibility, disinformation promoting violent pro-Kremlin, pro-invasion goals is one tactic used in the overall strategy of the illegal Russian invasion of Ukraine. Therefore, disinformation on the ongoing illegal invasion of Ukraine that harms the human rights of noncombatants living in areas threatened by war contributes to violating the right to life and private life. The European Union has already taken some steps to protect the human rights of those in Ukraine by sanctioning RT and Sputnik, another Russian state-controlled media outlet that was echoing false pro-invasion narratives, effectively banning these outlets from disseminating information in the European Union.¹⁰⁹ CNN reported that the American branch of RT, RT America, also shut down in early March. RT America was not shut down by legal action but because broadcasting providers terminated their contracts.¹¹⁰ The European Union’s sanction of RT and Sputnik is certainly a case of proportional censorship as the dissemination of false narratives that support the illegal Russian invasion of Ukraine is a central part of the Russian invasion strategy that is killing and displacing countless noncombatants that resided in Ukraine. However, this only

¹⁰⁶ When a hashtag is trending on Twitter, it indicates that there is a relatively huge amount of accounts tweeting that hashtag at a certain time.

Jones MO, ‘Analysis of #IStandWithPutin and #IStandWithRussia’ (Twitter, 3 March 2022)

<<https://twitter.com/marcowenjones/status/1499312091727020032>>

¹⁰⁷ Paz Perez E, *Strategic Disinformation: Russia, Ukraine, and Crisis Communication in the Digital Era* (ResearchGate 2022)

<https://www.researchgate.net/profile/Elisa-Paz/publication/362022036_Strategic_disinformation_Russia_Ukraine_and_crisis_communication_in_the_digital_era/links/62d187f2a6abd57c6ae78140/Strategic-disinformation-Russia-Ukraine-and-crisis-communication-in-the-digital-era.pdf>

¹⁰⁸ Abrams A, ‘Here’s What We Know So Far About Russia’s 2016 Meddling’ (*Time*, 18 April 2019)

<<https://time.com/5565991/russia-influence-2016-election/>> accessed 23 July 2022

¹⁰⁹ ‘EU Imposes Sanctions on State-Owned Outlets RT/Russia Today and Sputnik’s Broadcasting in the EU’ (2 March 2022)

<<https://www.consilium.europa.eu/en/press/press-releases/2022/03/02/eu-imposes-sanctions-on-state-owned-outlets-rt-russia-today-and-sputnik-s-broadcasting-in-the-eu/>> accessed 23 July 2022

¹¹⁰ Darcy O, ‘RT America Ceases Productions and Lays off Most of Its Staff’ (CNN, 4 March 2022)

<<https://www.cnn.com/2022/03/03/media/rt-america-layoffs/index.html>> accessed 23 July 2022

prevents these two outlets from spreading disinformation, which is problematic as many individuals also play roles in disseminating these false narratives.

Crucially, the Digital Services Act will allow the European Union to hold platforms that disseminate these false claims to be held accountable.¹¹¹ This is a sound approach to take, as manually censoring each instance of misinformation is an impossible task. Still, algorithmic changes that prevent posting, sharing, and amplifying harmful content are feasible and efficient ways to stop the disinformation before it reaches the public. In cases of disinformation related to the COVID-19 pandemic and the illegal Russian invasion of Ukraine, it may be necessary to tighten and censor some harmful information flows within the constraints of proportionality. Unfortunately, while human rights courts can play a role in enforcing the principle of proportionality on the restriction or censorship of online speech in states that belong to the Council of Europe, for example, the complete deregulation in places like the United States and the extreme censorship in places like China and Russia are detached from any legal processes to determine the proportionality of these derogations. This can also create risks for human rights, highlighting the need for proportional, balanced regulations on speech on Social Media platforms in emergency situations.

2: Conclusion

While disinformation has existed since the beginning of interpersonal communication, the massive capabilities and demand for information have created an unprecedented amount of disinformation that peddles many false narratives. Bad actors, from state-sponsored trolls and bots to individuals looking to profit off of the skepticism and alienation of others, take advantage of the average Social Media user's distrust of institutions and a high degree of reliance on heuristics such as confirmation bias, partisanship, the illusory truth effect, and the third-person effect for their own political or financial ends. The sheer volume of information and disinformation individuals consume on Social Media platforms creates an information disorder that obfuscates and distorts the objective truth, leading to risks for human rights in many cases. While appropriate legal resources already exist for potential violations stemming from claims of defamation and malinformation, the human rights to life, health, free and fair elections, and non-discrimination continue to be threatened by the dissemination of disinformation on Social Media. While there is a need to enact protections against these violations, any regulations on speech in the public sphere of Social Media must strike a delicate balance with the protection of freedom of expression, especially in cases where the press is targeted for regulation. A comparative review of the legal frameworks in place in the United States, China, and the European Union illustrates the effects different approaches to

¹¹¹ Johannes B and Manoury C, 'Digital Services Act: Commission Welcomes Political Agreement on Rules Ensuring a Safe and Accountable Online Environment' *European Commission - European Commission* (Brussels, 23 April 2022) <<https://ec.europa.eu/commission/presscorner/home/en>> accessed 23 July 2022

disinformation on Social Media platforms have: The United States treats Social Media platforms and the content they aggregate as private products that should not be interfered with while protecting the freedom of expression to the highest degree possible. This has resulted in the effective legalization of hate speech, the protection of Social Media platforms, and the dissemination of disinformation, regardless of the human rights risks. China treats Social Media platforms as a tool to exercise complete control over narratives within their state, leading to colossal oppression. The European Union recently updated an extremely obsolete framework with the intention of finding a middle ground that does not threaten free expression or allow the harm caused by the dissemination of disinformation on Social Media platforms. The Digital Services Act enables the legal authority of the European Union over the operations of Social Media platforms within the union, indicating that the European Union views Social Media platforms as public spheres that must be properly maintained to protect all relevant human rights. While this approach is very promising, the framework in place in the European Union is still in need of further additions to fully shield individuals from the harms of online disinformation. The role of disinformation in amplifying emergencies should also be combated: In two of the world's most pressing current crises, the COVID-19 pandemic and the illegal Russian invasion of Ukraine, disinformation clearly harms human rights, justifying proportional censorship in these cases.

The prevalence and effects of disinformation on Social Media platforms have raised some further issues. Primarily, the role Social Media platforms themselves play in the dissemination of disinformation and the subsequent human rights violations this produces must be investigated further. Another priority should be to undo the information disorder many individuals experience regularly. However, it is apparent from the content of this section that disinformation on Social Media platforms is the wellspring from which the vast majority of the other problems these platforms create are downstream.

Section 3: Online Extremism and the Role of Social Media Platforms in the Radicalization of Individuals

One of the most harmful consequences of disinformation disseminated on Social Media platforms is the systemic radicalization of users and the violent, hateful, and otherwise harmful acts radicalized individuals commit. In Section 2.3, there was a brief mention of the threats disinformation disseminated on Social Media platforms could pose to the right to non-discrimination, among others. Popular online disinformation narratives such as the white supremacist ‘Great Replacement theory’ have fostered hateful sentiments against vulnerable groups based on race/ethnicity, gender/sexuality, national origin, or other factors that often lead to hateful or violent acts. There are also important examples of political disinformation that lead to discriminatory violence against vulnerable groups. This thesis will discuss two main consequences of radicalization and extremism: Individual extremist acts, e.g., the person who committed the Christchurch, New Zealand mass shooting, and organized group extremism, e.g., the Rohingya genocide. This section will begin by defining radicalization and extremism and discuss the factors that contribute to radicalization. The section will go on to discuss how Social Media platforms create environments that accelerate and/or facilitate radicalization and extremism through an analysis of prominent examples of the two main types of consequences of radicalization and online extremism.

3.1: Definitions and Factors Contributing to Online Extremism and Radicalization

Like disinformation, extremism and radicalization have existed since the beginning of human society. Human nature innately searches for meaning and believable explanations for occurrences and trends that are hard to understand on the surface. For example, religion, ideology, and science all attempt to provide people with different lenses through which they can better make sense of the sometimes unexplainable things that all people experience. This aspect of the human condition can spawn extreme interpretations of the world that promote violence, hatred, or other harmful consequences. The Latin Church participated in a series of very bloody religious wars during the medieval period known as the Crusades, which generally aimed to bring Jerusalem and the surrounding area under their rule, and the Cold War spawned a series of proxy conflicts across the globe that continue to destabilize many regions into the present day. Science has also been used for harm in instances such as scientific racism, discrimination and imprisonment of people with disabilities, and anti-LGBTQ conversion therapy. As discussed in sections 1 and 2, the capabilities offered by Social Media platforms have caused a democratization of information that causes a societal information disorder and provides easier access to a broader array of fringe and extreme perspectives.

Similar to ‘fake news,’ ‘radicalization’ and ‘extremism’ have become terms that have suffered from definition problems due to increasing societal prominence and the subsequent vague and subjective usage. For the purposes of this thesis, radicalization is a phenomenon that generally describes the process an individual or group goes through to become more extreme, and extremism is the belief in oppositional attitudes that necessitate harmful acts against the group deemed the opposition.¹¹² Berger describes extremism as the “sweeping rationalization of violence” due to perceptions that the success or existential survival of the chosen ingroup necessitates harming or removing the scapegoated outgroup.¹¹³ Perhaps the most well-known example of this is the nazi ideology, which scapegoated Jews for all of the societal issues in post-World War I Germany, setting the stage for World War II and causing the Holocaust. Mandel compares radicalization and extremism to velocity and position, stating “...radicalization is the (positive) change in the degree of extremism expressed by an individual or group.” If the example of nazi ideology is used again, that would mean an individual or group becomes more radicalized the more they believe or echo nazi sentiments or narratives.

Vergani et al. reviewed 148 articles that discuss the factors and forms of individual radicalization across all contexts. They found that radicalized individuals or groups are either cognitively or behaviorally radicalized, and the main factors contributing to radicalization can generally be divided into push, pull, and personal factors. Cognitive radicalization is the acceptance of or support for extreme ideas and narratives. Behavioral radicalization is the commission of extreme acts, violence, terrorism, and joining an extremist group, e.g., ISIS or the proud boys. Push factors are generally structural, e.g., poverty or perceived persecution. Pull factors make extremism appealing to some, e.g., ideology and communal feelings. Personal factors are individual vulnerabilities to radicalization, e.g., personal experience, trauma, and psychological disorders. Vergani et al. stress the interrelatedness of the factors contributing to radicalization, and most cases of radicalization feature all three types of factors.¹¹⁴ Another critical factor that has recently emerged is also the degree of access to weaponry.¹¹⁵ In the case of individual online radicalization and extremism, instigating factors have already been discussed in detail, such as the overall digitization of

¹¹² Mandel DR, ‘Radicalization: What Does It Mean?’ [2009] Home-Grown Terrorism 101 <<https://ebooks.iospress.nl/doi/10.3233/978-1-60750-075-9-101>> accessed 24 July 2022

¹¹³ Another apt description of the oppositional attitude offered by Berger is an ‘us-versus-them’ attitude. A working definition of terrorism in this context is extremist action with a specific political, ideological, religious, or material goal. Non-terrorist extremism does exist, as there are many cases of extremist acts committed for purely personal reasons e.g. the Marjory Stoneman Douglas High School and recent Highland Park parade shooting. Berger JM, *Extremism* (The MIT Press 2018)

¹¹⁴ It should be noted that radicalization can also cause or trigger psychological disorders in and of itself. Vergani M and others, ‘The Three Ps of Radicalization: Push, Pull and Personal. A Systematic Scoping Review of the Scientific Evidence about Radicalization Into Violent Extremism’ (2020) 43 *Studies in Conflict & Terrorism* 854 <<https://www.tandfonline.com/doi/full/10.1080/1057610X.2018.1505686>> accessed 24 July 2022

¹¹⁵ A) Reeping PM and others, ‘State Gun Laws, Gun Ownership, and Mass Shootings in the US: Cross Sectional Time Series’ (2019) 364 *BMJ* 1542 <<https://www.bmj.com/content/364/bmj.1542>> accessed 24 July 2022
B) ‘Gun Violence’ (*Amnesty International*) <<https://www.amnesty.org/en/what-we-do/arms-control/gun-violence/>> accessed 24 July 2022

society, subsequent information disorder, and the role of Social Media platforms in that phenomenon. High levels of general dissatisfaction and distrust in institutions also create the conditions that increase radicalization. The Violence Project found that over 80% of individuals who commit mass shootings in the United States were in a “noticeable crisis” before the act.¹¹⁶ When the societal information disorder provides society’s most dysfunctional people with easy access to information that injects extreme distrust into their worldview, that can potentially radicalize them into insular online echo chambers that may encourage and eventually instigate extremist violence.

