UNIVERSITY OF EAST ANGLIA

LAW SCHOOL

Dissertation Module Code **:** LAW – 7000X

Module Title **:** LLM Dissertation

LLM Specialism **:** LLM In Information Technology and Intellectual Property Law

Candidate Name **:** Aysenur Ozsevik

Dissertation Title **:** How and to What Extent Do Social Media Platforms' Responses to Combating Disinformation Affect Freedom of Expression On Social Media?

Candidate ID Number **:** 100272777

Submission Date **:** 31.01.2023

Word Count **:** 9841

Acknowledgement

I would like to thank the Republic of Turkey Ministry of Education and the Information and Communication Technologies Authority of Turkey (“BTK”). for their financial assistance and funding during my master's degree.

I am also grateful to my family and my dear cousin Tugce, who have always supported me morally during this difficult process.

Table of contents

Abstract

Social media companies play a crucial role in combating disinformation, but their methods are often criticized for violating freedom of expression. This dissertation explores the methods used by social media companies to combat disinformation and the debates around their potential to violate freedom of expression. It also provides recommendations for ensuring that freedom of expression is protected while effectively combatting disinformation. The recommendations include a strong legal framework, promoting media literacy, ensuring transparency and accountability in decision-making, collaboration among stakeholders, and investment in technology. The dissertation concludes that a multi-faceted approach is necessary to effectively combat disinformation while also protecting freedom of expression.

1- Introduction

It can be said that social media platforms come into our lives with rising Web 2.0 technology is playing a major role in spreading disinformation. Notwithstanding, in response to the rise of disinformation through social media, again social media platforms have responded with different regulatory solutions in attempts to minimise or eliminate the adverse impact of information disorder but their efforts to address the problem have been woefully inadequate. Moreover, we witness on social media that the responses against disinformation have negative impacts on the right to freedom of expression, which is one of the basic human rights and even violates this right.

There is growing evidence that disinformation tends to thrive where the right to freedom of expression is constrained and where media quality, diversity and independence are weak. Conversely, where freedom of expression is protected, civil society, media members and all other people can challenge disinformation and present alternative perspectives. That makes the right to freedom of expression a powerful key of the situation for addressing disinformation. Consequently, disinformation cannot be addressed in the absence of freedom of expression concerns. Freedom of expression, therefore, should not be violated in the fight against disinformation, on the contrary, it should be protected and supported so that disinformation can be combated more effectively. Since I think that this outcome, which was reached in response to my research question, will create considerable awareness in the fight against disinformation, I found it worth answering. The focus of this dissertation is the impact on the freedom of expression of social media platforms' responses while combating disinformation.

The first part will be explained how disinformation emerges in social media and what its effects are, and then how social media platforms, one of the intermediaries, fight against disinformation and what their responses are. Following this introduction to the topic, freedom of expression's relationship to disinformation, an international framework for freedom of expression and its limitations, and the impacts of social media platforms' responses on freedom of expression will be detailed. Finally, answers will be sought to the questions of how freedom of expression is protected and how it is violated in the fight against disinformation, and alternative solutions for combating disinformation while respecting freedom of expression will be offered.

Principally, I will carry out theoretical desk research from the libraries and databases of academic institutions, States' statutes, rules and regulations, international treaties, relevant cases, civil society organizations' websites, and social media platforms' community standards.

2- Definitions of disinformation and social media

Octavian launched a nasty disinformation campaign more than 2,000 years ago to eliminate his competitor Mark Anthony and ultimately make him the first Roman emperor Augustus Caesar. [1] Since those early days, people have manufactured and manipulated information to wage wars, further political agendas, settle scores, harm the weak, and amass wealth. [2] Operation Bodyguard, a World War II misinformation campaign designed to hide the intended site of the D-Day invasion, serves as a classic illustration. [3] In a successful attempt to trick the Germans into believing a sizable force was preparing to attack Calais rather than Normandy, the Allies, among other lies, sent out phoney radio communications and fabricated false military bulletins. [4]

Disinformation has existed for a while. It is now possible for false or distorted information to be manufactured, spread, and magnified by different players for ideological, religious, or financial purposes at a size, velocity, and range that has never been seen before thanks to digital media. [5]

Fallis defines disinformation as false information that is intended to deceive someone. A function is "the action for which a person or item is particularly adapted or engaged," to use the dictionary definition.[6] For instance, a chair's purpose is to be sat on, much as a heart's job is to circulate blood. This analysis claims that disinformation may be distinguished by the fact that its purpose is to deceive people. Disinformation, as

