ICNL

www.icnl.org
info@icnl.org

A POLICY PROSPECTUS

Legal Responses to Disinformation

"Falsehood flies, and the Truth comes limping after it."

- Jonathan Swift

Introduction

In 2018, the Myanmar Army utilized Facebook to incite people to commit genocide.[1] During the Ebola epidemic in 2018, false claims about the Ebola virus, including that it was manufactured in a government lab, led to an attack on an Ebola evaluation center in the Democratic Republic of the Congo.[2] In the UK in April and May 2020, dozens of engineers from telecom companies were attacked, harassed and abused, and ninety cell phone towers were burned following online messages that 5G-towers contribute to spreading the virus that causes COVID-19. These are just three recent examples of disinformation campaigns and their consequences.

Disinformation is distorting politics, sowing confusion, undermining trust in democratic institutions, discrediting civil society and fueling a state of information disorder. An analysis by Buzzfeed News found that during the last few months of the 2016 U.S. presidential campaign, the 20 top-performing fake election stories generated more total engagement on Facebook than the 20 top-performing real stories from mainstream news outlets. The Reuters Institute analyzed a sample of 225 pieces of misinformation rated false or misleading by fact-checkers and published in English between January and the end of March 2020; it found that 38% of the false content was entirely fabricated and that social media is the source of 88% of all the misinformation.[3]

The goal of this briefer is to inform law and policymakers about available legal and regulatory measures to combat disinformation and empower civil society to advocate for effective regulatory efforts. The solutions presented in this prospectus seek to limit

[1] https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html
[2] https://www.sciencemag.org/news/2019/01/fighting-ebola-hard-congo-fake-news-makes-it-harder
[3] Brennen, J. et al. (2020), *Types, sources, and claims of COVID-19 misinformation*, https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation

the spread and amplification of disinformation and enable citizens to access true and false information, and then assess the validity of that information while fully respecting and protecting the freedom of expression.

The complete prohibition or elimination of disinformation is not a realistic regulatory goal. Disinformation has existed since at least the Roman Empire,[4] and various actors, from criminals to governments to opportunists, will often have an incentive to create and disseminate disinformation. As Guardian columnist Natalie Nougayrède has observed: "The use of propaganda is ancient, but never before has there been the technology to so effectively disseminate it." Therefore, a more realistic regulatory goal is to limit the spread of disinformation.

Definitions

Information disorder stems from the rise not only of disinformation, but also of misinformation and mal-information. While there are no universally agreed definitions of these terms, the elements that make-up these terms are generally accepted. For purposes of this briefer, we define disinformation, misinformation and mal-information as follows:

- *Disinformation* is the intentional dissemination of misleading and wrongful information. It is presented in such a way as to purposely mislead or is made with the intent to mislead. Put another way, disinformation is false or manipulated information that is knowingly shared to cause harm or is made with reckless disregard of likely harm. Disinformation often includes some truthful components or contains "half-truths." This makes it more difficult for the consumer to recognize something as disinformation.

  Political disinformation, or propaganda, as a subset of disinformation, is the intentional dissemination of false information seeking to shape perceptions around some aspect of political discourse.

  Disinformation = false information + intent to harm.

- *Misinformation* is the unintentional dissemination of misleading information. It is a claim that contradicts or distorts common understandings of verifiable facts, that people spread in error without intending to deceive others. Misinformation contains and describes false content, but the person sharing the content does not realize that it contains false or misleading information. Misinformation does not need to be wholly false; it can include information

[4] The late Roman emperor Caesar's heir Octavian distributed an allegedly false will from Mark Antony to Roman senators and citizens. It contained inflammatory claims, such as Antony's intention to leave legacies to his children with Cleopatra in Egypt including large pieces of Roman-held territory. This and other claims set the Roman people against Antony and enabled Octavian to become emperor Augustus Caesar; see, https://theconversation.com/the-fake-news-that-sealed-the-fate-of-antony-and-cleopatra-71287

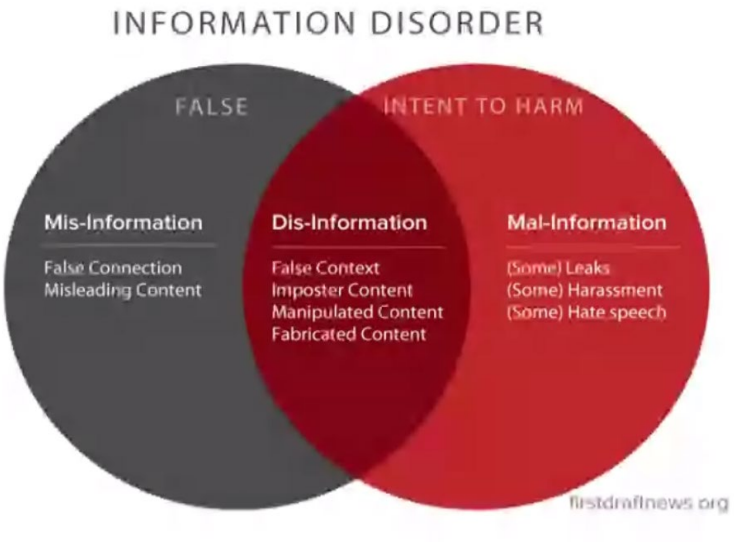
whose inaccuracy is unintentional, i.e., information reported in error.

Misinformation = false information + mistake.

- *Mal-information* is truthful information presented in deceptive ways in an attempt to mislead. Although the information is genuine, it is presented and shared in a manner intended to cause harm.

Mal-information = true information + intent to harm.