History shows that gender and general ideology are significant predictors of violent or offensive extremist acts. Statista data found that out of a sample of 132 mass shootings in the United States, 127 were committed by men, and two of the remaining five shootings were committed by a man and woman together.¹¹⁷ Interestingly, while this gender dynamic is replicated in Europe, it seems to be replicated at a less overwhelming rate, as about one in four or five extremists in the European Union is a woman.¹¹⁸ Extremist violence is also committed for reactionary ideologies, namely jihadist or far-right (usually white supremacist/nationalist/identitarian) belief systems. Extremist acts perpetrated for far-left ideologies are much rarer and less violent than acts of reactionary extremism.¹¹⁹ While many tend to distinguish far-right and jihadist ideologies, there is a false islamophobic distinction that attempts to signal that jihadist violence is committed for religious reasons while psychological disorders cause far-right violence. A report by Meleagrou-Hitchens, Crawford, and Wutke agreed with this view that both ideologies share violent reactionary perspectives, viewing any social or political progress as a threat to their ingroup privilege. They also shed some light on some reasons for the gender disparity in the commission of extreme acts as well as some of the fundamental contradictions inherent to these worldviews: The authors describe that these groups assume similar “chauvinistic and hyper-masculine collective identities” that glorify the ingroup and dehumanize outgroups by asserting that the outgroup is simultaneously inferior and an existential threat to the ingroup.¹²⁰ This inherent contradiction is a clear indication that radicalizing narratives depend on both a sense of personal insecurity and the confusion that information disorder causes.

¹¹⁶ The Violence Project, ‘Mass Shooting Data & Research’ (The Violence Project) <<https://www.theviolenceproject.org/>> accessed 24 July 2022

¹¹⁷ Statista Research Department, ‘Mass Shootings in the U.S.: Shooters by Gender 2022’ (*Statista*) <<https://www.statista.com/statistics/476445/mass-shootings-in-the-us-by-shooter-s-gender/>> accessed 24 July 2022

¹¹⁸ Orav A, Shreeves R and Radenovic A, ‘Radicalisation and Counter- Radicalisation: A Gender Perspective’ (European Parliamentary Research Service 2018) Members’ Research Service PE 581.955 <<https://www.europarl.europa.eu/EPRS/EPRS-Briefing-581955-Radicalisation-gender-perspective-rev-FINAL.pdf>>

¹¹⁹ Jasko K and others, ‘A Comparison of Political Violence by Left-Wing, Right-Wing, and Islamist Extremists in the United States and the World’ (2022) 119 Proceedings of the National Academy of Sciences e2122593119 <<https://pnas.org/doi/full/10.1073/pnas.2122593119>> accessed 24 July 2022

¹²⁰ Meleagrou-Hitchens A, Crawford B and Wutke V, ‘Rise of the Reactionaries: Comparing the Ideologies of Salafi-Jihadist and White Supremacist Extremism’ (National Counterterrorism, Innovation, and Education Center 2021) <<https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/Rise%20of%20the%20Reactionaries.pdf>>

3.1.1: QAnon and Other Conspiracy Theories

Often, the most harmful radicalizing narratives are deeply dependent on conspiracy theories, one of the most harmful forms of disinformation. The term ‘conspiracy theory’ suffers from definition problems like many of the other key terms in this thesis. It generally describes a theory that contradicts popular wisdom and alleges some sort of grand secret plot and coverup. Some conspiracy theories do turn out to be mostly, if not completely, true (e.g., MKULTRA, Iran-Contra), and many maintain that conspiracy theories play a role in governmental transparency.¹²¹ However, this thesis will use Wood et al.’s definition. They describe “a self-sustaining worldview comprised of a network of mutually supportive beliefs” that forms a “monological belief system.” They found that the belief in conspiracy theories is usually motivated by a general affinity for conspiracy theories themselves rather than any specific claims the conspiracy theory makes.¹²² DiMaggio finds that once again, rightward political attitudes are significantly predictive of belief in conspiracy theories.¹²³ As the rest of this subsection will explain, the outlandish ‘QAnon’ conspiracy theory is a great example of a harmful and radicalizing conspiracy theory. QAnon is a baseless narrative that mainly alleges that a secret global ‘cabal’ of pedophile elites systematically molests and murders children on a large scale for various far-fetched reasons. It has recently become one of the most prominent conspiracy theories on and off of Social Media platforms, especially in the United States. QAnon spawned out of the aforementioned ‘pizzagate’ theory in the early portion of Donald Trump’s presidency due to the cognitive dissonance of many Trump supporters that their candidate, who promised to improve their perceived conditions radically, did not solve the problems they continue to face. According to the theory, Donald Trump was fighting a secret battle with the ‘cabal,’ which allegedly included Hillary Clinton and many other political opponents while he was president. QAnon followers used this theory to justify their unchanging living conditions despite Trump's election. It is also an iteration of a ‘witch hunt’ or ‘satanic panic’ theory that attempts to grab as much attention as possible by alleging a massive scale of awful crimes committed by society’s most visible figures. While these general narrative themes exist, QAnon is an example of a ‘big tent’ conspiracy theory that is readily open to bringing in other conspiracy theorists through an attitude of ‘yes, and...’.¹²⁴ What is certain is that, as is the

¹²¹ A) Napolitano MG and Reuter K, ‘What Is a Conspiracy Theory?’ [2021] Erkenntnis <<https://doi.org/10.1007/s10670-021-00441-6>> accessed 24 July 2022

B) Linville, Tani. “Project MKULTRA and the Search for Mind Control: Clandestine Use of LSD Within the CIA.” *History Capstone Research Papers*, Apr. 2016, https://digitalcommons.cedarville.edu/history_capstones/6.

C) *Understanding the Iran-Contra Affairs*.

https://www.brown.edu/Research/Understanding_the_Iran_Contra_Affair/. Accessed 24 July 2022

¹²² Wood MJ, Douglas KM and Sutton RM, ‘Dead and Alive: Beliefs in Contradictory Conspiracy Theories’ (2012) *3 Social Psychological and Personality Science* 767 <<http://journals.sagepub.com/doi/10.1177/1948550611434786>> accessed 24 July 2022

¹²³ DiMaggio AR, ‘Conspiracy Theories and the Manufacture of Dissent: QAnon, the “Big Lie”, Covid-19, and the Rise of Rightwing Propaganda’ [2022] *Critical Sociology* 089692052110736

<<http://journals.sagepub.com/doi/10.1177/08969205211073669>> accessed 25 July 2022

¹²⁴ Zuckerman E, ‘QAnon and the Emergence of the Unreal’ [2019] *Journal of Design and Science* <<https://jods.mitpress.mit.edu/pub/tliexqdu>> accessed 24 July 2022

case with most conspiracy theories, it is based on deeply anti-semitic, homophobic, nationalistic, and otherwise discriminatory sentiments.¹²⁵ While the main claims as such are virtually baseless, QAnon depends on exploiting the cynicism of its followers by weaving a few strings of truth into its overall tapestry like many other disinformation narratives. Among many things, QAnon constantly berates its followers to categorically disbelieve mainstream media or evidence that disproves their claims, encouraging an attitude that the totality of information that does not confirm their views is a part of a grand charade designed to perpetuate the ‘cabal.’ This attitude has persisted despite the failure of nearly every prediction made by the proponents of the theory, which included narratives that there would be mass arrests of many prominent politicians and that Trump would maintain his presidential office despite losing the 2020 election.¹²⁶ This speaks to the fragility of both conspiracy narratives and their followers: Almost all verifiable information that does not expressly promote the conspiracy theory can collapse the entire narrative house of cards. Despite that, conspiracy theories often gain huge followings and instigate many extremist acts that are ready to dedicate their lives to the cause the conspiracy theory promotes. The complete distrust of mainstream media and a highly interactive online following of QAnon unmoors radicalized individuals from reality altogether. QAnon also showcases the rise of the interactive conspiracy theory. QAnon, for a multitude of factors alongside the interactive capacities Social Media offers, encourages followers to ‘do their own research’ to find the purportedly hidden truth. ‘Research’ as such means consuming expressly self-confirming content that grifters or proponents of the theory usually produce to further radicalize the audience. In reality, the constant assertion that the truth is never obvious and must be discovered gamifies radicalization.¹²⁷ They become deeply dependent on the deeply insular online echo chambers that are very hostile to outsiders and nonbelievers through which they consume QAnon-related content and interact with other followers.¹²⁸ Furthermore, these types of insular groups often substitute the level of radicalization for personal worth, idolizing those who are the most invested in the theory. The more ‘truth’ a follower ‘knows,’ the more righteous they are. This often results in job loss, family dysfunction, social ostracization, and obsessive behavior. Personal accounts compare the compulsive behavior of radicalized QAnon followers to cult members or addicts, completely unable to disengage these narratives to the detriment of themselves and those around them.¹²⁹ Despite the huge amount of false predictions that have

¹²⁵ Thompson J and Davis Thomander S, ‘What Drives Support for QAnon? Evidence from a Survey Experiment’ (University of Exeter and Stanford University 2021) <<https://files.osf.io/v1/resources/23qaj/providers/osfstorage/6181113dce8aee05592aee2?action=download&direct&version=1>>

¹²⁶ Tian E, ‘The QAnon Timeline: Four Years, 5,000 Drops and Countless Failed Prophecies’ (*bellingcat*, 29 January 2021) <<https://www.bellingcat.com/news/americas/2021/01/29/the-qanon-timeline/>> accessed 24 July 2022

¹²⁷ The persuasive appeal of information perceived as novel was discussed in Section 2.2. See footnotes 41 (Fabricated content) & 46.

¹²⁸ van Prooijen J-W, ‘Sometimes Inclusion Breeds Suspicion: Self-Uncertainty and Belongingness Predict Belief in Conspiracy Theories: Self-Uncertainty and Conspiracy Beliefs’ (2016) 46 *European Journal of Social Psychology* 267 <<https://onlinelibrary.wiley.com/doi/10.1002/ejsp.2157>> accessed 24 July 2022

¹²⁹ Jaffe G and Real JAD, ‘Life amid the Ruins of QAnon: “I Wanted My Family Back”’ (*Washington Post*) <<https://www.washingtonpost.com/nation/interactive/2021/conspiracy-theories-qanon-family-members/>> accessed 24 July 2022

been made by proponents and followers of the conspiracy theory, it has maintained its popularity. While it is inherently difficult to measure beliefs for conspiracy theories due to response bias, among other factors, recent studies show that anywhere from 10-20% of Americans believe the QAnon theory, and there is an emerging trend of right-wing politicians espousing QAnon narratives in states that are dominated by reactionary politics.¹³⁰ QAnon has also gained smaller levels of traction internationally.¹³¹ Donald Trump's presidency has come and gone, but the QAnon narrative continues to thrive, highlighting radicalization and conspiracy theories' addictive, hateful, and destructive allure.

3.2: The Role of Social Media Platforms in the Promotion of Online Extremism and Radicalization

Sections 1.4 and 2.2 discussed the historic dominance of the American legal framework, specifically Section 230 of the Communications Decency Act, for the regulation of expression on Social Media platforms. This necessitates the development and ownership of Social Media platforms by privately operated businesses that benefit their shareholders with little to no regard for the rest of their stakeholders.¹³² This has led the businesses operating Social Media platforms to optimize their products to maximize user engagement, as discussed in section 1.2. As with the past dominant avenues for media such as television and newspapers, Social Media platforms are essentially content aggregation platforms that depend almost solely on user engagement to generate revenue.¹³³ However, the differentiating factors between Social Media platforms and traditional media, such as the personalized stream of recommended content to each user have led Social Media platforms, especially the largest platforms, to serve and moderate content with algorithms. Social Media algorithms are software that is generally designed to aggregate a user's engagement history data (i.e. clicks, likes, shares, searches, etc.) to recommend content which the user is most likely to engage with.¹³⁴ This software is often

¹³⁰ A) 'New PRRI Report Reveals Nearly One in Five Americans and One in Four Republicans Still Believe in QAnon Conspiracy Theories' (*PRRI*)

<<https://www.prrri.org/press-release/new-prrri-report-reveals-nearly-one-in-five-americans-and-one-in-four-republicans-still-believe-in-qanon-conspiracy-theories/>> accessed 25 July 2022

B) Stapilus R, 'Stapilus: Idaho Voters Should Pay Attention to Candidates Espousing QAnon Beliefs' (*Idaho Press*) <https://www.idahopress.com/opinion/guest_opinions/stapilus-idaho-voters-should-pay-attention-to-candidates-espousing-qanon-beliefs/article_eccede09-0a7f-5373-a8b7-3168bf38becd.html> accessed 31 July 2022

¹³¹ Hoseini M and others, 'On the Globalization of the QAnon Conspiracy Theory Through Telegram' <<https://arxiv.org/abs/2105.13020>> accessed 25 July 2022

¹³² This is a simplified explanation of the most common motives and revenue streams of Social Media platforms. As discussed previously, Social Media platforms are operated in different ways, e.g. in places like China or Russia where governments are more authoritarian.