[1] Irene Khan, 'Disinformation and Freedom of Opinion and Expression Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (UN Human Rights Council 2021) A/HRC/47/25.
[2] ibid.
[3] Michael Farquhar, *A Treasury of Deception: Liars, Misleaders, Hoodwinkers, and the Extraordinary True Stories of History's Greatest Hoaxes, Fakes and Frauds* (1st Printing Edition, Penguin Books 2005).
[4] ibid.
[5] Jeffrey T Hancock, 'Digital Deception: Why, When and How People Lie Online'.
[6] HarperCollins Publishers, 'The American Heritage Dictionary Entry': <https://www.ahdictionary.com/> accessed 31 January 2023.

opposed to an innocent error, occurs from an individual who is purposefully trying to deceive. [7] [8]

There is not any globally agreed-upon definition of disinformation. As defined in the Final Report of the EU High-Level Expert Group on Fake News and Online Disinformation, disinformation encompasses all types of false, incorrect, or misleading information that is actively prepared, published, and promoted to damage the public or profit.[9] It excludes issues arising from the creation and distribution of illegal content online (for example, defamation, hate speech, and incitement to violence), which are subject to regulatory remedies under EU or national laws, as well as other types of deliberate but non-misleading factual distortions such as satire and parody. [10] Satire and parody are forms of expression that use exaggeration, irony, or humour to comment on society or politics. While they may contain elements of truth, they are not intended to be taken literally.

The Broadband Commission for Sustainable Development, on the other hand, has approached disinformation as false or misleading content with potential consequences, irrespective of the underlying intention or behaviours of producing and circulating messages. [11] According to this definition, irrespective of the underlying intent, it is sufficient to have a potential consequence of disinformation to define it as disinformation. However, in the EU definition, the intent is important, it should be deliberately aimed to deceive the public.

Due to the lack of a universally agreed definition of disinformation within the context of international legal norms, the idea of disinformation is susceptible to being confused with several other concepts. Some scholars have created a classification of an information disorder in which "disinformation" is defined as false information shared with the intent to harm, "misinformation" as the unintended spreading of false

[7] Paul S Piper, *Web Hoaxes, Counterfeit Sites, and Other Spurious Information on the Internet* (2002).
[8] James Fetzer, 'Disinformation: The Use of False Information' (2004) 14 Minds and Machines 231.
[9] European Commission and Content and Technology Directorate-General for Communication Networks, *A Multi-Dimensional Approach to Disinformation: Report of the Independent High Level Group on Fake News and Online Disinformation.* (2018) <https://data.europa.eu/doi/10.2759/739290> accessed 30 November 2021.
[10] ibid.
[11] Broadband Commission, 'Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression' (UNESCO 2020).

information, and "malinformation" as truthful information shared with the intent to harm. [12] Misinformation is often the result of a mistake or lack of knowledge, rather than a deliberate attempt to deceive. Malinformation can be used to damage an individual's reputation, exploit vulnerabilities, or incite violence. Furtherly, disinformation should not be restricted to the falsehood with possible harm merely in news content (as the term "fake news" implies), nor can it be confused with propaganda or hate speech. [13]

Disinformation and fake news are often used interchangeably, but there are some key differences between the two. Disinformation is spread deliberately to deceive people, while fake news is simply news that is false or fabricated. The intent behind fake news can be varied, from financial gain to sensationalism to political bias. Fake news can be completely fabricated or can include a mix of true and false information. Disinformation can come from a variety of sources, including state actors, non-state actors, and individuals, while fake news is often associated with clickbait websites or social media accounts that are created for the purpose of spreading false or misleading information for financial gain or other purposes.

The term 'propaganda' refers to the deliberate endeavour to shape the minds of individuals in order to achieve a desired result. [14] Propaganda is spreading with the aim of influencing people's ideas or behaviour to support a particular agenda or ideology. Disinformation can come from a variety of sources as mentioned above, while propaganda is often associated with governments, political parties, or interest groups. Disinformation is therefore a subtype of propaganda: whereas propaganda contains both genuine and untrue persuading content, disinformation consists only of false or altered information that is transmitted with the intent to cause harm. [15] Hereby, although disinformation overlaps with the notion of propaganda, it is distinctive.

---

[12] Claire Wardle and Hossein Derakhshan, 'Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making' <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html> accessed 31 January 2023.
[13] (n 11).
[14] John B. Whitton, 'Propaganda and International Law (Volume 72)', *Collected Courses of the Hague Academy of International Law* (Brill 1948) <https://referenceworks.brillonline.com/entries/the-hague-academy-collected-courses/propaganda-and-international-law-volume-72-A9789028610927_06> accessed 31 January 2023.
[15] Kate Jones, 'Online Disinformation and Political DiscourseApplying a Human Rights Framework'.