- The term "*fake news*", although widely used, has no accepted definition. It is rather a catch-all phrase for either news or information with which a given person, often someone in power, does not agree,[5] or fabricated content that is designed to look like actual news coverage of actual events. The term "fake news" is vulnerable to political manipulation; it is often deployed to mislead readers/viewers, which undermines legitimate journalism and reporting. As such, it is a term that should be avoided; it is better to define the content in question as disinformation, misinformation or mal-information.



Nature of the Problem

While not a new problem, disinformation today poses a new kind of threat because new technologies have enabled individuals and groups to spread messages faster and to a wider audience than ever before. Disinformation campaigns mobilize large numbers of individuals or groups to interact with the content, which then spurs others to share and post the content.

[5] UNESCO, *Journalism, 'Fake News' & Disinformation*, p. 43, (2018), https://unesdoc.unesco.org/ark:/48223/pf0000265552

SOCIAL MEDIA PLATFORMS

Social media platforms and internet companies cannot be relied upon to curb disinformation. Their first priority is to generate profit. Their business models thrive on engaging users with disruptive and exciting content.[6] The tools used by social media platforms, including behavioral data collection, analytics, advertising exchanges, tools for cluster detection and tracking social media sentiment, and various forms of Artificial Intelligence (AI)/machine learning, are not only integral to the use and operation of social media platforms, but are also being manipulated and harnessed by purveyors of disinformation. Algorithms that keep people engaged with the platforms and the massive collection of user data that enables micro-targeting of advertisements create a perfect ecosystem for disinformation.

At the same time, social media companies do not want to be viewed as propagators of disinformation and misinformation. Therefore, platforms including Facebook, Twitter and YouTube have developed polices and software solutions to identify and remove harmful content posted on their sites, the results of which are mixed, at best. Simultaneously, however, social media platforms continue to lobby against any changes in law that would remove the protection from liability for hosting harmful content which the platforms enjoy in the United States and in EU member states.

MESSAGING APPLICATIONS

Messaging applications like WhatsApp, Telegram and even SMS/text messaging are also used to spread disinformation. Although messaging applications do not have the same reach as social media, individuals use messaging apps to spread disinformation and misinformation, sometimes to large groups. For example, in 2018, misinformation spread via WhatsApp messages led to a mob attack and lynching in the village Murki, west of Hyderabad, in rural India.[7] Unlike social media, where posts are publicly available, recommended to other users through algorithms, and shared automatically, messaging apps carry messages from one individual to another individual or group of individuals. Even though the potential reach of disinformation and misinformation is less extensive than on social media, the results are similar.

Guarding against disinformation and misinformation on messaging apps must recognize that messaging apps have become a widespread method of communication; and that messaging apps often contain end-to-end encryption,[8] which protects

[6] See, e.g., https://www.washingtonpost.com/news/monkey-cage/wp/2018/08/06/its-no-accident-that-facebook-is-so-addictive/; https://singularityhub.com/2019/10/17/youtubes-algorithm-wants-to-keep-you-watching-and-thats-a-problem/

[7] https://www.reuters.com/article/us-india-killings/he-looked-like-a-terrorist-how-a-drive-in-rural-india-ended-in-a-mob-attack-and-a-lynching-idUSKBN1KJ09R

[8] Encryption is a mathematical "process of converting messages, information, or data into a form unreadable by anyone except the intended recipient," see, SANS Institute, "History of encryption" (2001); End-to-end encryption (E2EE) is a method of secure communication that prevents third-parties from accessing data while it's transferred from one end system or device to another.

individuals' privacy. [9] Encryption allows people to send personal identifying information to family and friends in a safer manner.[10] Weakening or prohibiting end-to-end encryption methods will impact all services that rely on it for protection of the privacy of its users and is not a viable option.[11]

GOVERNMENTAL RESPONSES

Legal and regulatory reforms have not been able to keep up with technological advances and the expanding role of social media and messaging apps. Laws have not been enacted to account for the spread of troll farms, fake accounts, bots, and other tools used by those spreading disinformation.

Many democratic governments have been reluctant to interfere in the social media landscape through regulation, sometimes out of fear of being accused of restricting free speech. Both the Government of the Netherlands, subject to disinformation related to the downing of Malaysia Airlines Flight 17, and Finland, mindful of its proximity to Russia, decided to invest in cyber security, media literacy programs, and cooperation with social media companies. Debates among policymakers increasingly revolve around how existing laws and policies fail to adequately address the specific challenges that come with the growth of social media companies, and more specifically the threat of online disinformation.[12]

Where governments have enacted laws, however, they have often proved problematic. "Anti-Fake News" laws that seek to directly counteract disinformation do so with general prohibitions on the dissemination of information based on vague and ambiguous concepts, including "false information", [13] and provide authorities with

[9] The right to privacy, as enshrined in Article 17(1) of the ICCPR, protects the right to privacy of one's communications, "[n]o one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence." Encryption provides individuals and groups with a zone of privacy to express and share their opinions. The rights to "privacy and freedom of expression are interlinked" and encryption plays a critical role in securing those rights. (United Nations Human Rights Council, A/HRC/29/32, "Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye" May 22, 2015, para. 16)

[10] For example, encryption enables a child to discuss medical issues with her parents over WhatsApp in private. Encryption is not only used for messaging apps but also to for digital services that require strong protection of privacy to prevent misuse or abuse of the exchanged information. Examples are online banking and mobile money, consulting COVID-19 test results online and exchange of sensitive diplomatic information.

[11] Even if encryption was prohibited, it would still be difficult if not impossible for governments to surveil every message sent via messaging apps in a lawful manner. See, *Case of Big Brother Watch And Others v. The United Kingdom*, European Court of Human Rights, *Applications nos. 58170/13, 62322/14 and 24960/15*, ruled that United Kingdom's bulk surveillance regime violates the European Convention on Human Rights rights to privacy and freedom of expression, the interception regime "does not meet the 'quality of law' requirement and is incapable of keeping the 'interference' to what is 'necessary in a democratic society.'"