¹³³ The main revenue streams for most Social Media platforms is selling advertising, the value of which is directly tied to the volume and intensity of user engagement on the platform.

McFarlane G, 'How Facebook (Meta), Twitter, Social Media Make Money From You' (*Investopedia*) <<https://www.investopedia.com/stock-analysis/032114/how-facebook-twitter-social-media-make-money-you-twtr-lnkd-fb-goog.aspx>> accessed 25 July 2022

¹³⁴ O'Brien C, 'How Do Social Media Algorithms Work?' (*Digital Marketing Institute*) <<https://digitalmarketinginstitute.com/blog/how-do-social-media-algorithms-work>> accessed 25 July 2022

augmented with machine learning capabilities that aim to allow the software to optimize its own predictive and influential power.¹³⁵ The absolute focus on the maximization of engagement has harmful consequences. Studies have shown that Social Media use and engagement are both significantly correlated to depression, and Social Media overuse can increase polarization and result in addiction.¹³⁶ Leaks from internal Meta (Facebook's parent company) communications show that the company's management has been aware of its platform's intensely polarizing and depressing effects since at least 2016. Despite that, they proceeded not to address the mental health issues the platform may cause or fact-check any political advertising on the platform before the 2020 United States presidential election.¹³⁷ However, as a private business beholden to its shareholders, Meta and all other owners of Social Media platforms have no intention or incentive to meaningfully self-regulate.¹³⁸ Much like the resistance of oil or coal firms to climate change initiatives, the businesses directing the operations of Social Media platforms are resistant to calls for change to this lucrative model despite the obvious harm it causes. In addition to the depression the algorithm causes, there is significant evidence of the existence of 'pipelines,' or radicalizing progressions of extreme ideological content recommended by algorithms. This is not merely incidental, as the recommendation algorithms of Social Media platforms often seem to find that radicalization is a considerable driver of engagement and revenue for the platform. DiMaggio found that Social Media consumption, especially right-wing media consumption, was a significant statistical predictor of the acceptance of modern conspiracy theories.¹³⁹ This is another consequence of engagement and inability to self-regulate that Social Media platforms are aware of and do not meaningfully address in most cases: The aforementioned documents leaked from Meta included an internal presentation which found that 64% of people who joined extremist Facebook Groups did so because those Groups were recommended to them by Facebook's recommendation algorithms.¹⁴⁰ Facebook's moderation of Groups on their platform was lackluster at best, and groups emerged as common sources of disinformation and radicalizing narratives. Meta announced in early 2021 that they would no longer algorithmically recommend groups to people, but the damage had been done at this point as over 600 million

¹³⁵ Lai T and Zheng X, 'Machine Learning Based Social Media Recommendation,' *2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)* (2015)

¹³⁶ A) Yoon S and others, 'Is Social Network Site Usage Related to Depression? A Meta-Analysis of Facebook-Depression Relations' (2019) 248 *Journal of Affective Disorders* 65
<<https://www.sciencedirect.com/science/article/pii/S0165032718321700>> accessed 25 July 2022

B) Hillard J, 'Social Media Addiction' (*Addiction Center*)
<<https://www.addictioncenter.com/drugs/social-media-addiction/>> accessed 30 July 2022

¹³⁷ Seetharaman JH and D, 'Facebook Executives Shut Down Efforts to Make the Site Less Divisive' *Wall Street Journal* (26 May 2020)
<<https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499>> accessed 25 July 2022

¹³⁸ Farthing R and Sooriyakumaran D, 'Why the Era of Big Tech Self-Regulation Must End' 92 *AQ - Australian Quarterly* 3 <<https://search.informit.org/doi/abs/10.3316/agispt.20211005054383>> accessed 25 July 2022

¹³⁹ *Ibid* 123.

¹⁴⁰ *Ibid* 133.

users in total used the feature.¹⁴¹ Some of the recommendation algorithms Social Media platforms utilize to maximize user engagement can deeply radicalize users. However, the private business interests that operate most Social Media platforms are hesitant to take any action which may hurt their bottom line. Social Media platforms continue to facilitate the victimization of users who fall into radicalizing pipelines and those that are harmed by extremism facilitated by these platforms.

3.2.1: Radicalization Pipelines on Youtube

The political influence and harmful effects Facebook can generate have been widely reported on due to public events such as the Cambridge Analytica scandal and the congressional whistleblower hearings in the United States.¹⁴² While disinformation narratives and those that disseminate them are by no means limited to a specific platform, a more illustrative example of a platform with significant radicalizing effects is Youtube. Youtube is a video-sharing platform owned and operated by Google that gives a share of advertising revenue to users with larger followings regardless of their views. Munger & Phillips point out some unique aspects of Youtube that gave rise to the popularity of radicalizing content on the platform: There is very little up-front commitment necessary to publish videos on the platform, videos are generally a better medium for communicating emotion, and Youtube's recommendation algorithm functions in a way that is predictable to experienced users, incentivizing sensationalization and benefitting fringe content which has less commitment to objectivity.¹⁴³ Youtube also has an autoplay feature that will play the most highly recommended video shortly after the conclusion of the video being watched. Youtube notoriously hosts a multitude of prominent far-right channels whose videos enjoy sizable audiences and revenue despite frequently espousing white supremacist narratives, among other harmful behaviors. These channels usually revolve around a single 'micro-celebrity' that asserts that their narratives are more trustworthy than contradictory mainstream media narratives.¹⁴⁴ Shoddy moderation allows the content on these channels to exist in close, symbiotic proximity with mainstream content as long as no slurs are used.¹⁴⁵ This has led Youtube to facilitate one of the most prominent online radicalization pipelines: Users on Youtube can easily go from unrelated content or mainstream media content to content made by channels that act as 'gateways' to further

¹⁴¹ Abrams Z, 'Stress of Mass Shootings Causing Cascade of Collective Traumas' (<https://www.apa.org>, 11 July 2022) <<https://www.apa.org/monitor/2022/09/news-mass-shootings-collective-traumas>> accessed 28 July 2022

¹⁴² Lewis P and Hilder P, 'Leaked: Cambridge Analytica's Blueprint for Trump Victory' *The Guardian* (23 March 2018)

<<https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory>> accessed 25 July 2022

¹⁴³ Munger K and Phillips J, 'Right-Wing YouTube: A Supply and Demand Perspective' (2022) 27 *The International Journal of Press/Politics* 186 <<http://journals.sagepub.com/doi/10.1177/1940161220964767>> accessed 26 July 2022

¹⁴⁴ Lewis R, "'This Is What the News Won't Show You': YouTube Creators and the Reactionary Politics of Micro-Celebrity' (2020) 21 *Television & New Media* 201

<<http://journals.sagepub.com/doi/10.1177/1527476419879919>> accessed 26 July 2022

¹⁴⁵ Lewis B, 'Alternative Influence' (*Data & Society*, 18 September 2018)

<<https://datasociety.net/library/alternative-influence/>> accessed 25 July 2022

radicalization.¹⁴⁶ This appears to be due to the fact that gateway channels on Youtube receive much more interaction from audiences than the average video.¹⁴⁷ These gateway channels are often considered ideologically ‘alt-lite’.¹⁴⁸ Alt-lite narratives are usually composed of thinly veiled homophobic, misogynistic, and/or xenophobic narratives that rail against social justice, political correctness, and inclusion. Some notable alt-lite online spaces are the so-called ‘manosphere,’ a semi-connected group of spaces that promote misogyny, hyper-masculinity, and anti-feminism, and the ‘intellectual dark web,’ a staunchly reactionary group of Youtube micro-celebrities that defends the purported rationality of society’s discriminatory biases.¹⁴⁹ Jordan Peterson, the most prominent figure in the intellectual dark web, constantly decries ‘cultural marxism,’ a reference to the nazi propaganda narrative of ‘cultural bolshevism.’ The ‘manosphere’ is found to directly lead significant portions of its audience to alt-right content.¹⁵⁰ The story of Buckey Wolfe, a QAnon follower, and affiliate of the white nationalist extremist group ‘the proud boys,’ is a tragic but informative look into how the radicalizing forces of Social Media platforms have very serious offline consequences. In early 2019, Buckey Wolfe murdered his brother during what appears to be a mental health crisis. He called the police and admitted to his crime, pleading: “Kill me, kill me, I can’t live in this reality,” and “God told me he was a lizard”.¹⁵¹ Wolfe fits the profile of a victim of online radicalization: He is a white male that had been diagnosed with schizophrenia at some point before the murder. This is borne out by the findings of QAnon researcher Logan Strain (aka. Travis View), who documented Buckey's personal radicalization: Buckey began by watching nonpolitical mainstream content. Suddenly, Buckey started to watch many videos from an anti-feminist gateway channel, which brought him to videos promoting conspiracy theories and eventually content from prominent far-right figures (e.g. Alex Jones and Stephen Crowder, among others). He continued to consume content from increasingly extreme channels that eventually espoused explicitly white supremacist narratives. Once he started to consume these openly extreme narratives, he also began to watch QAnon videos as well as videos on narratives about shape-shifting reptiles disguising themselves as

¹⁴⁶ Mamié R, Horta Ribeiro M and West R, ‘Are Anti-Feminist Communities Gateways to the Far Right? Evidence from Reddit and YouTube,’ *13th ACM Web Science Conference 2021* (Association for Computing Machinery 2021) <<https://doi.org/10.1145/3447535.3462504>> accessed 25 July 2022

¹⁴⁷ Ribeiro MH and others, ‘Auditing Radicalization Pathways on YouTube,’ *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (ACM 2020) <<https://dl.acm.org/doi/10.1145/3351095.3372879>> accessed 25 July 2022

¹⁴⁸ ‘Alt-lite’ is a variation on the term ‘Alt-right.’ ‘Alt-right’ generally describes more novel elements of the far-right movement that are especially prominent online. ‘Alt-lite’ describes ideologies that represent a middle ground between mainstream ideologies and alt-right ideologies that generally object to open hatred but advocate for similar narratives through dog-whistles and a generally biased perspective.

‘From Alt Right to Alt Lite: Naming the Hate’ <<https://www.adl.org/resources/background/alt-right-alt-lite-naming-hate>> accessed 25 July 2022

¹⁴⁹ Finlayson A, ‘Neoliberalism, the Alt-Right and the Intellectual Dark Web’ (2021) 38 *Theory, Culture & Society* 167 <<http://journals.sagepub.com/doi/10.1177/02632764211036731>> accessed 30 July 2022

¹⁵⁰ *Ibid* 141.

¹⁵¹ Green SJ, “‘God Told Me He Was a Lizard’: Seattle Man Accused of Killing His Brother with a Sword’ *Seattle Times* (Seattle, Washington, USA, 8 January 2019) <<https://www.seattletimes.com/seattle-news/law-justice/god-told-me-he-was-a-lizard-seattle-man-accused-of-killing-his-brother-with-a-sword/>>

humans.¹⁵² While Buckey’s story is one extreme example of many, and his psychological disorders certainly played a role in the crime, there is no mistaking the obvious effect conspiracy theories found after a radicalizing experience on Youtube played in this murder. The stark contrast of his online activities before and after his exposure to the gateway channels is shocking and indicates the important role mainstream Social Media platforms play in radicalization and the commission of extremely violent acts.

3.2.2: Stochastic Extremism and the ‘Chan’ Forums

No discussion of online extremism is complete without mentioning the style of alt-right narratives found in places like 4chan or 8chan, notoriously some of the most toxic Social Media platforms. These platforms are collectively referred to as the ‘chan’ forums and are regarded as one of if not the final platforms radicalized individuals will frequent before they become committed extremists. While each platform in this grouping have different characteristics than the others, they all share the same ‘image board’ format where pictures are posted with accompanying text. The chan platforms offer users little to no restrictions on posting and user anonymity, attracting extremists and others looking for unregulated hateful (usually far-right) communities online. While the group of platforms that constitute the chan forums is constantly in flux for a number of reasons, they form a slightly fragmented but overall distinct subculture.¹⁵³ The most prominent examples are the aforementioned 4chan and 8chan platforms. 8chan is notorious for its central role in a series of mass shootings in 2019, after which it was shut down. This included the mass shootings in Christchurch, San Diego, and El Paso. Each of the individuals who committed these atrocities made posts that included racist imagery and captions with links to long manifestos detailing their motives for their crime and/or livestream links so fellow users could watch the atrocity take place. All of these manifestos directly referenced or alluded to the aforementioned ‘Great Replacement Theory’ discussed in section 2.2. This narrative fits all of the indicators of extremism: It claims that ‘white nations’ are being ‘invaded’ by non-white immigrants who will bring about the ruin of the nation in some way or another.¹⁵⁴ While these narratives and extremist violence they incite are not exclusive to the chan platforms and users, they are very common on the chan platforms. Other examples of horrific violence being committed for these reasons include the Norway attacks (2011) and the recent Buffalo shooting. Surviving screenshots of the 8chan posts preceding shootings show that the most immediate responses to the posts were overwhelmingly positive and even encouraging. In fact, the overwhelmingly like-minded and encouraging attitude these communities have to

¹⁵² Strain L, ‘The “Likes” on the Youtube Page of Buckey Wolfe’ (*Twitter*, 10 January 2019) <https://twitter.com/travis_view/status/1083437810634248193> accessed 26 July 2022

¹⁵³ Generally, the chan forums include: 4chan, 8kun (formerly 8chan), Infinitychan, 9chan, 16chan, ShitChan, EndChan, and NeinChan among others.