In contrast, the term 'hate speech' refers to words or other forms of expression that are designed to humiliate, degrade, or instigate violence or prejudice against a specific group of individuals based on their race, nationality, ethnic origin, faith, gender identity, or other traits. Even though both disinformation and hate speech can be detrimental and have harmful outcomes, they are distinct ideas with different juristic and social repercussions. [16]

It's important to indicate that these definitions are not fixed and the distinction between disinformation and other types of false information can be challenging in some cases, especially in the online environment where it's easier to manipulate and spread false information and can be harder to trace the origin of the information. Also, the impact of false information types can be similar, regardless of whether it is classified as disinformation or others. As a result, this study will use the definition of disinformation from the EU High-Level Expert Group's Final Report on Fake News and Online Disinformation.

After stating what should be understood by the concept of disinformation, it would be appropriate to provide context about the current status of disinformation and its prevalence in various media forms. Disinformation as defined can be spread through various channels, including social media, news outlets, and messaging apps. Disinformation can take many forms, including fabricated news stories, doctored images and videos, and false narratives that are designed to appeal to the emotions of the audience. One of the main impacts of disinformation is the spread of false information, which can lead to confusion and misunderstanding among the public. For example, false information about a political candidate can influence voters' decisions, while false information about a product or service can harm consumers.

Disinformation can also erode trust in institutions and information sources, making it more difficult for individuals to separate fact from fiction. This can lead to a "post-truth" society, in which emotions and personal beliefs are more influential than facts and evidence in shaping public opinion and decision-making. Another impact of disinformation is that it can be used as a tool in political and military operations to gain advantages over an opponent. It can be used to influence elections, sow discord among

[16] (n 11).

rival groups, and even incite violence. It can also be used by non-state actors such as criminal and terrorist organizations to achieve their goals. Overall, disinformation is a serious threat to society as it can undermine democracy and stability by spreading false information and eroding trust in institutions and information sources.

3- Disinformation Reponses by Social Media Platforms

The subject of this study is social media companies' responses to disinformation, and they can be sorted as follows. Social media companies use a variety of responses to combat disinformation, including:

a- Content moderation:

Content moderation is the process of reviewing and removing content that violates a company's terms of service or community guidelines. Social media companies use a combination of automated and human moderation to review content.

They use software, such as machine learning algorithms, to automatically flag content that appears to violate the company's terms of service or community guidelines. This can include content that contains false, incorrect, or misleading information. The software can be trained to recognize certain keywords, phrases, or images that are associated with violating content. The software can also be trained to recognize patterns in images, videos and audio to detect violence, pornography, and other sensitive material. However, relying solely on automated moderation can lead to false positives or negatives, meaning that the system may flag content that is not actually violative or fail to flag content that is.

They also employ human moderators who review flagged content to determine if it should be removed. These moderators may also review content that has been reported by users. Human moderation allows for more nuanced judgment and can be more effective in understanding context and intent. Human moderators can also provide a more nuanced judgment when it comes to determining whether a post is violative or not, and can also make decisions on whether to remove, hide or flag a post. However, human moderation can be costly, and it can be difficult to find enough qualified personnel to review all flagged content.

As the amount of content being uploaded to social media platforms is vast, companies use different techniques to scale their content moderation process:

- Pre-moderation: All content is reviewed before it is posted. This is a more effective way of moderating content as it ensures that violative content does not reach the public. However, pre-moderation can be very time-consuming and can slow down the process of content being posted.
- Post-moderation: Content is reviewed after it has been posted. This method is less effective as it allows violative content to reach the public before it is removed. It also requires a lot of resources to review all the content that has been posted.
- Community moderation: Users are able to report content they believe is false, incorrect or misleading. This allows users to flag content they find violative and allows the social media company to focus on the most reported content. Once a piece of content is reported, a team of human moderators will review it to determine if it should be removed or labelled. Some social media companies use crowdsourcing to moderate content, by allowing users to vote on the accuracy of certain pieces of content. Pieces of content that receive a large number of negative votes may be flagged for removal or labelled as potentially false. Companies often establish community guidelines to set expectations for behaviour and content on their platforms. These guidelines may prohibit the sharing of false, incorrect, or misleading information, and users who violate these guidelines may be subject to enforcement actions such as account suspension or removal of content. These are some examples of how community moderation works, but the specifics of how a particular social media company implements community moderation can vary.