[12] See for example these discussions about the difference between platform and publisher (https://socialmediahq.com/if-social-media-companies-are-publishers-and-not-platforms-that-changes-everything/), about intermediary liability for communications platforms (https://cyberlaw.stanford.edu/focus-areas/intermediary-liability) and anti-trust or competition law and digital services companies (https://www.americanprogress.org/issues/economy/reports/2020/07/28/488201/using-antitrust-law-address-market-power-platform-monopolies/)

[13] The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States

broad powers to act as 'arbiters of truth', in violation of Article 19 of the ICCPR. Cybercrime laws or penal codes often contain provisions criminalizing categories of speech online, including disinformation, but again do so too often with overly broad and vague language. Such laws and provisions do not comply with Article 19 of the ICCPR, and therefore impermissibly restrict freedom of expression.[14] The COVID-19 pandemic has led to enactment of emergency rules that often include high penalties or long jail sentences for spreading 'false information' or 'rumors' about COVID-19 on social media, while lacking sufficient safeguards to protect freedom of speech.

While "anti-fake news" laws have proven ill-equipped to combat disinformation, legal reform is nonetheless a necessary part of the solution. Social media platforms and other internet companies have not previously acted, and are unlikely to act, in the interest of the public without regulation by law. The challenge is to formulate legal solutions that are fully consistent with international standards relating to the freedom of expression.

Premises

Government regulation is needed to limit the spread of disinformation, misinformation and mal-information. But governments should not be the arbiters of truth, and citizens should be able to access true and false information, and then assess the validity of that information. To achieve this, regulation must comply with international norms relating to the freedom of expression and privacy. Restrictions on expression are only permissible when they satisfy each element of Article 19's test, meaning that any restrictions on expression must be provided by law and necessary in a democratic society in furtherance of legitimate government aims.[15]

We recognize that the problem of disinformation cannot be solved through legal or regulatory measures alone. Education initiatives, technological advances and other forms of multi-stakeholder initiatives will be needed. The focus of this paper, however, is on legal and regulatory measures that may, to some extent, help curb disinformation while still respecting the freedom of expression.

(OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on "Fake News", Disinformation and Propaganda*, para. 2(a) (March 3, 2017).

[14] See for these types of laws in Sub-Saharan Africa and analyses: www.disinformationtracker.org.

[15] See, e.g. United Nations Human Rights Council, A/HRC/17/27, "Report of UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue" May 16, 2011, para. 69. (1) the restriction must be provided by law, which is clear and accessible to everyone (i.e., adheres to principles of predictability and transparency); (2) the restriction must pursue one of the purposes set out in article 19(3) of the ICCPR, namely: (i) to protect the rights or reputations of others; (ii) to protect national security or public order, or public health or morals (principle of legitimacy); and (3) the restriction must be proven as necessary and the least restrictive means required to achieve the purported aim (i.e., adheres to principles of necessity and proportionality).

Legal and Regulatory Responses

We now turn to three different categories of laws that governments may adopt and apply to limit the spread of disinformation:

1. Existing laws that, while not designed to combat disinformation, could be helpful tools to do so;
2. Laws targeting disinformation or some aspect of disinformation, recently enacted in a limited number of jurisdictions, that may be replicated in other jurisdictions; and
3. Newly proposed legal and regulatory approaches that, to our knowledge, have not yet been enacted anywhere, but may provide innovative solutions.

1. Existing Laws

The existing laws in many countries may be an important part of the legal response to disinformation. While not necessarily designed to address disinformation in the digital realm, they could be applied to combatting disinformation, and in some cases have been, as highlighted below. A recent report highlighted that in the Netherlands, a large number of types of disinformation already fall partly under an existing legal category, such as misleading advertising, libel or unlawful press publication, and as such are already regulated.

1.1. TORT LAW

Tort laws[16] can be used to provide reparations to victims and serve as a deterrent for engaging in disinformation. As most countries have tort laws, the causes of action outlined in this section could be utilized without needing to enact new laws, although amendments to existing tort law may be needed to ensure coverage for harms caused by disinformation.

1.1.1. Intentional infliction of emotional distress / Intention to cause harm

In the US, the intentional infliction of emotional distress (IIED) provides an avenue to address the harmful impact of disinformation. IIED occurs when one acts abominably or outrageously with intent to cause another to suffer severe emotional distress, such as issuing a threat of future harm. Although courts usually refrain from assigning liability under IIED to a person who speaks harmfully about public figures, two recent examples highlight how this cause of action can impose liability against those who wage disinformation campaigns against individuals.

[16] We use "tort" here to mean a "civil" or personal wrong to a private individual or enterprise, and the legal remedy for a tort victim lies in the injured party bringing a private lawsuit against the injurer. Most countries allow for such lawsuits, but depending on the jurisdiction and legal system (Common Law or Civil Law), the term "tort" might be replaced by another term or title, for example "Liability in Damages," Civil Code of Germany § 823.

In the first example, a publisher of a neo-Nazi website was ordered to pay more than $14 million in damages for encouraging "an online anti-Semitic harassment and intimidation campaign" against a woman in 2016. A federal magistrate judge in Montana found the publisher of the neo-Nazi website liable under IIED, invasion of privacy and Montana's Anti-Intimidation Act (cyber-bullying).

In the second example, Taylor Dumpson, the first black woman to serve as student government president at American University (AU), sued the founder and editor of the Daily Stormer, Andrew Anglin, in 2017. The Daily Stormer instituted a vitriolic campaign against Ms. Dumpson. Ms. Dumpson sued under IIED because the Daily Stormer's campaign against her "interfered with her enjoyment of places of public accommodation because she no longer felt safe on the AU campus." The U.S. District Court for the District of Columbia agreed with Ms. Dumpson's argument that because the AU campus is accessible to the public, it should be considered a "public accommodation." Racist online trolling activity can interfere with one's equal access to a public accommodation. She was awarded $700,000 in damages.