Baele SJ, Brace L and Coan TG, ‘Variations on a Theme? Comparing 4chan, 8kun, and Other Chans’ Far-Right “/Pol” Boards’ (2021) *15 Perspectives on Terrorism* 65

¹⁵⁴ Walton M, ‘Defending White America: The Apocalyptic Meta-Narrative of White Nationalist Rhetoric’ [2020] *Theses and Dissertations* <<https://scholarsarchive.byu.edu/etd/8491>>

extremist acts seems to have created an effect similar to the gamification of the discovery of novel disinformation as seen in QAnon communities.¹⁵⁵ Furthermore, each of these shootings, especially the Christchurch shooting, seems to have had an influential role on the subsequent shootings. The Christchurch shooting itself seems to have been inspired by the 2015 Charleston and aforementioned Norway shootings.¹⁵⁶ Baele et al. explain the ‘tarrant effect’ (named after the person who committed the Christchurch shooting): After the shooting, the use of racial slurs on 8chan increased for over a month, and the individual who committed the attack was ‘sainted’ as a martyr by peers on the platform.¹⁵⁷ The perpetrator’s manifesto was titled ‘The Great Replacement’.¹⁵⁸ This happened in a similar capacity inside ‘incel’ communities, which are also prominent on chan platforms, after the Isla Vista shootings in 2014.¹⁵⁹ As mentioned before, the 8chan platform was forced to shut down after the spate of attacks in 2019, but a group of copycat platforms including a near-reboot in 8kun sprung up soon afterwards.¹⁶⁰ The unregulated environment of the chan platforms provide extremists a forum for highly conspiratorial and/or explicitly extremist discussions which include narratives like Pizzagate, QAnon, and explicit white supremacist narratives among others, enabling them to create deeply insular and outwardly hostile environments that justify and glorify extremist acts. The tarrant effect is an effective and extreme instance of the trend of copycat extremists, who do not just mimic previous perpetrators of extremist acts but “relate to them, are inspired by them, and want to outdo them” according to Murray.¹⁶¹ This has given rise to extremist acts that seem not to be terroristic in nature but purely motivated by a desire to commit horrific acts of violence. The recent July 4th shootings in Highland Park is one of these cases, as the perpetrator left a large online footprint that included extensive participation in many fringe online communities that fetishized and shared content depicting graphic violence.¹⁶² These communities spawn extremists in a fashion that Amman

¹⁵⁵ Evans R, ‘The El Paso Shooting and the Gamification of Terror’ (*bellingcat*, 4 August 2019) <<https://www.bellingcat.com/news/americas/2019/08/04/the-el-paso-shooting-and-the-gamification-of-terror/>> accessed 27 July 2022

¹⁵⁶ Am AB and Weimann G, ‘Fabricated Martyrs: The Warrior-Saint Icons of Far-Right Terrorism’ (2020) 14 *Perspectives on Terrorism* 130

¹⁵⁷ A) Baele SJ, Brace L and Coan TG, ‘The “Tarrant Effect”: What Impact Did Far-Right Attacks Have on the 8chan Forum?’ [2020] *Behavioral Sciences of Terrorism and Political Aggression* 1 <<https://www.tandfonline.com/doi/full/10.1080/19434472.2020.1862274>> accessed 27 July 2022

B) *Ibid* 156.

¹⁵⁸ Evans R, ‘Shitposting, Inspirational Terrorism, and the Christchurch Mosque Massacre’ (*bellingcat*, 15 March 2019) <<https://www.bellingcat.com/news/rest-of-world/2019/03/15/shitposting-inspirational-terrorism-and-the-christchurch-mosque-massacre/>> accessed 27 July 2022

¹⁵⁹ DiBranco A, ‘Male Supremacist Terrorism as a Rising Threat’ [2020] *International Centre for Counter-terrorism* <<https://icct.nl/publication/male-supremacist-terrorism-as-a-rising-threat/>> accessed 27 July 2022

¹⁶⁰ *Ibid* 153.

¹⁶¹ Murray JL, ‘Mass Media Reporting and Enabling of Mass Shootings’ (2017) 17 *Cultural Studies ↔ Critical Methodologies* 114

¹⁶² It must be noted that the perpetrator may have manipulated his own digital footprint to portray himself or signal in a certain way. What is certain about the perpetrator of the Highland Park shooting is that he was an avid Trump supporter.

Sullivan B, ‘Highland Park Suspect’s Online History Reveals a Fascination with Violence’ (NPR, 5 July 2022) <<https://text.npr.org/1109844728>> accessed 27 July 2022

and Meloy call “stochastic.” This trend of stochastic extremism creates a constant, serious risk to public safety that can be extremely difficult to preempt.¹⁶³ These devastating and somewhat unpredictable outbursts of violence are committed by highly motivated individuals whose worldviews, cultivated online, justify atrocities due to perceived existential threats. Extremist acts are often extremely deadly and create a sense of collective trauma and anxiety that an attack could be imminent.¹⁶⁴ The overall increase in extremist violence due to online extremism and the radicalizing effects of Social Media platforms subjects the population to frequent stochastic extremism.

3.2.3: The Facilitation of Coordinated Extremism by Social Media Platforms

While radicalization pipelines on Social Media platforms factor into a trend of stochastic extremists that generally act alone as ‘lone wolves,’ Social Media platforms can also facilitate more organized extremism. As mentioned in section 3.1.1, 64% of all users joining extremist Facebook groups were recommended to do so by Facebook’s platform, the residual effects of which are still lingering. As mentioned in section 1.3, the organizing power Social Media platforms enable is exemplified by overwhelmingly positive examples like the Black Lives Matter movement and frightening examples of coordinated extremist acts. These acts’ overall nature and scale vary widely: The January 6th Capitol Riots illustrate the spectrum of participation in coordinated extremist acts. As an event featuring Donald Trump directly before what the administration saw as their last realistic opportunity to overturn the election, there was significant discussion on mainstream and extreme Social Media platforms about the event. This resulted in a large crowd that included non-radicalized individuals attending out of a desire to see Donald Trump speak as well as many far-right extremist elements.¹⁶⁵ This included nationally syndicated militant white nationalist groups like the proud boys, who made advance preparations to rush the Capitol building and QAnon followers.¹⁶⁶ In Germany, extreme anti-vaccine groups use Telegram, a group messaging platform, to disseminate disinformation and conspiracy narratives, resulting in widespread protests against anti-pandemic public health measures. Extremists take advantage of the fact that Social Media platforms offer both easy access to large, possibly mainstream audiences, and the ability to stay relatively hidden on anonymous,

¹⁶³ Amman M and Meloy JR, ‘Stochastic Terrorism’ (2021) 15 Perspectives on Terrorism 2

¹⁶⁴ Abrams Z, ‘Stress of Mass Shootings Causing Cascade of Collective Traumas’ (<https://www.apa.org>, 11 July 2022) <<https://www.apa.org/monitor/2022/09/news-mass-shootings-collective-traumas>> accessed 28 July 2022

¹⁶⁵ Harton HC, Gunderson M and Bourgeois MJ, “‘I’ll Be There with You’: Social Influence and Cultural Emergence at the Capitol on January 6.” [2022] Group Dynamics: Theory, Research, and Practice <<http://doi.apa.org/getdoi.cfm?doi=10.1037/gdn0000185>> accessed 28 July 2022

¹⁶⁶ A) Reneau N and others, ‘Proud Boys Led Major Breaches of Capitol on Jan. 6, Video Investigation Finds’ *The New York Times* (17 June 2022) <<https://www.nytimes.com/2022/07/12/us/politics/proud-boys-jan-6.html>> accessed 28 July 2022

B) Stratton J, “‘All about a Revolution,’ or Duped by QAnon? Read the Interview between Doug Jensen and the FBI” (*KCCI*, 11 April 2022) <<https://www.kcci.com/article/iowan-interview-between-doug-jensen-and-the-fbi-qanon/39694179>> accessed 28 July 2022

unmoderated platforms to mobilize less radicalized individuals.¹⁶⁷ However, the most harmful cases of radicalization and extremism on Social Media platforms appear to occur when Social Media platforms facilitate popular support and participation in institutional discrimination. The Rohingya community of Myanmar is pursuing legal actions against Meta in courts in the United States and the United Kingdom for Facebook's role in the genocide committed against the group.¹⁶⁸ The Burmese military continues to systematically target the Rohingya community in Myanmar, committing scores of mass rapes, killings, and forced resettlements, provoking United Nations investigators to call for prosecutions of top military commanders in Myanmar for crimes including genocide.¹⁶⁹ The state media apparatus was a proponent of the nationalist rhetoric that led to extreme violence against Rohingya communities, which played an important role in mainstreaming anti-Rohingya nationalist narratives. Facebook quickly became ubiquitous after its 2011 launch in Myanmar, and became an important forum for the widespread popularization of anti-Rohingya narratives. In 2018 Meta (still Facebook then) commissioned an independent organization (Business for Social Responsibility) to investigate the impacts their platforms had on the human rights of members of Rohingya communities. This report found that "Facebook has become a means for those seeking to spread hate and cause harm, and posts have been linked to offline violence," with a small minority of users actively attempting to radicalize and incite extremism on the platform.¹⁷⁰ Despite the highly publicized findings of the report and the commitments made by Meta to address their complicity in this genocide, Facebook continues to show advertisements containing anti-Rohingya hate speech to its Burmese userbase.¹⁷¹ Damningly, the class action suit filed by the Rohingya community against Meta in San Francisco states that Meta was "willing to trade the lives of the Rohingya people for better market penetration in a small country in south-east Asia... In the end, there was so little for Facebook to gain from its continued presence in Burma, and the consequences for the Rohingya people could not have been more dire. Yet, in the face of this knowledge, and possessing the tools to stop it, it simply kept marching forward."¹⁷²

¹⁶⁷ Grelich K and others, 'Far-Right Extremists' Use Of Social Media Platforms To Communicate And Spread Radicalized Beliefs' (*The Counterterrorism Group*, 6 December 2021) <<https://www.counterterrorismgroup.com/post/far-right-extremist-use-of-social-media-platforms-to-communicate-and-spread-radicalized-beliefs>> accessed 28 July 2022

¹⁶⁸ Milmo D, 'Rohingya Sue Facebook for £150bn over Myanmar Genocide' *The Guardian* (6 December 2021) <<https://www.theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence>> accessed 28 July 2022

¹⁶⁹ 'Myanmar Military Leaders Must Face Genocide Charges – UN Report' (*UN News*, 27 August 2018) <<https://news.un.org/en/story/2018/08/1017802>> accessed 28 July 2022

¹⁷⁰ Business for Social Responsibility, 'Human Rights Impact Assessment: Facebook in Myanmar' (Meta 2018) <https://about.fb.com/wp-content/uploads/2018/11/bsr-facebook-myanmar-hria_final.pdf>

¹⁷¹ Lee R, 'Extreme Speech in Myanmar: The Role of State Media in the Rohingya Forced Migration Crisis' (2019) 13 *International Journal of Communication* 22 <<https://ijoc.org/index.php/ijoc/article/view/10123>> accessed 28 July 2022

¹⁷² Milmo D, 'Rohingya Sue Facebook for £150bn over Myanmar Genocide' *The Guardian* (6 December 2021) <<https://www.theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence>> accessed 28 July 2022

3: Conclusion

Social Media platforms provide all of their users with instantaneous access to any and all content they may wish to consume. However, this information disorder can lead some individuals to disinformation that injects confusion and distrust into their worldview. Many individuals find radicalizing narratives on mainstream Social Media platforms like Twitter, Facebook, and Youtube, which lead them into more extreme online spaces in a downwards spiral of disinformation that could, in the case of society's most disturbed individuals, lead to acts of extremist violence. Generally, the most harmful radicalization and extremism is done by those with fringe reactionary ideologies, notably jihadism or far-right/white supremacist ideologies that assume "chauvinistic and hyper-masculine collective identities".¹⁷³ These convoluted narratives directly encourage extremist acts of violence by claiming that the dehumanization and destruction of a scapegoated outgroup is necessary for the survival of the designated ingroup. Conspiracy theories like QAnon exemplify the addictive attraction some develop to conspiracy theories, as that narrative persists despite years without any true predictions. The inherent contradictions of extremist narratives and conspiracy theories indicate how radicalizing narratives thrive on information disorder and poor living conditions. Social Media platforms play a significant role in the radicalization of users not only by distributing extreme narratives but also by ignoring the consequences of excessive Social Media engagement. Social Media use, in general, has been linked with depression and anxiety. The recommendation algorithms Social Media platforms employ to maximize engagement can put users in a 'pipeline' to extremist content. However, the owners of these platforms have done little to address these risks as it could have an adverse effect on user engagement and their bottom line. The role a Youtube pipeline to extreme conspiracy narratives played in the frightening story of Buckley Wolfe, the trend of chan platform-related far-right extremist shootings, and Facebook's promotion of anti-Rohingya sentiments in Myanmar are all prominent examples of the consequences of radicalization perpetuated by Social Media platforms. Social Media platforms must end their complicity in radicalization and the perpetration of stochastic extremism and organized acts of extremism.