Content moderation is a powerful response for combating disinformation because it helps to identify and remove false or misleading information from social media platforms. But at the same time, it is a challenging task that requires balancing multiple considerations, including protecting users from harm, promoting free expression, and ensuring compliance with laws and regulations, and it's not perfect. Social media companies have faced criticism from users, advocacy groups, and lawmakers for not

doing enough to remove harmful content, and for removing content that shouldn't have been removed. Companies should continuously work to improve their content moderation process to strike the right balance between protecting users and allowing free expression.

b- Fact-checking:

Fact-checking is a response that social media companies use to combat disinformation by identifying and flagging false, incorrect, or misleading information on their platforms. They can partner with fact-checking organizations to help identify and flag disinformation on their platforms. Fact-checking organizations can provide expertise in identifying false, incorrect or misleading information and can work with social media companies to develop systems for identifying and flagging disinformation.

Social media companies can use machine learning algorithms to automatically identify and flag disinformation on their platforms. These algorithms can be trained to recognize patterns and characteristics of disinformation and can flag content for review by human moderators.

They can also review flagged content and determine if it is false, incorrect, or misleading. If it is, they can take steps to reduce its spread or remove it from their platforms. In addition, this content can be labelled to warn users that the information has been fact-checked and found to be disinformation.

Although fact-checking has a significant function for combating disinformation because it helps to identify and flag false, incorrect, or misleading information on social media platforms, it seems more meaningful to approach it as a stimulant that will strengthen the fight process rather than a saviour that can completely solve the problem. This can help to reduce the spread of disinformation and to protect users from being misled by disinformation.

c- Algorithm adjustments:

Algorithm adjustments is one of the responses that social media companies use to combat disinformation. Social media companies may adjust their algorithms to reduce the visibility of content that has been flagged as disinformation. For example, they may use machine learning algorithms to detect patterns of behaviour that are associated with

disinformation campaigns, such as the use of bots to amplify a message. They can also use algorithms to identify accounts that are spreading disinformation and limit the reach of their content by reducing its visibility in news feeds search results and recommendations.

They can also use algorithms to promote credible information and to surface news articles from reputable sources. For instance, they may use algorithms to identify news articles from reputable sources, such as major news organizations, and surface them more prominently in news feeds search results and recommendations. This can help users to find accurate and reliable information and reduce the spread of disinformation.

Algorithms can also be used to check the factuality of information shared on their platform. As mentioned above, it can be partnered with fact-checking organizations to flag information that has been found to be false or misleading. They can also use machine learning algorithms to identify patterns that are associated with false or misleading information, such as text, images, or videos. Once information is flagged as being disinformation, the company can take actions such as reducing its visibility, providing a warning, or blocking it.

They can use their algorithms to demote content that is low-quality, misleading, or spammy. For example, they may use signals such as click-through rates, time spent on a page, and engagement metrics to identify low-quality content. Once identified, the algorithm can demote this content by reducing its visibility in news feeds, search results, and recommendations, making it less likely to be seen by users.

It should not be forgotten that while these algorithm adjustments can be effective in reducing the spread of disinformation, they are not a silver bullet solution. Social media companies need to continue to refine and adapt their algorithms to keep up with the constantly changing disinformation landscape.

d- Transparency and labelling:

On the one hand, social media companies may label or tag content that is determined to be disinformation in order to provide users with additional context about the content they are viewing. For example, they may use fact-checking organizations to flag information that has been found to be disinformation. They can also use machine

learning algorithms to identify patterns that are associated with disinformation, such as text, images, or videos. Once information is flagged as being false, incorrect or misleading, the company can provide a warning or a label to the content. Besides some social media companies may label the state-controlled media so users can understand the source of information and evaluate it accordingly.

On the other hand, they can disclose information about the sponsors of political advertisements and the target audience of these advertisements. This allows users to see who is behind an ad and to understand its intended audience. This can help users to assess the credibility of an ad and to determine whether it is trying to manipulate them.

They can also disclose information about the reach and engagement of advertisements. This allows users to see how many people saw an ad and how many people interacted with it. This can help users to understand the impact of an ad and to determine whether it is having a significant effect on public opinion.

Lastly, they may disclose information about the content they remove from their platforms. This allows users to see what types of content are being removed and why. This can help users to understand the policies of the platform and to determine whether the content is being removed in a fair and unbiased manner.

e- Banning or limiting the reach of accounts:

Social media companies can ban accounts that are found to be spreading disinformation. This means that the account will no longer be able to post content on the platform. This can be a powerful tool to limit the spread of disinformation, but it is also a drastic measure that should be used judiciously.

They can also temporarily suspend accounts that are found to be spreading disinformation. This means that the account will not be able to post content on the platform for a specific period of time. This can be a less drastic measure than banning an account and can be used to give the account holder a chance to change their behaviour.