1.1.2. Defamation (Libel and Slander)

Defamation is a statement that injures a third party's reputation. This tort includes both libel (written statements) and slander (spoken statements).

Three examples in the US demonstrate the use of defamation to combat disinformation. First, Leonard Pozner, whose 6-year-old son Noah was killed in the 2012 Sandy Hook massacre, was awarded $450,000 in damages by a Wisconsin jury from a defamation lawsuit filed in response to conspiracy theorists claiming the Sandy Hook tragedy never occurred.

Second, comedian and writer Dean Obeidallah was awarded $4.1 million in a lawsuit he filed against The Daily Stormer for libel and intentional infliction of emotional distress due to the Daily Stormer falsely claiming that Mr. Obeidallah was a terrorist who had masterminded the bombing in Manchester, England in May 2017.

Third, in January and February 2021, Dominion Voting Systems and Smartmatic Corporation filed defamation suits seeking $1.4 billion and $2.7 billion in damages, respectively, against several people and entities, including Fox News and Rudy Giuliani, for allegedly coordinating and executing a disinformation campaign centered on claiming widespread fraud in the 2020 US Presidential election.[17]

Using defamation in response to disinformation is not limited to the United States. In South Africa, three journalists lodged a complaint in May 2018 against the British public relations firm, Bell Pottinger, for defamation, alleging that Bell Pottinger conducted a

[17] https://www.npr.org/2021/01/25/960302843/dominion-voting-systems-suing-giuliani-for-defamation; https://www.bloomberg.com/news/articles/2021-02-04/fox-news-faces-2-7-billion-defamation-case-for-election-disinfo; https://www.nytimes.com/2021/02/06/business/media/conservative-media-defamation-lawsuits.html

disinformation social media campaign that used a slew of bots[18] and other fake accounts to portray the journalists as biased, having no integrity and engaging in fake news.[19] As of February 2021, this case has not yet been decided.

In India, in 2018, Fatima Nafees, the mother of a Jawaharlal Nehru University (JNU) student, Najeeb Ahmed, who went missing in October 2016, filed a civil defamation suit in the Delhi High Court against certain media houses for linking her son with the terrorist organization ISIS.[20] In the story, journalist Raj Shekhar Jha, based on conversations with anonymous sources, claimed Ahmed's internet browsing history showed that he was looking for information on ISIS's "ideology, execution and network," and that his searches included "ways to join ISIS" and other similar inquiries.[21] The Delhi High Court demanded that all defamatory content be taken down.[22]

In Finland, a court sent a man to prison for harassing Jessikka Aro, a reporter well-known for exposing pro-Kremlin disinformation activities. Several men were found guilty of defamation and harassment, including the founder of a right-wing, pro-Kremlin website. He and others had targeted Aro for years by sending her text messages impersonating her father, publishing stories claiming she was drug addict, disclosing her personal information and issuing death threats.[23]

1.1.3. Unlawful act

In the Netherlands, tort law provisions allow courts to adjudicate whether actions and expressions on social media are lawful or not. This has been used to hold a social media company responsible for the spread of a campaign intended to do financial harm using false information.

On November 11, 2019, a court in the Netherlands held Facebook liable for authorizing the publication of false Bitcoin ads with John de Mol, a famous European media entrepreneur and TV producer, leading to €1.7 million in damages of victims. The court held that Facebook is obligated to take all necessary and reasonable measures to prevent and discourage unlawful ads and that Facebook's ad review systems and preventive measures were not sufficient.[24]

---

[18] Bots are autonomous programs on a network, usually the Internet, that can interact with computer systems or users, and are generally designed to respond or behave like humans.
[19] https://www.leighday.co.uk/News/News-2018/May-2018/South-African-journalists-launch-UK-defamation-act ; https://www.politicsweb.co.za/news-and-analysis/sa-editors-launch-defamation-claim-against-now-def; *see also*, https://www.ft.com/content/4702368a-5d22-11e8-ad91-e01af256df68
[20] https://2019.hrln.org/high-court-issues-notice-to-media-houses-in-the-defamation-case-filed-by-fatima-nafees-najeebs-mother-bringbacknajeeb/
[21] https://thewire.in/law/delhi-hc-directs-media-to-remove-news-videos-linking-jnu-student-najeeb-ahmed-to-isis.
[22] https://www.sabrangindia.in/article/media-outlets-forced-remove-false-and-defamatory-news-about-najeeb-ahmed
[23] https://www.dw.com/en/court-in-finland-finds-pro-kremlin-trolls-guilty-of-harassing-journalist/a-45944893
[24] https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBAMS:2019:8415&showbutton=true&keyword=facebook

In Brazil, in 2013, Giovanna Lancellotti, an actress, filed a lawsuit against Facebook requesting it to delete 59 false profiles and 10 groups which hosted offensive and harmful content. The lawsuit was filed only after Lancellotti was unsuccessful in her attempts to work with Facebook directly to remove the profiles and groups. The Court reasoned that the relationship between Lancellotti and Facebook Brazil was comparable to that between customer and user and, as such, consumer protection law could be applied. Accordingly, Facebook Brazil's service became defective as soon as Lancellotti's privacy and image were harmed, entitling her to compensation.[25]

1.2. CYBER-BULLYING/CYBER-STALKING

Cyber-bullying and cyber-stalking laws are similar to anti-harassment laws. These statutes prohibit harassing individuals online. They have become more prevalent in recent years. The criminalization of cyber-bullying or cyber-stalking may exist as a stand-alone law, as part of a country's criminal code, or as part of a cybercrime law. For example, the US has a federal cyber-stalking law, 18 U.S. Code § 2261A, which makes it illegal for anyone with the intent to kill, injure, harass or intimidate to use an electronic communication service that places that person in reasonable fear of the death of or serious bodily injury or would be reasonably expected to cause substantial emotional distress to a person.