¹⁷³ Ibid 120.

Conclusion

In section 1, this thesis began by discussing the radical paradigm shift in mass communications caused by widespread access and adoption of mobile phones, wireless data, and Social Media platforms, among other new technologies. The unprecedented and universal capabilities Social Media platforms offer for anyone to distribute content globally are causing major societal shifts that must be understood. The relative destruction of the barriers to entry to mass communications has grown, diversified, and democratized media and news. Social Media platforms flung the Overton window wide open in a few years, ushering in a massive wave of outlets, some offering fringe or extreme perspectives, disrupting traditional 'pro-custom' news outlets. Social Media has fundamentally intertwined itself with politics, and these platforms have become a modern and generally more representative instance of Habermas' public sphere. It also adds a participatory dimension to media consumption: Social Media platforms offer individuals the opportunity to interact with or respond to the content they consume like never before, allowing any individual to become a political actor rather than a participant in collective action. Paradoxically, while these platforms exist as public spheres in society, they are operated as private businesses. While they were probably justified in doing so, Twitter showed that Social Media platforms are so powerful that they can strongarm the sitting President of the United States without resistance. Ironically, as most firms operating Social Media platforms are based in the United States, Section 230 of the American Communications Decency Act absolves any responsibility for their role in harm their platform may create. While platforms were essentially able to operate with impunity worldwide, the European Union is attempting to revamp its legal framework to regain regulatory control of these platforms within the borders of the union by passing the Digital Services act, which aims to mark the beginning of a new 'age of accountability' for Social Media platforms.

In section 2, the thesis established that the root cause of the harm caused by Social Media platforms is disinformation. Due to a combination of factors, most notably a deep distrust of institutions and the commonplace use of heuristics to determine the truth, disinformation narratives have become popular on Social Media platforms. The democratization of media has allowed bad actors who range from state-sponsored trolls and bots to individuals looking to peddle disinformation narratives for personal or political gain. Individuals experience an 'information disorder' due to the pure volume of content with uncertain veracity that generally confuses them. Information disorder can threaten, for example, the human rights to life, health, free and fair elections, non-discrimination, and private life. Established legal procedures exist in the case of risks posed by malinformation, harassment, and defamation. On the other hand, minimal or authoritarian protections stand between disinformation and the human rights to life, health, free and fair elections, and non-discrimination from the dissemination of disinformation on Social Media. States must better protect these rights across the board. Recently, the Russian invasion of Ukraine and the COVID-19 pandemic clearly illustrated the potentially deadly consequences of disinformation and the crucial need to enact measures to eliminate these risks.

A comparison of legal frameworks in the United States, China, and the European Union exemplifies the different effects legislative anti-disinformation philosophies have on the balance between the protection of the right to free speech and the protection of the rights affected by disinformation on Social Media platforms: The laissez-faire American approach has essentially legalized hate speech and authoritarian regulations in China all but explicitly forbid all dissent. Europe is attempting to find a suitable middle ground by establishing a 'co-regulation' model with the Digital Services Act. The 'middle-ground' approach is still under development, and as such, the Digital Services Act is the foundation on which to build an updated and effective legal framework.

In Section 3, this thesis discussed the central role Social Media platforms themselves play in the creation of information disorder and radicalization. Many individuals encounter radicalizing reactionary narratives on mainstream Social Media platforms like Twitter, Facebook, and Youtube. Reactionary extremists often peddle their narratives on mainstream Social Media platforms, which can draw a small minority of users into online spaces that actively facilitate radicalization and extremism. Extremism necessitates harm to a scapegoat group to preserve the chosen ingroup due to "chauvinistic and hyper-masculine collective identities," meaning that these ideologies directly incite their followers to violence. The radicalization of users on Social Media platforms is one of the major factors in the trend of public acts of extremism. Generally, perpetrators of extreme acts are men. Perpetrators usually fall into two groups: stochastic extremists such as the 'lone-wolf shooter' and members of an extremist group such as the proud boys. As we have seen from the systemic persecution of the Rohingya in Myanmar, these extremist groups can even represent central authorities. While suitable protections exist for the harms Social Media platforms can cause, there are few solutions to the problems posed by disinformation and its downstream effects in place anywhere.

This thesis ends by repeating the call to the owners of Social Media platforms to immediately end their complicity in radicalization as well as the perpetration of stochastic and organized acts of extremism. Steps should also be taken to address the structural causes of information disorder like widespread distrust in public institutions and dissatisfaction with material living conditions. While nothing has stopped the widespread disinformation on Social Media platforms, any countermeasure that adds a marginal advantage deserves some consideration. Some solutions already mentioned, e.g., moderation, media literacy education, and inoculation, have demonstrated some efficacy and merit inclusion in a broader package of countermeasures. It is most likely a fool's errand to attempt to regulate all Social Media platforms considering the amount of deeply insular online extremist groups. However, as Cloudflare's shutdown of 8chan illustrated, platforms like the chan boards that intentionally facilitate extremism can and should be addressed by their network providers. Many copycat platforms emerged after 8chan's shutdown. However, taking action against platforms that directly incite violence is still required. Cloudflare's action was taken too late as the perpetrators gained widespread attention after their crimes, inspiring extremist acts multiple years in the

future. However, mainstream platforms like Twitter, Facebook, and Youtube must address their roles in disseminating disinformation.

Legal solutions like the Digital Services Act that attempt to truly balance freedom of expression and the protection of the population from the effects of disinformation on Social Media are direly needed. The two main categories these solutions fall under are a priori and a posteriori solutions. Essentially, solutions attempt to prevent users from consuming disinformation, and the subsequent effects, before or after it is consumed. Generally, we should aim to stop radicalization before it occurs, as radicalization can quickly bring about adverse outcomes. Inoculation is a promising tool that, if correctly implemented, should help to prevent the initial steps of radicalization. Media literacy education is another example of a practical measure that would make users more resilient to disinformation. While all of these measures would ameliorate the problem, the cornerstone of any regulation effort must completely reinvent the moderation and recommendation algorithms of Social Media platforms. Benesch presents the most intriguing solution, proposing that the firms that operate Social Media platforms should adopt a version of international human rights law customized for their purposes. Benesch reasons that, since also stating that the preamble of the Universal Declaration of Human Rights calls upon "all organs of society" to do their part to uphold Human Rights, mainstream Social Media platforms should follow the mandates in Articles 19 and 20 of the International Covenant on Civil and Political Rights primarily.¹⁷⁴ Even if legal complications make this problematic for mainstream Social Media platforms to implement fully, the principles of Articles 19 and 20 of the International Covenant of Civil and Political Rights should inform the reimagination of the moderation and recommendation features of Social Media platforms.

¹⁷⁴ Benesch S, 'But Facebook's Not a Country: How to Interpret Human Rights Law for Social Media Companies' (Social Science Research Network 2020) SSRN Scholarly Paper 3692701 <<https://papers.ssrn.com/abstract=3692701>> accessed 29 July 2022

Bibliography

- Abrams A, 'Here's What We Know So Far About Russia's 2016 Meddling' (Time, 18 April 2019) <<https://time.com/5565991/russia-influence-2016-election/>> accessed 23 July 2022
- Abrams Z, 'Stress of Mass Shootings Causing Cascade of Collective Traumas' (<https://www.apa.org>, 11 July 2022) <<https://www.apa.org/monitor/2022/09/news-mass-shootings-collective-traumas>> accessed 28 July 2022
- Abu Arqoub O and others, 'Mapping the Scholarship of Fake News Research: A Systematic Review' (2022) 16 *Journalism Practice* 56 <<https://doi.org/10.1080/17512786.2020.1805791>> accessed 14 May 2022
- Ahmed I, 'The Disinformation Dozen: Why Platforms Must Act On Twelve Leading Online Anti-Vaxxers' (Center for Countering Digital Hate 2021) <<https://counterhate.com/wp-content/uploads/2022/05/210324-The-Disinformation-Dozen.pdf>> accessed 18 July 2022
- Aichner T and others, 'Twenty-Five Years of Social Media: A Review of Social Media Applications and Definitions from 1994 to 2019' (2021) 24 *Cyberpsychology, Behavior, and Social Networking* 215 <<https://www.liebertpub.com/doi/full/10.1089/cyber.2020.0134>> accessed 3 May 2022
- Am AB and Weimann G, 'Fabricated Martyrs: The Warrior-Saint Icons of Far-Right Terrorism' (2020) 14 *Perspectives on Terrorism* 130
- Amman M and Meloy JR, 'Stochastic Terrorism' (2021) 15 *Perspectives on Terrorism* 2
- Applebaum A, 'Tweet from 3:38 PM, Jul 31, 2021' (Twitter) <<https://twitter.com/anneapplebaum/status/1421480489748254723>> accessed 22 July 2022
- Araoz, Jennifer. "Opinion | Jeffrey Epstein Raped Me When I Was 15." *The New York Times*, 14 Aug. 2019. NYTimes.com, <https://www.nytimes.com/2019/08/14/opinion/jeffrey-epstein-jennifer-araoz.html>.
- Baele SJ, Brace L and Coan TG, 'The "Tarrant Effect": What Impact Did Far-Right Attacks Have on the 8chan Forum?' [2020] *Behavioral Sciences of Terrorism and Political Aggression* 1 <<https://www.tandfonline.com/doi/full/10.1080/19434472.2020.1862274>> accessed 27 July 2022
- Baele SJ, Brace L and Coan TG, 'Variations on a Theme? Comparing 4chan, 8kun, and Other Chans' Far-Right "Pol" Boards' (2021) 15 *Perspectives on Terrorism* 65
- Béland D and others, 'Trump, Bolsonaro, and the Framing of the COVID-19 Crisis: How Political Institutions Shaped Presidential Strategies' (2021) 184 *World Affairs* 413 <<http://journals.sagepub.com/doi/10.1177/004382002111053314>> accessed 16 July 2022
- Benesch S, 'But Facebook's Not a Country: How to Interpret Human Rights Law for Social Media Companies' (Social Science Research Network 2020) SSRN Scholarly Paper 3692701 <<https://papers.ssrn.com/abstract=3692701>> accessed 29 July 2022
- Berger JM, *Extremism* (The MIT Press 2018)
- Bradd S, 'Infodemic' (World Health Organization, n.d.) <<https://www.who.int/health-topics/infodemic>> accessed 18 July 2022
- Bryant, Nicholas A. *The Franklin Scandal: A Story of Powerbrokers, Child Abuse and Betrayal*. First edition (revised for softcover), Trine Day, 2012.
- Bump P, 'Analysis | Ivermectin Is the Signature Example of Politics Trumping Health' *Washington Post* (31 March 2022)

<<https://www.washingtonpost.com/politics/2022/03/31/ivermectin-is-signature-example-politics-trumping-health/>> accessed 18 July 2022

Business for Social Responsibility, 'Human Rights Impact Assessment: Facebook in Myanmar' (Meta 2018) <https://about.fb.com/wp-content/uploads/2018/11/bsr-facebook-myanmar-hria_final.pdf>

Darcy O, 'RT America Ceases Productions and Lays off Most of Its Staff' (CNN, 4 March 2022) <<https://www.cnn.com/2022/03/03/media/rt-america-layoffs/index.html>> accessed 23 July 2022

DiBranco A, 'Male Supremacist Terrorism as a Rising Threat' [2020] International Centre for Counter-terrorism <<https://icct.nl/publication/male-supremacist-terrorism-as-a-rising-threat/>> accessed 27 July 2022

DiMaggio AR, 'Conspiracy Theories and the Manufacture of Dissent: QAnon, the "Big Lie", Covid-19, and the Rise of Rightwing Propaganda' [2022] Critical Sociology 089692052110736 <<http://journals.sagepub.com/doi/10.1177/08969205211073669>> accessed 25 July 2022