It may also be limited the reach of accounts that are found to be spreading disinformation. This means that the account's content will be seen by fewer people. This can be accomplished by reducing the visibility of the account in search results, reducing

the visibility of the account's content in the news feed, or reducing the visibility of the account's content in the platform's advertising.

Furthermore, it can be identified the network of accounts that are spreading disinformation and limit the reach of those accounts. This means that the accounts that are found to be spreading disinformation will be seen by fewer people and will be less effective at spreading disinformation. In addition to this, they can limit the functionality of accounts that are spreading disinformation, such as disabling comments, disabling the ability to create groups or pages, or disabling the ability to send direct messages.

In spite of the fact that banning or limiting the reach of accounts can be effective in reducing the spread of disinformation, it does not prevent the creator of the account from creating a new one, so companies may need to take a multifaceted approach.

f- Educating users:

Educating users is one of responses that social media companies use to combat disinformation. It can be addressed more than one different form of education on the subject.

Firstly, social media companies can provide education to users on how to fact-check information they come across on the platform. This can include providing tips on how to identify credible sources, how to check the veracity of images and videos, and how to spot the signs of disinformation.

Secondly, they may provide education to users on how to critically evaluate the information they see on the platform. This can include teaching users how to identify the source of information, how to evaluate the credibility of the source, and how to spot the signs of disinformation.

Thirdly, users can be also educated on how to spot disinformation on the platform. This can include teaching users how to identify the signs of disinformation, such as sensational headlines, and how to evaluate the credibility of the information they see. Moreover, public campaigns and outreach programs can be launched to educate users on disinformation and how to spot it. This can include creating public awareness campaigns, hosting webinars, and creating educational materials.

Finally, social media companies can also integrate educational features within their platforms, such as providing explanations on how certain features work, or providing a feature that allows users to report disinformation or check the credibility of a piece of content.

It is crucial to state that educating users is an ongoing process and it is not a one-time solution. Social media companies need to continue to provide education to users to keep them informed of the latest disinformation tactics, and to empower users to be able to identify and report disinformation. Additionally, these education efforts can be complemented with other responses like those previously mentioned.

g- Collaboration with other organizations:

Collaboration with other organizations is the last response to be cited in this paper that social media companies use to tackle disinformation. The following has been mentioned several times already, social media companies can partner with fact-checking organizations to help identify and flag disinformation on their platforms. Additionally, it can be appropriated for collaborating with civil society organizations to empower citizens to identify and report disinformation. This can include providing tools and resources that citizens can use to identify and report disinformation and working with civil society organizations to create public awareness campaigns.

They can also collaborate with academic researchers to better understand the spread of disinformation on their platforms and to develop new strategies for tackling it. Researchers can provide insight into the psychology and sociology of disinformation, as well as the technical aspects of its spread.

It may be worked with government agencies to help combat disinformation. For example, they can share information with government agencies about disinformation campaigns and can work together to develop strategies for combating them.

Moreover, they can join industry coalitions to share information and best practices with other companies in the industry. This can include sharing data on disinformation campaigns and developing shared standards for identifying and flagging disinformation. This response can help social media companies to combat disinformation by providing

them with additional resources, expertise, and information, which can improve their ability to identify and flag disinformation on their platforms.

In addition to all these, cooperation can be made with external grievance mechanisms. For instance, The Oversight Board is an independent body created by Facebook to make decisions on complex and difficult content moderation cases. The board is made up of experts in a variety of fields, including human rights, free speech, and technology, and its goal is to provide transparency and accountability in how Facebook handles content on its platform. The Board makes final and binding decisions on whether specific content should be allowed or removed from the platform since late 2021. There are two types of content decisions about which users can appeal to the supervisory board. One is the decisions made by Facebook regarding their own content, and the other is the decisions made by Facebook regarding the content they report belonging to others. In either case, an appeal must first go through Facebook's review request process. Unfortunately, not every content and content decision can be appealed at the moment, but more options are being added every day. [17]

Overall, while it is important for social media companies to take steps to combat disinformation, they must also ensure that their efforts do not unduly restrict freedom of expression or other rights. Before moving on to the concerns about how these responses may undermine freedom of expression, it would be appropriate to provide detailed information about freedom of expression.

4- Definition and International Framework of Freedom of Expression

Freedom of expression is protected as a human right under international law. It is enshrined in several international human rights treaties, including the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights (ICCPR).

Article 19 of the International Covenant on Civil and Political Rights (ICCPR) as a human rights instrument guarantees that individuals have the right to express their opinions and ideas freely and without interference from the government or other actors.

[17] 'Oversight Board | Independent Judgement. Transparency. Legitimacy.' <https://oversightboard.com/> accessed 31 January 2023.