Australia's Criminal Code, section 474.17, makes it a crime to use "a carriage service to menace, harass or cause offence." This provision was used against an individual for using Facebook to threaten to rape an acquaintance.[26] In the United Kingdom, cyber harassment can be prosecuted under the Protection from Harassment Act of 1997 or the Malicious Communications Act of 1988. Similarly, Singapore's Protection from Harassment Act 2014 criminalizes harassment, cyber-bullying and unlawful stalking, among other things.

In Ireland, a law amending the Harassment, Harmful Communications and Related Offences Bill from 2017 was passed in December 2020. This law is primarily meant to criminalize the sharing of intimate images without consent and also includes the offence 'distributing, publishing or sending threatening or grossly offensive communications' with the intent to cause harm. It appears that this law could impose liability on people who conduct disinformation campaigns, whether or not with the use of (fake) pornographic images.

These laws, however, are sometimes abused. For example, in Uganda, human rights activist Stella Nyanzi was charged under the Computer Misuse Act of 2011, and subsequently detained for allegedly "cyber-harassing" President Museveni because she called him a "a pair of buttocks." Therefore, to be effective, such laws must be narrowly

[25] https://globalfreedomofexpression.columbia.edu/cases/lancellotti-v-facebook/
[26] https://www.theguardian.com/australia-news/2015/dec/08/man-who-allegedly-made-online-threats-expected-to-plead-not-guilty

drafted and implemented based on objective standards. Otherwise, such laws invite arbitrary and subjective decision-making, which is likely to lead to violation of the freedom of expression.

1.3. FRAUD

Fraud statutes may exist for both offline and online "fraud." Depending on the definition in a specific jurisdiction and the content that was shared, fraud statutes could be used to punish actors that intend to harm via false information. For example, if an actor impersonates someone else or forges a document as part of a disinformation campaign, those actions may run afoul of fraud statutes.

2. Laws Targeting Disinformation

The two laws discussed in this section are recently enacted or draft laws that are intended to directly combat the rise of disinformation. While perhaps premature to judge the effectiveness of these laws, they are worth watching. If effective, these laws could be replicated in more jurisdictions to limit the spread of disinformation.

2.1. ANTI-BOT LAWS

A piece of false information is not effective in reaching its purposes if the information is not amplified and read by a large number of readers. The actors pushing disinformation use various methods to amplify the information, often using so-called automated bots, such as fake accounts on social media that are programmed to look human and cause a certain message. Another method is to have many people running multiple accounts to amplify certain information in a coordinated way, so-called hybrid campaigns. Reducing the influence of these types of operations has become a major focus for social media companies. Anti-bot laws can be used to limit the spread of disinformation because these laws make it more difficult to push content through bots.

California became the first state in the United States, and possibly the first jurisdiction in the world, to try to reduce the power of bots through an "Anti-bot law."[27] The law requires that bots (or the person controlling them) reveal their "artificial identity" when they are used to sell a product or influence a voter. The law defines a "bot" as "an automated online account where all or substantially all of the actions or posts of that account are not the result of a person." This definition ensures that use of simple technological tools like vacation responders and scheduled tweets will not be unintentionally impacted.

The Anti-bot law makes it illegal "for any person to use a bot to communicate or interact with another person in California online, with the intent to mislead the other person about its artificial identity for the purpose of knowingly deceiving the person about the content of the communication in order to incentivize a purchase or sale of goods or

[27] The "Bolstering Online Transparency," or B.O.T. Act.

services in a commercial transaction or to influence a vote in an election." The only exception is where the person discloses its use of the bot in a manner that is "clear, conspicuous, and reasonably designed to inform persons with whom the bot communicates or interacts."

The law targets large platforms—those with 10 million or more unique monthly United States visitors. This limitation seems largely appropriate: limiting the law to large platforms ensures that it will not unduly burden small businesses or community-run forums.

The law expressly provides that it "does not impose a duty on service providers of online platforms, including, but not limited to, Web hosting and Internet service providers." Without this limitation, internet platforms would have been in a very difficult situation of trying to ascertain whether or not a user was a "bot." This would have likely resulted in censorship of speech, especially as bad actors learned to cheat the system, as in Copyright/Digital Millennium Copyright Act (DCMA) notice and take-down requests.[28]

This law aims to solve problems caused by bots deployed at scale on large platforms and could provide an interesting model for other states in the U.S. – and other countries – to consider. At the same time, however, the law has been criticized for its ambiguity[29] and its ultimate effectiveness is still unknown since it was enacted only recently, in July 2019.

2.2. TRANSPARENCY LAWS

Disinformation is often disguised. The entity producing the troublesome content is not known. These entities, sometimes with ties to governmental agencies, mask their identities. Transparency laws aim to make social media users aware of where content comes from and which entity is supporting the production and publication of that content.

Enabling users to understand where content originates from or who paid for it is particularly relevant during elections and other political events, as online advertising strategies have become a crucial part of political campaigns to influence public debate.