Erlich A and Garner C, 'Is Pro-Kremlin Disinformation Effective? Evidence from Ukraine' [2021] The International Journal of Press/Politics 194016122110452 <<http://journals.sagepub.com/doi/10.1177/19401612211045221>> accessed 22 July 2022

Evans R, 'Shitposting, Inspirational Terrorism, and the Christchurch Mosque Massacre' (bellingcat, 15 March 2019) <<https://www.bellingcat.com/news/rest-of-world/2019/03/15/shitposting-inspirational-terrorism-and-the-christchurch-mosque-massacre/>> accessed 27 July 2022

Evans R, 'The El Paso Shooting and the Gamification of Terror' (bellingcat, 4 August 2019) <<https://www.bellingcat.com/news/americas/2019/08/04/the-el-paso-shooting-and-the-gamification-of-terror/>> accessed 27 July 2022

Farthing R and Sooriyakumaran D, 'Why the Era of Big Tech Self-Regulation Must End' 92 AQ - Australian Quarterly 3 <<https://search.informit.org/doi/abs/10.3316/agispt.20211005054383>> accessed 25 July 2022

Fathaigh R, Helberger N and Appelman N, 'The Perils of Legally Defining Disinformation' (2021) 10 Internet Policy Review <<https://policyreview.info/articles/analysis/perils-legally-defining-disinformation>> accessed 4 July 2022

Finlayson A, 'Neoliberalism, the Alt-Right and the Intellectual Dark Web' (2021) 38 Theory, Culture & Society 167 <<http://journals.sagepub.com/doi/10.1177/02632764211036731>> accessed 30 July 2022

Fletcher R, Robertson CT and Nielsen RK, 'How Many People Live in Politically Partisan Online News Echo Chambers in Different Countries?' (2021) 1 Journal of Quantitative Description: Digital Media <<https://journalqd.org/article/view/2585>> accessed 30 April 2022

Frank Hoffman, 'Conflict in the 21st Century: The Rise of Hybrid Wars' (Potomac Institute for Policy Studies 2007) <<https://potomacinstitute.org/reports/19-reports/1163-conflict-in-the-21st-century-the-rise-of-hybrid-wars>> accessed 19 July 2022

Gabbatt A, 'Fox News Suddenly Goes Quiet on "Great Replacement" Theory after Buffalo Shooting' The Guardian (17 May 2022) <<https://www.theguardian.com/us-news/2022/may/17/buffalo-shooting-fox-news-tucker-carlson-great-replacement-theory>> accessed 14 July 2022

Gaskins B and Jerit J, 'Internet News: Is It a Replacement for Traditional Media Outlets?' (2012) 17 The International Journal of Press/Politics 190 <<http://journals.sagepub.com/doi/10.1177/1940161211434640>> accessed 3 May 2022

Golombek M, *Pandemic Politics As A Holocaust 2.0? : The Perception Of The Holocaust In Telegram-Communications Of German Covid-Protectors* (2022) <<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-478375>> accessed 28 July 2022

- Goyanes M, Borah P and Gil de Zúñiga H, 'Social Media Filtering and Democracy: Effects of Social Media News Use and Uncivil Political Discussions on Social Media Unfriending' (2021) 120 *Computers in Human Behavior* 106759 <<https://linkinghub.elsevier.com/retrieve/pii/S0747563221000819>> accessed 1 May 2022
- Gravel Institute, 'Nancy Pelosi just sent out her first email fundraising off the decision overturning Roe v. Wade.' (Twitter, 24 June 2022) <<https://twitter.com/GravelInstitute/status/1540347369111392256/photo/1>> accessed 26 July 2022
- Green SJ, "'God Told Me He Was a Lizard": Seattle Man Accused of Killing His Brother with a Sword' *Seattle Times* (Seattle, Washington, USA, 8 January 2019) <<https://www.seattletimes.com/seattle-news/law-justice/god-told-me-he-was-a-lizard-seattle-man-accused-of-killing-his-brother-with-a-sword/>>
- Grelich K and others, 'Far-Right Extremists' Use Of Social Media Platforms To Communicate And Spread Radicalized Beliefs' (*The Counterterrorism Group*, 6 December 2021) <<https://www.counterterrorismgroup.com/post/far-right-extremist-use-of-social-media-platforms-to-communicate-and-spread-radicalized-beliefs>> accessed 28 July 2022
- Gruwell L, 'Constructing Research, Constructing the Platform: Algorithms and the Rhetoricity of Social Media Research.' (2018) 6 *Present Tense* 1 <<http://www.presenttensejournal.org/wp-content/uploads/2018/06/Gruwell.pdf>> accessed 30 April 2022
- Guess A, Nagler J and Tucker J, 'Less than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook' (2019) 5 *Science Advances* eaau4586 <<https://www.science.org/doi/10.1126/sciadv.aau4586>> accessed 12 May 2022
- Haas B, 'China Bans Winnie the Pooh Film after Comparisons to President Xi' *The Guardian* (7 August 2018) <<https://www.theguardian.com/world/2018/aug/07/china-bans-winnie-the-pooh-film-to-stop-comparisons-to-president-xi>> accessed 6 July 2022
- Habermas J, Lennox S and Lennox F, 'The Public Sphere: An Encyclopedia Article (1964)' [1974] *New German Critique* 49 <<https://www.jstor.org/stable/487737?origin=crossref>> accessed 1 May 2022
- Habgood-Coote J, 'Stop Talking about Fake News!' (2019) 62 *Inquiry* 1033 <<https://www.tandfonline.com/doi/full/10.1080/0020174X.2018.1508363>> accessed 29 June 2022
- Harton HC, Gunderson M and Bourgeois MJ, "'I'll Be There with You": Social Influence and Cultural Emergence at the Capitol on January 6.' [2022] *Group Dynamics: Theory, Research, and Practice* <<http://doi.apa.org/getdoi.cfm?doi=10.1037/gdn0000185>> accessed 28 July 2022
- Hauser C, 'Alex Jones Retracts Chobani Claims to Resolve Lawsuit' *The New York Times* (17 May 2017) <<https://www.nytimes.com/2017/05/17/us/alex-jones-chobani-lawsuit.html>> accessed 1 July 2022
- Hillard J, 'Social Media Addiction' (Addiction Center) <<https://www.addictioncenter.com/drugs/social-media-addiction/>> accessed 30 July 2022
- Hitlin P and Shutava N, 'Trust in Government: A Close Look at Public Perceptions of the Federal Government and Its Employees' (Partnership for Public Service and Freedman Consulting, LLC 2022) <<https://ourpublicservice.org/wp-content/uploads/2022/03/Trust-in-Government.pdf>> accessed 15 July 2022
- Horbyk R and others, 'Fake News as Meta-Mimesis: Imitative Genres and Storytelling in the Philippines, Brazil, Russia and Ukraine' <<https://aaltodoc.aalto.fi:443/handle/123456789/111278>> accessed 17 May 2022
- Hoseini M and others, 'On the Globalization of the QAnon Conspiracy Theory Through Telegram' <<https://arxiv.org/abs/2105.13020>> accessed 25 July 2022
- Hughes HC and Waismel-Manor I, 'The Macedonian Fake News Industry and the 2016 US Election' (2021) 54 *PS: Political Science & Politics* 19

<<https://www.cambridge.org/core/journals/ps-political-science-and-politics/article/macedonian-fake-news-industry-and-the-2016-us-election/79F67A4F23148D230F120A3BD7E3384F>> accessed 7 June 2022

Jaffe G and Real JAD, 'Life amid the Ruins of QAnon: "I Wanted My Family Back"' (Washington Post) <<https://www.washingtonpost.com/nation/interactive/2021/conspiracy-theories-qanon-family-members/>> accessed 24 July 2022

James L, 'Seven Cartoon Characters Peppa Pig Joined on China's Banned List' (South China Morning Post, 3 May 2018) <<https://www.scmp.com/culture/film-tv/article/2144561/after-peppa-pigs-ban-china-here-are-eight-other-cartoons-and>> accessed 5 July 2022

Jasko K and others, 'A Comparison of Political Violence by Left-Wing, Right-Wing, and Islamist Extremists in the United States and the World' (2022) 119 Proceedings of the National Academy of Sciences e2122593119 <<https://pnas.org/doi/full/10.1073/pnas.2122593119>> accessed 24 July 2022

Johannes B and Manoury C, 'Digital Services Act: Commission Welcomes Political Agreement on Rules Ensuring a Safe and Accountable Online Environment' European Commission - European Commission (Brussels, 23 April 2022) <<https://ec.europa.eu/commission/presscorner/home/en>> accessed 23 July 2022

John RR, 'Freedom of Expression in the Digital Age: A Historian's Perspective' (2019) 4 Church, Communication and Culture 25 <<https://doi.org/10.1080/23753234.2019.1565918>> accessed 28 April 2022

Jones MO, 'Analysis of #IStandWithPutin and #IStandWithRussia' (Twitter, 3 March 2022) <<https://twitter.com/marcowenjones/status/1499312091727020032>>

Karp P, 'Scott Morrison Says He Doesn't Back QAnon and Refuses to "Cancel" Supporter Friend' The Guardian (21 June 2021) <<https://www.theguardian.com/australia-news/2021/jun/21/scott-morrison-says-he-doesnt-back-qanon-and-refuses-to-cancel-supporter-friend>> accessed 1 July 2022

Kartsev D, 'The Heirs of Ancient Rus Vladimir Putin Runs Roughshod over Ukrainian History in New Essay' (Meduza) <<https://meduza.io/en/feature/2021/07/12/the-heirs-of-ancient-rus>> accessed 22 July 2022

Kemp S, 'Digital 2022: April Global Statshot Report' (Kepios 2022) <<https://datareportal.com/reports/digital-2022-april-global-statshot>> accessed 28 April 2022

Kirby EJ, 'The City Getting Rich from Fake News' BBC News (5 December 2016) <<https://www.bbc.com/news/magazine-38168281>> accessed 23 May 2022

Kizilova K and Norris P, 'Assessing Russian Public Opinion on the Ukraine War' (2022) 281 Russian Analytical Digest <<https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/RAD281.pdf>> accessed 26 July 2022

Klawier T, Prochazka F and Schweiger W, 'Public Knowledge of Alternative Media in Times of Algorithmically Personalized News' [2021] New Media & Society 146144482110210 <<http://journals.sagepub.com/doi/10.1177/14614448211021071>> accessed 30 April 2022

Klepper D, 'Russia's Information War Expands through Eastern Europe' (AP News, 15 July 2022) <<https://apnews.com/article/russia-ukraine-nato-bulgaria-misinformation-eastern-europe-0e87db7fef9263a465d6cf40d3287efe>> accessed 22 July 2022

Kruse LM, Norris DR and Flinchum JR, 'Social Media as a Public Sphere? Politics on Social Media' (2018) 59 The Sociological Quarterly 62 <<https://doi.org/10.1080/00380253.2017.1383143>> accessed 1 May 2022

Lai T and Zheng X, 'Machine Learning Based Social Media Recommendation', 2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM) (2015)