[18] It also guarantees the right to access and receive information and ideas from a variety of sources, including traditional and new media. [19] This includes the right to access information that may be considered controversial or unpopular. [20] It refers to the right to express opinions and ideas without censorship, restraint, or fear of retribution. [21] The ICCPR is an international treaty adopted by the United Nations General Assembly in 1966, and it has been ratified by 179 countries.

From the perspective of disinformation, it is an issue that should be highlighted that content is within the scope of this right regardless of whether it is true or false. All information and ideas are protected within the framework of the right to freedom of expression. [22] According to article 10 of the ECHR, even if there is a strong suspicion that the information is false, the discussion or dissemination of this information is not prohibited. [23] In this context, people can make unfounded ideas and statements, including the right to parody and satire. Statements considered to be disinformation may only be restricted under the criteria set out in Article 19 (3) of the International Covenant on Civil and Political Rights (ICCPR).

Article 19 (3) of the International Covenant on Civil and Political Rights (ICCPR) is a provision that allows for certain restrictions on the rights guaranteed in Article 19 (1) and (2), which guarantees freedom of expression and the right to seek, receive, and impart information and ideas. The provision states that while individuals have the right to express their opinions and ideas freely and access information, there may be certain limitations imposed by law if they are necessary to protect the rights of others (e.g. protection of privacy or reputation), national security, public order (i.e. preventing incitement to violence or other forms of harm), or public health and morals. These limitations must be strictly proportionate to the specific need they are addressing and can't be used as a pretext to silence dissenting voices or censor unwanted information. [24]

[18] International Covenant on Civil and Political Rights.
[19] ibid.
[20] ibid.
[21] ibid.
[22] OHCHR, 'General Comment No.34 on Article 19: Freedoms of Opinion and Expression' (2011) <https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no34-article-19-freedoms-opinion-and> accessed 22 August 2022.
[23] *Salov v Ukraine* [2005] ECtHR 65518/01.
[24] David Kaye, UN Human Rights Council Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression and UN Human Rights Council Secretariat, 'Report of the

The restriction must be the least restrictive possible and the burden of proof is on the government to demonstrate that the restriction is necessary and proportionate.

The UN Human Rights Committee, which oversees the implementation of the ICCPR, has emphasized that any restriction on freedom of expression must be based on a "pressing social need" and should be the least restrictive possible. [25] It also said that there is a need to balance the interest in protecting the rights or reputations of others against the interest of protecting the freedom of expression. In summary, Article 19(3) allows for certain limitations on freedom of expression and access to information, but only if they are strictly necessary and proportional, imposed by law and not used as a pretext for censorship or silencing dissenting voices. [26] As a result, freedom of expression may be restricted if all three of these three-part tests are carried out. These three conditions can be entitled as legality, necessity, and legitimacy respectively. [27] Other provisions of the ICCPR must also be complied with, including any restriction of speech and non-discrimination.

There are other international treaties and articles that address the right to freedom of expression and access to information. For example, the Universal Declaration of Human Rights (UDHR) also includes a right to freedom of expression in Article 19, which states: "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers."[28]

The European Convention on Human Rights (ECHR) also includes a right to freedom of expression in Article 10, which states: "Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers."[29]

Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression ': <https://digitallibrary.un.org/record/1631686> accessed 22 August 2022.

[25] (n 22).

[26] International Covenant on Civil and Political Rights.

[27] Evelyn Mary Aswad, 'The Future of Freedom of Expression Online' 17 TECHNOLOGY REVIEW.

[28] Universal Declaration of Human Rights.

[29] European Convention on Human Rights.

Additionally, the American Convention on Human Rights (ACHR) also includes a right to freedom of expression in Article 13, which states: "Everyone has the right to freedom of thought and expression. This right includes freedom to seek, receive, and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing, in print, in the form of art, or through any other medium of one's choice."[30]

It is worth mentioning that these international treaties and articles all have similar provisions that allow for certain limitations on the rights they guarantee, but only if they are strictly necessary and proportionate, imposed by law and not used as a pretext.