Some types of transparency measures already exist. For example, the United States, France, and Ireland require social network companies to collect and disclose information to users about who paid for an advertisement or piece of sponsored content, and to share information about the audience that advertisers target. Under these laws, all digital producers and disseminators are legally required to identify

[28] The Digital Millennium Copyright Act (DMCA) is a 1998 United States copyright law that implements two 1996 treaties of the World Intellectual Property Organization (WIPO). The DMCA created a notice and takedown process that copyright holders use to have user-uploaded material that infringes their copyrights removed from websites. In recent years, authoritarian governments have used the DMCA to target civil society organizations and human rights defenders that criticize government official and policies, see, e.g., https://www.eff.org/deeplinks/2014/05/state-censorship-copyright-spanish-firm-abuses-DMCA

[29] https://www.wired.com/story/law-makes-bots-identify-themselves/

themselves and their beneficial owners on platforms and sites in clear terms. The precise details differ by law and jurisdiction, but these laws are generally part of existing consumer protection laws and/or data privacy laws.

Political advertising has traditionally been an area where many governments have taken a hands-off approach due to the importance of protecting political speech. After the 2016 US Presidential election, however, US lawmakers realized that disinformation was causing harm to the democratic, political process. In response, the US Congress drafted the Honest Ads Act.

Introduced in the US Senate, the Honest Ads Act is the furthest the United States has gone in addressing political disinformation. The Act would require digital platforms that have more than 50 million unique monthly visitors to maintain a record of advertisers who have spent more than $500 on advertisements in the previous year. This record would be required to be made public and include a copy of the advertisement itself.

The Act addresses a loophole in existing campaign finance laws, which regulate television and radio advertisements, but not internet advertisements. This loophole has allowed foreign entities to purchase online ads that mention political candidates. The Honest Ads Act would help close that loophole by subjecting internet advertisements to the same rules as advertisements aired on television and radio. The Act would also increase overall transparency by allowing the public to see who bought an online political advertisement, no matter who that entity is.

In the EU, the European Commission published two bills, the Digital Services Act (DSA) and the Digital Markets Act (DMA) in December 2020, which are expected to be enacted within two to four years. These bills contain a range of transparency requirements for companies that host content, including social media companies that operate in the EU market. The aim of these bills is to reduce the spread of illegal content and to address the drivers of disinformation campaigns.

The DSA bill contains requirements for social media companies on content moderation practices, including automated processes, and rules for notice and action procedures. The DSA also outlines transparency requirements regarding the disclosure of the entity or entities supporting targeted or political advertisements. Finally, the DSA includes provisions enabling users to opt out of surveillance-based advertisements.

The DMA bill sets rules for how very large companies collect, use and share the data of their users. This is likely to have a major impact on the advertising models of social media companies and other companies that rely on and purchase the data collected by social media companies.

Both bills seek to strengthen the oversight procedures in the EU and subject wrongful conduct to fines, as well as to sanctions like break-ups of companies (DMA) and take-downs of an entire platform-based service (DSA).

Additionally, the European Democracy Action Plan outlines the efforts underway to improve the existing EU toolbox for countering foreign interference. This policy document also states the ambition to overhaul the voluntary Code of Practice on Disinformation into a co-regulatory framework of obligations and accountability of online platforms in line with the abovementioned DSA.

In addition to the DSA and DMA, every member state in the EU is responsible for national elections and therefore can regulate political advertising within their borders. Member States of the EU have adopted a national approach, such as the 2018 law in France, which requires large social media companies, such as Facebook, Twitter and YouTube, to adhere to the following conduct during the three months preceding general elections:

- Provide users with "honest, clear and transparent information" about the identity and corporate address of anyone who paid to promote informational content related to a "debate of national interest;"
- Provide users with "honest, clear and transparent information" about the use of personal data in the context of promoting content related to a "debate of national interest;" and
- Make public the total amount of payments received for the promotion of informational content when these amounts are above a certain threshold.[30]

The full effects of this law are not yet known, but thus far results have been mixed. Most notably, Twitter banned the French government-sponsored #OuiJeVote (Yes, I Vote) campaign, which encouraged voters to register for the European elections, because it appeared to violate the law's advertising transparency standards.[31]

Transparency laws for political advertising may help combat disinformation, but there are likely to be issues regarding the laws' reach and scope, and we need to learn more regarding their implementation and impact.

3. Proposed Regulatory Responses

We now turn to novel ways to address disinformation – that is, proposed regulatory responses that have not yet been enacted or implemented. The goals of each of these newly proposed regulatory responses is to combat disinformation while protecting the

[30] Law No. 2018-1202 of 22 December 2018 Regarding the Fight Against Information Manipulation, https://www.legifrance.gouv.fr/affichTexte do;jsessionid=3EA914DFE69980E3FBB01324A666B5D1.tplgfr22s_1?cidTexte=JORFTEXT000037847559&categorieLien=id

[31] https://www.bbc.com/news/world-europe-47800418

freedom of expression. The regulatory solutions outlined below all seek to impose mandatory requirements on or create accountability mechanisms for social media platforms and messaging apps.

Social media platforms are hosts of user-generated content rather than producers of content and, as intermediaries, are largely shielded from liability when content on their platforms violates the law. Platforms have used the argument that any kind of government regulation would stifle freedom of speech and claim that self-regulation is more effective to combat disinformation and other harm. However, the increased pressure on companies to prevent harm more effectively has led to Mark Zuckerberg calling for government regulation over social media platforms.[32]

His remarks were widely taken as an overarching effort to avoid liability for Facebook for actions perceived by some as criminal and negligent.[33] Governments around the globe are aiming to set standards for technology and social companies that align with the deep impact these companies have on our daily lives. As mentioned, the European Commission is currently drafting a new legislative proposal, the Digital Services Act, that aims to establish a regulatory framework to help address the threat of disinformation.

3.1. REQUIREMENT TO UPHOLD TERMS OF SERVICE OR COMMUNITY STANDARDS

As part of the conditions of membership to use their services, social media platforms require users to acknowledge the right of the company to restrict a users' speech and abide by the rules set by the social media platform. These rules are known as "terms of service"[34] or "community standards."[35] However, social media companies implement their terms of service or community standards in arbitrary and subjective ways,[36] and without any transparency. As a result, disinformation is allowed to flourish, especially from popular accounts.