- Lankford A and Madfis E, 'Don't Name Them, Don't Show Them, But Report Everything Else: A Pragmatic Proposal for Denying Mass Killers the Attention They Seek and Deterring Future Offenders' (2018) 62 *American Behavioral Scientist* 260 <<http://journals.sagepub.com/doi/10.1177/0002764217730854>> accessed 14 July 2022
- Lee R, 'Extreme Speech in Myanmar: The Role of State Media in the Rohingya Forced Migration Crisis' (2019) 13 *International Journal of Communication* 22 <<https://ijoc.org/index.php/ijoc/article/view/10123>> accessed 28 July 2022
- Levy R, 'Social Media, News Consumption, and Polarization: Evidence from a Field Experiment' (2021) 111 *American Economic Review* 831 <<https://www.aeaweb.org/articles?id=10.1257/aer.20191777>> accessed 29 April 2022
- Lewandowsky S and van der Linden S, 'Countering Misinformation and Fake News Through Inoculation and Prebunking' (2021) 32 *European Review of Social Psychology* 348 <<https://www.tandfonline.com/doi/full/10.1080/10463283.2021.1876983>> accessed 1 July 2022
- Lewis B, 'Alternative Influence' (Data & Society, 18 September 2018) <<https://datasociety.net/library/alternative-influence/>> accessed 25 July 2022
- Lewis P and Hilder P, 'Leaked: Cambridge Analytica's Blueprint for Trump Victory' *The Guardian* (23 March 2018) <<https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory>> accessed 25 July 2022
- Lewis R, "'This Is What the News Won't Show You': YouTube Creators and the Reactionary Politics of Micro-Celebrity' (2020) 21 *Television & New Media* 201 <<http://journals.sagepub.com/doi/10.1177/1527476419879919>> accessed 26 July 2022
- Ling J, 'How 4chan's Toxic Culture Helped Radicalize Buffalo Shooting Suspect' *The Guardian* (18 May 2022) <<https://www.theguardian.com/us-news/2022/may/18/4chan-radicalize-buffalo-shooting-white-supremacy>> accessed 14 July 2022
- Linville, Tani. "Project MKULTRA and the Search for Mind Control: Clandestine Use of LSD Within the CIA." *History Capstone Research Papers*, Apr. 2016, https://digitalcommons.cedarville.edu/history_capstones/6.
- Liu J and Zhao J, 'More than Plain Text: Censorship Deletion in the Chinese Social Media' (2021) 72 *Journal of the Association for Information Science and Technology* 18 <<https://onlinelibrary.wiley.com/doi/10.1002/asi.24390>> accessed 11 July 2022
- Lubbers E, 'There Is No Such Thing as the Denver Guardian, despite That Facebook Post You Saw' (*The Denver Post*, 6 November 2016) <<https://www.denverpost.com/2016/11/05/there-is-no-such-thing-as-the-denver-guardian/>> accessed 22 May 2022
- Mamié R, Horta Ribeiro M and West R, 'Are Anti-Feminist Communities Gateways to the Far Right? Evidence from Reddit and YouTube', 13th ACM Web Science Conference 2021 (Association for Computing Machinery 2021) <<https://doi.org/10.1145/3447535.3462504>> accessed 25 July 2022
- Mandel DR, 'Radicalization: What Does It Mean?' [2009] *Home-Grown Terrorism* 101 <<https://ebooks.iospress.nl/doi/10.3233/978-1-60750-075-9-101>> accessed 24 July 2022
- McCarthy N, 'Where Concern Is Highest About Fake News On The Internet [Infographic]' (*Forbes*) <<https://www.forbes.com/sites/niallmccarthy/2019/06/12/where-concern-is-highest-about-fake-news-on-the-internet-infographic/>> accessed 7 June 2022
- McEvoy J, 'Fox News, Newsmax Walk Back Election Fraud Claims After Voting Machine Manufacturer Threatens Legal Action' (*Forbes*) <<https://www.forbes.com/sites/jemimamcevoy/2020/12/21/fox-news-newsmax-walk-back-election-fraud-claims-after-voting-machine-manufacturer-threatens-legal-action/>> accessed 1 July 2022

McFarlane G, 'How Facebook (Meta), Twitter, Social Media Make Money From You' (Investopedia) <<https://www.investopedia.com/stock-analysis/032114/how-facebook-twitter-social-media-make-money-you-twtr-lnkd-fb-goog.aspx>> accessed 25 July 2022

McGraw M and Sam S, 'It's Been Exactly One Year since Trump Suggested Injecting Bleach. We've Never Been the Same.' (Politico, 23 April 2021) <<https://www.politico.com/news/2021/04/23/trump-bleach-one-year-484399>> accessed 18 July 2022

Meadows CZ, Tang L and Zou W, 'Managing Government Legitimacy during the COVID-19 Pandemic in China: A Semantic Network Analysis of State-Run Media Sina Weibo Posts' (2022) 15 Chinese Journal of Communication 156 <<https://doi.org/10.1080/17544750.2021.2016876>> accessed 13 July 2022

Meleagrou-Hitchens A, Crawford B and Wutke V, 'Rise of the Reactionaries: Comparing the Ideologies of Salafi-Jihadism and White Supremacist Extremism' (National Counterterrorism, Innovation, and Education Center 2021) <<https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/Rise%20of%20the%20Reactionaries.pdf>>

Michael Kofman and Matthew Rojansky, 'A Closer Look at Russia's "Hybrid War"' (Wilson Center and Kennan Institute 2015) 7 <<https://www.files.ethz.ch/isn/190090/5-KENNAN%20CABLE-ROJANSKY%20KOFMAN.pdf>> accessed 19 July 2022

Mill JS, *On Liberty, Utilitarianism and Other Essays* (Mark Philp and Frederick Rosen eds, 2nd edn, Oxford University Press 2015) <<https://oxfordworldsclassics.com/view/10.1093/owc/9780199670802.001.0001/isbn-9780199670802>> accessed 4 May 2022

Milmo D, 'Rohingya Sue Facebook for £150bn over Myanmar Genocide' *The Guardian* (6 December 2021) <<https://www.theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence>> accessed 28 July 2022

Mirsky A, 'Dying to Kill The Russian Neo-Nazis Fighting Vladimir Putin's War to "Denazify" Ukraine' (Meduza) <<https://meduza.io/en/feature/2022/07/15/dying-to-kill>> accessed 22 July 2022

Moore RC and Hancock JT, 'A Digital Media Literacy Intervention for Older Adults Improves Resilience to Fake News' (2022) 12 Scientific Reports 6008 <<https://www.nature.com/articles/s41598-022-08437-0>> accessed 13 July 2022

Moravec P, Minas R and Dennis AR, 'Fake News on Social Media: People Believe What They Want to Believe When It Makes No Sense at All' (Social Science Research Network 2018) SSRN Scholarly Paper ID 3269541 <<https://papers.ssrn.com/abstract=3269541>> accessed 12 May 2022

Mundt M, Ross K and Burnett CM, 'Scaling Social Movements Through Social Media: The Case of Black Lives Matter' (2018) 4 Social Media + Society 205630511880791 <<http://journals.sagepub.com/doi/10.1177/2056305118807911>> accessed 30 May 2022

Munger K and Phillips J, 'Right-Wing YouTube: A Supply and Demand Perspective' (2022) 27 The International Journal of Press/Politics 186 <<http://journals.sagepub.com/doi/10.1177/1940161220964767>> accessed 36 July 2022

Munn L, 'Alt-Right Pipeline: Individual Journeys to Extremism Online' [2019] First Monday <<https://firstmonday.org/ojs/index.php/fm/article/view/10108>> accessed 18 July 2022

Murray JL, 'Mass Media Reporting and Enabling of Mass Shootings' (2017) 17 Cultural Studies ↔ Critical Methodologies 114

Naeem SB and Bhatti R, 'The Covid-19 "Infodemic": A New Front for Information Professionals' (2020) 37 Health Information and Libraries Journal 233 <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7323420/>> accessed 18 July 2022

- Napolitano MG and Reuter K, 'What Is a Conspiracy Theory?' [2021] *Erkenntnis* <<https://doi.org/10.1007/s10670-021-00441-6>> accessed 24 July 2022
- O'Brien C, 'How Do Social Media Algorithms Work?' (Digital Marketing Institute) <<https://digitalmarketinginstitute.com/blog/how-do-social-media-algorithms-work>> accessed 25 July 2022
- O'Dea S, 'Mobile Coverage by Technology Worldwide 2011-2027' (Statista 2022) <<https://www.statista.com/statistics/1016292/mobile-coverage-by-technology-worldwide/>> accessed 2 May 2022
- Orav A, Shreeves R and Radenovic A, 'Radicalisation and Counter- Radicalisation: A Gender Perspective' (European Parliamentary Research Service 2018) Members' Research Service PE 581.955 <<https://www.europarl.europa.eu/EPRS/EPRS-Briefing-581955-Radicalisation-gender-perspective-rev-FINAL.pdf>>
- Osman Segar N, 'As Hospitals Restrict Visitors, What Constitutes A "Good Death"?' (2022) 41 *Health Affairs* <<https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2021.01857>> accessed 16 July 2022
- Osmundsen M and others, 'Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter' (2021) 115 *American Political Science Review* 999 <https://www.cambridge.org/core/product/identifier/S0003055421000290/type/journal_article> accessed 1 July 2022
- Park CS and Gil de Zúñiga H, 'Learning about Politics from Mass Media and Social Media: Moderating Roles of Press Freedom and Public Service Broadcasting in 11 Countries' (2021) 33 *International Journal of Public Opinion Research* 315 <<https://academic.oup.com/ijpor/article/33/2/315/6031067>> accessed 30 May 2022
- Paul C and Matthews M, 'The Russian "Firehose of Falsehood" Propaganda Model Why It Might Work and Options to Counter It' (RAND Corporation 2016) <<https://www.rand.org/pubs/perspectives/PE198.html>> accessed 23 July 2022
- Pavia JR and Brack N, 'Racism in Japanese and U.S. Wartime Propaganda' (1994) 56 *The Historian* 671 <<https://www.jstor.org/stable/24449072>>
- Paz Perez E, *Strategic Disinformation: Russia, Ukraine, and Crisis Communication in the Digital Era* (ResearchGate 2022) <https://www.researchgate.net/profile/Elisa-Paz/publication/362022036_Strategic_disinformation_Russia_Ukraine_and_crisis_communication_in_the_digital_era/links/62d187f2a6abd57c6ae78140/Strategic-disinformation-Russia-Ukraine-and-crisis-communication-in-the-digital-era.pdf>
- Pennycook G and Rand DG, 'The Psychology of Fake News' (2021) 25 *Trends in Cognitive Sciences* 388 <<https://www.sciencedirect.com/science/article/pii/S1364661321000516>> accessed 12 May 2022
- Putin V, 'Article by Vladimir Putin "On the Historical Unity of Russians and Ukrainians"' (President of Russia, 12 July 2021) <<http://en.kremlin.ru/events/president/news/66181>> accessed 22 July 2022
- Radhika Gharpure and others, 'Knowledge and Practices Regarding Safe Household Cleaning and Disinfection for COVID-19 Prevention — United States, May 2020' (United States Center for Disease Control 2020) <<https://www.cdc.gov/mmwr/volumes/69/wr/pdfs/mm6923e2-H.pdf>>
- Rapp DN and Salovich NA, 'Can't We Just Disregard Fake News? The Consequences of Exposure to Inaccurate Information' (2018) 5 *Policy Insights from the Behavioral and Brain Sciences* 232 <<http://journals.sagepub.com/doi/10.1177/2372732218785193>> accessed 1 July 2022
- Reeping PM and others, 'State Gun Laws, Gun Ownership, and Mass Shootings in the US: Cross Sectional Time Series' (2019) 364 *BMJ* 1542 <<https://www.bmj.com/content/364/bmj.1542>> accessed 24 July 2022
- Reilly P, *Digital Contention in a Divided Society Social Media, Parades and Protests in Northern Ireland*. (2021)

Reneau N and others, 'Proud Boys Led Major Breaches of Capitol on Jan. 6, Video Investigation Finds' *The New York Times* (17 June 2022) <<https://www.nytimes.com/2022/07/12/us/politics/proud-boys-jan-6.html>> accessed 1 August 2022

Repnikova M, 'China's Lessons for Fighting Fake News' (Foreign Policy) <<https://foreignpolicy.com/2018/09/06/chinas-lessons-for-fighting-fake-news/>> accessed 1 July 2022

RFE/RL's Kazakh Service, 'Thousands Detained At Anti-War Protests Across Russia' RadioFreeEurope/RadioLiberty (6 March 2022) <<https://www.rferl.org/a/russia-1000-protesters-arrested-ukraine-invasion/31738786.html>>

Ribeiro MH and others, 'Auditing Radicalization Pathways on YouTube', Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (ACM 2020) <<https://dl.acm.org/doi/10.1145/3351095.3372879>> accessed 25 July 2022

Sauer P and Michaelson R, "'It Was Game over": Russian Journalists Flee to Istanbul after Putin's Shutdown' *The Guardian* (18 March 2022) <<https://www.theguardian.com/world/2022/mar/18/dozhd-tv-staff-russian-channel-in-exile-istanbul>> accessed 22 July 2022

Seetharaman JH and D, 'Facebook Executives Shut Down Efforts to Make the Site Less Divisive' *Wall Street Journal* (26 May 2020) <<https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-1159050749>> accessed 25 July 2022

Shao C and others, 'The Spread of Low-Credibility Content by Social Bots' (2018) 9 *Nature Communications* 4787 <<http://arxiv.org/abs/1707.07592>> accessed 3 June 2022

Slodysko B, 'Democrats Pull in \$80 Million Following Supreme Court Ruling on Roe' (PBS NewsHour, 2 July 2022) <<https://www.pbs.org/newshour/politics/democrats-pull-in-80-million-following-supreme-court-ruling-on-roe>> accessed 14 July 2022

Stapilus R, 'Stapilus: Idaho Voters Should Pay Attention to Candidates Espousing QAnon Beliefs' (Idaho Press) <https://www.idahopress.com/opinion/guest_opinions/stapilus-idaho-voters-should-pay-attention-to-candidates-espousing-qanon-beliefs/article__ecede09-0a7f-5373-a8b7-3168bf38becd.html> accessed 25 July 2022

Statista Research Department, 'Mass Shootings in the U.S.: Shooters by Gender 2022' (*Statista*) <<https://www.statista.com/statistics/476445/mass-shootings-in-the-us-by-shooter-s-gender/>> accessed 24 July 2022

Stempel J, 'Fox News Parent Must Face Defamation Lawsuit over Election Coverage' *Reuters* (22 June 2022) <<https://www.reuters.com/world/us/fox-news-parent-must-face-defamation-lawsuit-over-election-coverage-2022-06-21/>> accessed 1 July 2022