Furthermore, it can be said that there are also many other international and regional instruments that address the right to freedom of expression and access to information, such as:

- The African Charter on Human and Peoples' Rights (ACHPR), which includes a right to freedom of expression in Article 9.[31]
- The ASEAN Human Rights Declaration (AHRD), which includes a right to freedom of expression in Article 14.[32]
- The Arab Charter on Human Rights, which includes a right to freedom of expression in Article 32.[33]
- The OAS Charter of the Organization of American States (OAS Charter), which includes a right to freedom of expression in Article 13.[34]
- The SAARC Human Rights Charter, which includes a right to freedom of expression in Article 19. [35]

In addition to these international and regional instruments, many countries also have national laws and constitutions that protect the right to freedom of expression and access to information. Consequently, the right to freedom of expression and access to information is considered a fundamental human right and is protected by various

[30] The American Convention on Human Rights.
[31] The African Charter on Human and Peoples' Rights.
[32] The ASEAN Human Rights Declaration.
[33] The Arab Charter on Human Rights.
[34] The OAS Charter of the Organization of American States.
[35] The SAARC Human Rights Charter.

international and regional instruments, and is an important aspect of democratic societies, as it allows for the free flow of ideas and information and ensure that citizens are able to hold their governments accountable.

5- The Impact of Responses of Social Media Companies on Freedom of Expression

Social media companies have been under increasing pressure to combat the spread of disinformation on their platforms. To address this issue, they have implemented a variety of responses as said above. Although these responses are typically beneficial, they fall short of addressing the issues created by disinformation. [36] Moreover, the methods used by social media companies to combat disinformation can potentially violate freedom of expression.

The first potential issue is that content moderation, a method used by social media companies to combat disinformation, has the potential to violate freedom of expression in several ways. Content moderation can violate freedom of expression in several ways. First, it can result in the removal or censorship of legitimate speech, including speech that is critical of governments, corporations, or other powerful actors.[37] Second, it can result in unequal treatment of different types of speech, with some types of speech being allowed while others are censored. Third, it can result in over-censorship, with social media companies becoming overly cautious in their efforts to prevent the spread of false information, thereby stifling freedom of expression. For example, the Human Rights Watch has stated that "content moderation policies should be designed to minimize restrictions on freedom of expression, to be transparent and accountable, and to provide for effective redress mechanisms for those whose content is restricted or removed." Social media companies must ensure that their content moderation policies are transparent, accountable, and minimally restrictive to freedom of expression.

The seeming discrepancy in how they implement their policies in different parts of the world is a serious issue currently. [38] While the United States has profited from

[36] Khan (n 1).
[37] Kaye, Expression and Secretariat (n 24).
[38] 'Intervozes Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation'

election or voting information centres and advertisement libraries, the majority of nations do not appear to receive the same degree of investment. [39] Even in the United States, Spanish sources are subsidised less than their English-language counterparts. There appears to be a minority of corporations that are adopting relatively short-term territorial election disinformation rules in locations where they operate. [40] In addition, the option to label content appears to be accessible in its whole just in the United States, and to a far lesser extent in Latin American nations. [41] The world's giant U.S.-based companies are apparently driven by American and European interests. [42] They do not devote enough efforts to comprehending the localized circumstances that fuel online falsehoods in other regions, particularly third world countries. [43] In regions where disinformation becomes pervasive, a comprehensive awareness of the regional democratic, sociological, and financial environment, linguistic skills, and strong engagement with local communities are required.[44]

The second potential issue is that fact-checking can be seen as censorship. Fact-checking may limit the spread of false information, but it can also limit the spread of opinions, opinions which may not align with the opinions of the fact-checkers or the social media company. This censorship can restrict the diversity of ideas and perspectives available, leading to the restriction of freedom of expression. [45] Furthermore, when social media companies engage in fact-checking, they are

<https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/Intervozes-Coletivo-Brasil.pdf>.

[39] 'Privacy International Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation'
<https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/Privacy-International.pdf>.

[40] 'Facebook Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation' (Facebook)
<https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/4-Companies/Facebook.pdf>.

[41] 'Derechos Digitales Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation'
<https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/Derechos-Digitales.pdf>.

[42] Khan (n 1).

[43] 'Center for Law and Democracy Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation'
<https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/UN-SR-on-FOE-CLD-Submission-Disinformation-Mar21-final.pdf>.

[44] Khan (n 1).

[45] Kaye, Expression and Secretariat (n 24).

essentially making a determination of what is true and what is false. This can lead to the suppression of opinions and ideas that fact-checkers or social media companies do not agree with, which can restrict the diversity of viewpoints available to the public. Additionally, the fact-checking process itself can be subjective and influenced by the personal biases and perspectives of the fact-checkers, leading to further concerns about the impartiality and accuracy of the process. This can also call into question the legitimacy of the fact-checking process and the decisions made.

Another significant problem is that there are fewer fact-checking organizations in many regions around the globe. [46] Concerns were raised regarding the intentional inactivity of organizations in less wealthy locations, as revealed by whistle-blowers. [47] If substantiated, the inequalities would indicate vastly divergent content filtering standards, damaging the public sphere in underdeveloped countries. [48] Overall, the potential for fact-checking to restrict freedom of expression highlights the need for caution and balance when implementing these measures.