Social media companies should take a stronger stance against disinformation and uphold internal policies to tackle harmful untruths spreading across the platform. The law can assist, by requiring that social media companies develop and implement their

[32] Mark Zuckerberg, *The Internet needs new rules. Let's start in these four areas*., March 19, 2019,"Every day we make decisions about what speech is harmful, what constitutes political advertising, and how to prevent sophisticated cyberattacks. These are important for keeping our community safe. But if we were starting from scratch, we wouldn't ask companies to make these judgments alone. I believe we need a more active role for governments and regulators. By updating the rules for the internet, we can preserve what's best about it – the freedom for people to express themselves and for entrepreneurs to build new things – while also protecting society from broader harms." https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html

[33] https://www.latimes.com/business/story/2019-07-24/facebook-ftc-facebook-5-billion-fine

[34] Terms of Service = the set of rules and regulations a provider attaches to a software service or Web-delivered product.

[35] Community Standards = This used by various types of social media platforms to govern what content it will allow on its platform. Facebook is the most famous example: https://www.facebook.com/communitystandards/. Many other sites that have interactive comment sections will also create such standards.

[36] https://www.nytimes.com/2020/09/18/technology/facebook-tried-to-limit-qanon-it-failed.html

terms of service or community standards in an open and transparent way, with oversight from government or third-party regulators.

Indeed, the anticipated EU Digital Services Act may seek to increase the legal responsibility of large online platforms to such an extent that the costs for the failure to enforce internal rules will be greater than the income generated by allowing false information to circulate.

3.2. INDEPENDENT REGULATORY AGENCY

States could establish accountability mechanisms to oversee how social media companies create and implement policies relating to whether content is permissible. Specifically, states could establish an independent regulatory agency for this purpose.

As a reference point, in the EU, regional agencies have been established to support the implementation of EU law. The European Data Protection Board (EDPB) ensures that the data protection law is applied consistently across the EU and can be called upon to make binding decisions on disputes regarding cross-border processing of data.[37] The European Regulators Group for Audiovisual Media Services (ERGA) consists of representatives of national independent regulatory bodies and works to ensure consistent implementation of audiovisual media law.[38]

Similarly, a regulatory agency could be empowered to ensure that social media platforms are complying with either their own internal policies or national laws on issues like fact-checking, advertisement disclosures, use of bots, due diligence, consumer responsibilities and worker protections. Social media companies struggle to respond appropriately to disinformation campaigns, or 'coordinated inauthentic behavior' as Facebook refers to cross-border information operations that are often politically motivated and presumably run by or on behalf of state actors. [39] The platforms currently are subject to little judicial scrutiny, because courts and adjudicators are ill-equipped to deal with the specific nature of pieces of disinformation spreading with tremendous speed and across multiple borders.

The advantage of an independent regulatory agency is that it need not require individual injury. In other words, the agency could be tasked with overseeing social media platforms regardless of whether a person or group of people suffers a specific injury. It is likely that the agency would also need investigatory powers, as the agency would be able to act more quickly than a traditional court. The law establishing such an independent, regulatory agency would likely address a range of issues, including, but

[37] https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/enforcement-and-sanctions/enforcement/what-european-data-protection-board-edpb_en
[38] ERGA's first contribution to upcoming debates about the Digital Services Act, https://erga-online.eu/wp-content/uploads/2020/06/PressRelease1_ERGA_Plenary_June2020_published.pdf
[39] https://about.fb.com/news/2020/07/removing-political-coordinated-inauthentic-behavior/

not limited to, the structure of the agency, its available powers (including investigation and enforcement), and the accountability of the agency.

Of course, there are risks in creating such a regulatory agency. If not sufficiently independent or staffed with appropriate personnel, this agency could lead to government overreach or unwarranted interference in social media platforms and curtail the freedom of expression. In Tanzania, for example, the Tanzania Communications Regulatory Authority has broad powers to control and prohibit what information is published on the internet and has used that power to imprison bloggers that criticized government policies. In Nepal, the Press Council is an independent statutory and quasi-judicial authority that was established in 1970 with one of the objectives to take necessary action if found violation of journalist code of conduct. This council has shut down 17 online news portals for allegedly publishing disinformation related to Covid-19 in April 2020.

The efficacy of such a regulatory approach will be greatly dependent on staffing and the extent to which a so-called independent agency can truly act independently – that is, free of political interference, and guided solely by objective standards. In countries not sufficiently bound by a rule-of-law tradition, the creation of such an agency may simply serve to undermine the freedom of expression.

3.3. ADMINISTRATIVE TRIBUNALS

An alternative approach would be to establish administrative tribunals to hear (newly created) private rights of action with statutory damages (to avoid necessity of demonstrating pecuniary harm) for instances of platforms violating regulatory requirements. The statutory damages could be linked to the number of platform users, as this would create an effective deterrent.

The advantages of administrative tribunals to hear these claims in the first instance (versus reliance on the court system) would be: (a) to avoid overwhelming the court system with new claims; (b) to enable adjudicators to have adequate expertise on disinformation and related technical issues; and (c) to authorize a national-level law enforcement agency to bring these suits, with damages awarded invested in education, etc.

The concern is whether an administrative tribunal could operate independently, competently, and free from political interference.

3.4. COMPLAINT AND REVIEW MECHANISMS

A fourth potential regulatory solution would be to require platforms with a minimum number of users to establish a transparent complaint-and-review or notice-and-action mechanism for content, to enable platform users to submit a complaint about certain

content for follow-up. An example of this is the NetzDG law in Germany.[40] In order to be effective and to enhance the platforms' accountability, independent and transparent review of the complaint mechanisms and their operation should be enabled by law.