Storsul T, 'Deliberation or Self-Presentation?: Young People, Politics and Social Media' (2014) 35 *Nordicom Review* 17 <<https://sciendo.com/article/10.2478/nor-2014-0012>> accessed 1 May 2022

Strain L, 'The "Likes" on the Youtube Page of Buckley Wolfe' (Twitter, 10 January 2019) <https://twitter.com/travis_view/status/1083437810634248193> accessed 26 July 2022

Stratton J, "'All about a Revolution," or Duped by QAnon? Read the Interview between Doug Jensen and the FBI' (*KCCI*, 11 April 2022) <<https://www.kcci.com/article/iowan-interview-between-doug-jensen-and-the-fbi-qanon/39694179>> accessed 28 July 2022

Sullivan B, 'Highland Park Suspect's Online History Reveals a Fascination with Violence' (NPR, 5 July 2022) <<https://text.npr.org/1109844728>> accessed 27 July 2022

- Svetlova N, ‘‘The TV Is Winning’’ Many Ukrainians Now Share a Common Experience: Their Relatives in Russia Refuse to Believe Their Accounts of the War.’ (Meduza)
<<https://meduza.io/en/feature/2022/03/14/the-tv-is-winning>> accessed 22 July 2022
- Tewksbury D and Rittenberg J, *News on the Internet: Information and Citizenship in the 21st Century* (Oxford University Press 2012)
<<http://www.oxfordscholarship.com/view/10.1093/acprof:osobl/9780195391961.001.0001/acprof-9780195391961>> accessed 29 April 2022
- Thomala LL, ‘Weibo Corporation: DAUs 2022’ (Statista)
<<https://www.statista.com/statistics/1058070/china-sina-weibo-dau/>> accessed 13 July 2022
- Thompson J and Davis Thomander S, ‘What Drives Support for QAnon? Evidence from a Survey Experiment’ (University of Exeter and Stanford University 2021)
<<https://files.osf.io/v1/resources/23qaj/providers/osfstorage/6181113dce8aee05592aee2?action=download&direct&version=1>>
- Tian E, ‘The QAnon Timeline: Four Years, 5,000 Drops and Countless Failed Prophecies’ (bellingcat, 29 January 2021) <<https://www.bellingcat.com/news/americas/2021/01/29/the-qanon-timeline/>> accessed 24 July 2022
- Understanding the Iran-Contra Affairs. https://www.brown.edu/Research/Understanding_the_Iran_Contra_Affair/. Accessed 24 July 2022
- United Nations Human Rights Office of the High Commissioner, *Human Rights and Elections – A Handbook on International Human Rights Standards on Elections* (2021)
<<https://www.ohchr.org/sites/default/files/Documents/Publications/Human-Rights-and-Elections.pdf>> accessed 27 June 2022
- van Prooijen J-W, Spadaro G and Wang H, ‘Suspicion of Institutions: How Distrust and Conspiracy Theories Deteriorate Social Relationships’ (2022) 43 *Current Opinion in Psychology* 65
<<https://www.sciencedirect.com/science/article/pii/S2352250X21000828>> accessed 20 May 2022
- van Prooijen J-W, ‘Sometimes Inclusion Breeds Suspicion: Self-Uncertainty and Belongingness Predict Belief in Conspiracy Theories: Self-Uncertainty and Conspiracy Beliefs’ (2016) 46 *European Journal of Social Psychology* 267 <<https://onlinelibrary.wiley.com/doi/10.1002/ejsp.2157>> accessed 24 July 2022
- Vasilyeva A, Kommersant correspondent Anya Vasilyeva reports from Detsky Mir on Lubyanka Street: (2022)
<<https://t.me/Crexcrexcrex/3059>>
- Vergani M and others, ‘The Three Ps of Radicalization: Push, Pull and Personal. A Systematic Scoping Review of the Scientific Evidence about Radicalization Into Violent Extremism’ (2020) 43 *Studies in Conflict & Terrorism* 854
<<https://www.tandfonline.com/doi/full/10.1080/1057610X.2018.1505686>> accessed 24 July 2022
- Vinokour M, ‘Russia’s Media Is Now Totally in Putin’s Hands’ (Foreign Policy)
<<https://foreignpolicy.com/2022/04/05/russia-media-independence-putin/>> accessed 22 July 2022
- Vosoughi S, Roy D and Aral S, ‘The Spread of True and False News Online’ (2018) 359 *Science* 1146
<<https://www.science.org/doi/10.1126/science.aap9559>> accessed 2 July 2022
- Walker M and Matsa KE, ‘News Consumption Across Social Media in 2021’ (Pew Research Center 2021)
<https://www.pewresearch.org/journalism/wp-content/uploads/sites/8/2021/09/PJ_2021.09.20_News-and-Social-Media_FINAL.pdf> accessed 28 April 2022
- Walker M, ‘Americans Favor Mobile Devices over Desktops and Laptops for Getting News’ (Pew Research Center)
<<https://www.pewresearch.org/fact-tank/2019/11/19/americans-favor-mobile-devices-over-desktops-and-laptops-for-getting-news/>> accessed 29 April 2022

- Walton M, 'Defending White America: The Apocalyptic Meta-Narrative of White Nationalist Rhetoric' [2020] Theses and Dissertations <<https://scholarsarchive.byu.edu/etd/8491>>
- Wardle C and Hossein D, 'Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making' (Council of Europe Publishing) <<https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>>. Accessed 26 May 2022.
- Waszak PM, Kasprzycka-Waszak W and Kubanek A, 'The Spread of Medical Fake News in Social Media – The Pilot Quantitative Study' (2018) 7 Health Policy and Technology 115 <<https://www.sciencedirect.com/science/article/pii/S2211883718300881>> accessed 16 July 2022
- Watson A, 'Fake News Consumption Frequency in the EU by Country 2022' (Statista, 12 May 2022) <<https://www.statista.com/statistics/1076701/fake-news-frequency-europe/>> accessed 7 June 2022
- Winter, Dan Mangan, Tom. "Ghislaine Maxwell Trial: Sex Crimes Accuser Testifies Jeffrey Epstein Took Her to Meet Trump When She Was 14." CNBC, 1 Dec. 2021, <https://www.cnbc.com/2021/12/01/ghislaine-maxwell-trial-jeffrey-epstein-took-teen-girl-to-meet-trump-accuser-says.html>.
- Wiseman Few J, 'Rush to Pass "Fake News" Laws during Covid-19 Intensifying Global Media Freedom Challenges' International Press Institute (3 October 2020) <<https://ipi.media/rush-to-pass-fake-news-laws-during-covid-19-intensifying-global-media-freedom-challenges/>> accessed 18 July 2022
- Wood MJ, Douglas KM and Sutton RM, 'Dead and Alive: Beliefs in Contradictory Conspiracy Theories' (2012) 3 Social Psychological and Personality Science 767 <<http://journals.sagepub.com/doi/10.1177/1948550611434786>> accessed 24 July 2022
- Yablokov I, 'Russian Disinformation Finds Fertile Ground in the West' (2022) 6 Nature Human Behaviour 766 <<https://www.nature.com/articles/s41562-022-01399-3>> accessed 23 July 2022
- Yip M, 'Netizens Debate Bo Xilai despite Censorship' BBC News (11 April 2012) <<https://www.bbc.com/news/world-asia-china-17673673>> accessed 11 July 2022
- Yoon S and others, 'Is Social Network Site Usage Related to Depression? A Meta-Analysis of Facebook–Depression Relations' (2019) 248 Journal of Affective Disorders 65 <<https://www.sciencedirect.com/science/article/pii/S0165032718321700>> accessed 25 July 2022
- Zhu M, 'Jury Still out on China's Legal Reform and Judicial Independence' (South China Morning Post, 3 April 2021) <<https://www.scmp.com/economy/article/3128422/jury-still-out-chinas-legal-reform-and-judicial-independence>> accessed 14 July 2022
- Zuckerman E, 'QAnon and the Emergence of the Unreal' [2019] Journal of Design and Science <<https://jods.mitpress.mit.edu/pub/tliexqdu>> accessed 24 July 2022
- 'Digital Services Act: Agreement for a Transparent and Safe Online Environment' (European Parliament, 23 April 2022) <<https://www.europarl.europa.eu/news/en/press-room/20220412IPR27111/digital-services-act-agreement-for-a-transparent-and-safe-online-environment>> accessed 1 May 2022
- 'Disinformation About Russia's Invasion of Ukraine - Debunking Seven Myths Spread by Russia | EEAS Website' <https://www.eeas.europa.eu/delegations/china/disinformation-about-russias-invasion-ukraine-debunking-seven-myths-spread-russia_en?s=166> accessed 22 July 2022

- ‘Disinformation: Commission Welcomes the New Stronger and More Comprehensive Code of Practice on Disinformation’ European Commission (Brussels, 16 June 2022)
<https://ec.europa.eu/commission/presscorner/detail/en/IP_22_3664> accessed 14 July 2022
- ‘E-Commerce Directive | Shaping Europe’s Digital Future’
<<https://digital-strategy.ec.europa.eu/en/policies/e-commerce-directive>> accessed 6 July 2022
- ‘EU Imposes Sanctions on State-Owned Outlets RT/Russia Today and Sputnik’s Broadcasting in the EU’ (2 March 2022)
<<https://www.consilium.europa.eu/en/press/press-releases/2022/03/02/eu-imposes-sanctions-on-state-owned-outlets-rt-russia-today-and-sputnik-s-broadcasting-in-the-eu/>> accessed 23 July 2022
- ‘From Alt Right to Alt Lite: Naming the Hate’ (Anti-Defamation League)
<<https://www.adl.org/resources/backgrounder/alt-right-alt-lite-naming-hate>> accessed 25 July 2022
- ‘Gun Violence’ (Amnesty International) <<https://www.amnesty.org/en/what-we-do/arms-control/gun-violence/>> accessed 24 July 2022
- ‘Hate Speech and Hate Crime’ (Advocacy, Legislation & Issues, 12 December 2017)
<<https://www.ala.org/advocacy/intfreedom/hate>> accessed 2 July 2022
- ‘How Expensive Is a Smartphone in Different Countries?’ (Alliance for Affordable Internet 2021)
<<https://a4ai.org/how-expensive-is-a-smartphone-in-different-countries/>> accessed 28 May 2022
- ‘Hungary: Press Freedom Threatened as Orbán Handed New Powers’ International Press Institute (30 March 2020)
<<https://ipi.media/hungary-press-freedom-threatened-as-orban-handed-new-powers/>> accessed 18 July 2022
- ‘Increasingly Normal: Guns Seen Outside Vote-Counting Centres’
<<https://www.aljazeera.com/news/2020/11/7/increasingly-normal-guns-seen-outside-vote-counting-centres>> accessed 1 July 2022
- ‘Lisa Birgit Holst’ (Snopes.com) <<https://www.snopes.com/lisa-birgit-holst/>> accessed 15 May 2022.
- ‘Mass Shooting Data & Research’ (The Violence Project) <<https://www.theviolenceproject.org/>> accessed 24 July 2022
- ‘Myanmar Military Leaders Must Face Genocide Charges – UN Report’ (*UN News*, 27 August 2018)
<<https://news.un.org/en/story/2018/08/1017802>> accessed 28 July 2022
- ‘New PRRI Report Reveals Nearly One in Five Americans and One in Four Republicans Still Believe in QAnon Conspiracy Theories’ (PRRI)
<<https://www.prii.org/press-release/new-prii-report-reveals-nearly-one-in-five-americans-and-one-in-four-republicans-still-believe-in-qanon-conspiracy-theories/>> accessed 25 July 2022
- ‘On Limiting Access to the Social Network Instagram’ (Federal Service for Supervision of Communications, Information Technology, and Mass Media, 11 March 2022) <<https://rkn.gov.ru/news/rsoc/news74176.htm>> accessed 25 July 2022
- ‘Satisfaction With the United States’ (Gallup 2022)
<<https://news.gallup.com/poll/1669/General-Mood-Country.aspx>> accessed 24 May 2022
- ‘Tackling Disinformation Online: The Digital Services Act Opens the Era of Accountability’ (EU DisinfoLab)
<<https://www.disinfo.eu/advocacy/tackling-disinformation-online-the-digital-services-act-opens-the-era-of-accountability/>> accessed 10 July 2022
- ‘Tackling Online Disinformation: A European Approach’ (European Commission, 26 April 2018)
<<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236>>

'Trump-Russia Steele Dossier Analyst Charged with Lying to FBI' BBC News (5 November 2021)
<<https://www.bbc.com/news/world-us-canada-59168626>> accessed 26 May 2022

'Ukraine: Kyiv TV Tower Struck by Russian Missile' BBC News
<<https://www.bbc.com/news/av/world-europe-60579847>> accessed 19 July 2022