The mechanism would be triggered through the submission, by any user, of a complaint requesting review of content on a variety of designated bases.[41] The individual or entity posting the content could then be given an appropriate, but time-limited, opportunity to respond. Where the poster of the questioned content responds, the law could envision one of three approaches: (a) the platform could then be required to conduct an investigation; or (b) the complaint could be referred to a private review mechanism set up and funded by platforms with government oversight; or (c) the complaint could be referred to the relevant body for adjudication.

Where the poster of questioned content does not respond to a complaint, or review of a complaint determines that the complaint is well-founded and the content violates either the law or terms of service, the content would either be removed, or platforms would be required to prominently tag the content as disputed.

3.5. OBLIGATORY INVESTMENTS IN EDUCATION, LITERACY, AND HYGIENE OF USERS

Education and literacy are widely acknowledged as crucial to be able to navigate the information space and recognize disinformation. The responsibility to increase digital hygiene, integrate digital skills in education programs and generate awareness and increase literacy is left to citizens, educational institutes, civil society and government agencies. It is the large online platforms, however, that provide the main channels to spread and amplify disinformation and this brings a responsibility to equip users with the necessary skills to use the platforms responsibly.

Because these investments do not necessarily generate income for platforms, they may need to be compelled to do this by law. As one option, social media companies could be taxed to support education initiatives. For example, social media companies could be required to pay a percentage of advertising revenue into an education fund; this fund could then be used by government agencies to create and implement courses on media literacy for all segments of the population.

3.6. TRANSPARENCY REQUIREMENTS FOR SOCIAL MEDIA CONTENT MODERATION

Social media companies should provide information regarding the source and truthfulness of content. Disinformation only becomes an effective campaign if the false

[40] See Tworek & Leerssen, Transatlantic Working Group, An Analysis of Germany's NetzDG Law (April 15, 2019), https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf

[41] At its narrowest, this could be limited to hate speech and incitement; at its broadest, this could extend to any false information.

information, created and published with the intent to do harm, spreads across the internet and finds an audience.

First, for example, a Facebook page was named the "Concerned Sons of Texas." This page had a large following and consistently published disinformation. However, it was actually run out of a Russian troll farm. If Facebook properly labelled it as a Russian account, would it have so effectively disseminated disinformation?[42]

Second, social media platforms can engage in robust fact-checking. Under this program, platforms would disclose information relevant to evaluating the credibility of information. These extensive disclosures would be sent to particular regulators or expert research groups who might then work to enforce rules and inform the public at large without amplifying the false content. ICNL is aware of several types of fact-checking initiatives already underway and note that while fact-checking can be helpful, research has shown that it is not a panacea.

3.7. LIMIT DISINFORMATION ON MESSAGING APPS

Social messaging apps are generally used to communicate between individuals or groups of relatively small sizes. Although disinformation and misinformation can be spread in smaller groups, its impact will of course be much larger if shared with and among large groups. Some regulatory actions can be taken to limit the spread of false information via communications messengers.

3.7.1. Limit the Amounts of Forwards or People per Chat

Disinformation and misinformation are often spread by people forwarding a message they receive in one chat to several people in several other chats. WhatsApp introduced a limit of 5 forwards one person can make per message, presumably with the intent to raise the barriers for unwittingly spreading disinformation by requiring more action by the user and allowing time to recognize information as fake. Currently, WhatsApp allows up to 256 people per chat. This creates an environment where a large number of people can be exposed to disinformation and misinformation. Laws could limit the number of times a message can be forwarded or the number of people allowed into a specific chat.

3.7.2. Automatic Message on Forwards

Messaging apps could require that an automatic message be shown before a message is forwarded and/or after forwarded message is received. For example, WhatsApp has this message on its COVID page: "Think about the messages that you receive, because not everything you receive about coronavirus may be accurate. Verify the facts with other trusted official sources, fact checkers, or via the International Fact-Checking Network (IFCN) fact checking chatbot at +1 (727)291-2606. If you aren't sure something's

[42] This type of disclosure is different from "real name requirements," which go against international norms for the freedom of expression; the ability to speak anonymously is part of the freedom of expression.

true, don't forward it."[43] This type of message could be automatically included with any forwarded message.

3.7.3. Complaint Mechanism for Disinformation

Governments and messaging apps could set-up complaint or flagging mechanisms so that when someone receives a message containing what he or she believes to be disinformation or misinformation, that person could forward the suspected message or messages to a body that seeks to moderate disinformation and misinformation. This body could be either government-led or platform-led. For example, a law could require every message app with a certain number of users to form a "Disinfo Account" where users can "report" any suspected messages, images or videos that they've received containing disinformation or misinformation. This account could then verify the message – as with fact-checking – and, if found to be disingenuous, alert the original sender. If users are habitually sending messages that are verified to be disinformation or misinformation, they could have their accounts blocked.

Conclusion

No single solution can serve as a panacea to end disinformation. Rather, several different types of legal and regulatory responses need to be implemented to address the rise and spread of disinformation. The legal and regulatory framework relating to the freedom of expression on the internet has not kept pace with the technological advances that have created new avenues for disinformation and misinformation to spread. This Policy Prospectus has outlined existing laws in many countries that can be utilized to stem disinformation, discussed new laws and regulations that are being enacted to curb the spread of disinformation, and examined some measures that social media companies should adopt, either voluntarily or via legislation, to respond to disinformation on their platforms. Although the prospectus has not addressed every potential solution, we hope it will be useful for governments and civil society to advance meaningful reforms that address disinformation while protecting the freedom of expression.

[43] https://www.whatsapp.com/coronavirus/?lang=fb