The third potential issue is that algorithm adjustments can involve the manipulation of the algorithms that govern the visibility and spread of information on social media platforms. For example, social media companies might adjust their algorithms to suppress certain types of content, prioritize certain types of content over others, or otherwise manipulate the information landscape in an effort to combat disinformation. However, these adjustments can have unintended consequences, including the suppression of legitimate speech and the unequal treatment of different types of speech. For example, if algorithms are adjusted to prioritize news from mainstream media sources over alternative media sources, this could result in the suppression of diverse perspectives and stifle freedom of expression. Additionally, if algorithms are adjusted to suppress certain types of content, this could result in the suppression of dissenting voices, violating the right to free expression.

---

[46] Mahsa Alimardani and Mona Elswah, 'Trust, Religion, and Politics: Coronavirus Misinformation in Iran' <https://papers.ssrn.com/abstract=3634677> accessed 31 January 2023.
[47] Khan (n 1).
[48] Craig Silverman Dixit Ryan Mac, Pranav, '"I Have Blood On My Hands": A Whistleblower Says Facebook Ignored Global Political Manipulation' *BuzzFeed News* (14 September 2020) <https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo> accessed 31 January 2023.

Furthermore, the process of adjusting algorithms can be opaque, with users and content creators often having little or no understanding of how the algorithms are affecting the visibility and spread of their content. This lack of transparency can further undermine freedom of expression, as users and content creators are unable to make informed decisions about what they post and share. Additionally, if the algorithms are not transparent or subject to public oversight, it can be difficult to know why certain content or perspectives are being promoted or suppressed, leading to a lack of accountability and trust in the system. This issue is compounded when the automatic filters and algorithms on which social media companies rely extensively fail to recognise nuance and comprehend the context. The problem is becoming more noticeable because of the technological limitations of automatic filters and algorithms and the politicisation of disinformation. This, combined with the lack of transparency in content moderation, demonstrates that the risk of removing content authorised by international law is rather considerable. [49]

The fourth potential issue is that transparency and labelling can result in censorship or suppression of speech if the criteria for labelling or flagging content as false or misleading are too broad or subjective. For example, if a social media company decides to label or flag content based on the political views of the speaker or the message being conveyed, it could result in the censorship of political speech or the suppression of certain perspectives. Similarly, if the criteria for labelling or flagging content are vague or ill-defined, it could result in the arbitrary or inconsistent application of these labels, which could also have a chilling effect on freedom of expression. [50]

Most social media platforms publish transparency reports two times per year, but they don't provide more detailed and insightful information about the steps they've made to combat false or misleading material. [51] [52] For instance, the Facebook Transparency

[49] Khan (n 1).
[50] 'Chilling Effects: The Impact of Social Media Platforms' Policies and Practices on Freedom of Expression'.
[51] 'Association for Progressive Communications Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation'.
[52] 'Working Groups - Forum on Information & Democracy' (*Forum Information & Democracy*) <https://informationdemocracy.org/working-groups/concrete-solutions-against-the-infodemic/> accessed 31 January 2023.

Report solely details the deletion of fraudulent accounts but not the content. [53] Like this, there is no information given regarding the quantity of content getting flagged, nor are there any appeals concerning decisions to delete information or restrict user profiles in accordance with disinformation policy. There is no information given about how users interact with false or misleading information, such as the number of shares, visits, reaches, complains, or deletion orders. [54] It is challenging to evaluate the efficacy of the initiatives taken by the platforms and the impact on human rights due to the general absence of transparency surrounding the methods and procedures used by companies for content filtering. [55] Increased transparency is indeed required when it comes to agreements between businesses and governments, particularly when those agreements call for the deletion of content or other limitations or the promotion of official statements. [56]

Additionally, there are concerns about the potential for these labels or flags to become a form of "shadow censorship," where the mere presence of a label or flag on a piece of content could discourage others from sharing or engaging with it, even if the content itself is not false or misleading. [57] This could have a significant impact on the visibility and reach of certain perspectives or ideas, leading to a suppression of free expression. In conclusion, clear and specific criteria and clear processes for appeal and review of labelling decisions should be given for labelling or flagging content, and it should be regularly reviewed.

The fifth potential issue is that banning or limiting the reach of accounts can potentially violate freedom of expression because it restricts the dissemination of information and limits the ability of individuals or groups to share their views and ideas. This restriction could have a chilling effect on free speech, as people may be discouraged from speaking out on certain topics or issues for fear of facing similar

[53] Khan (n 1).
[54] 'March 2021 Coordinated Inauthentic Behavior Report' (*Meta*, 6 April 2021) <https://about.fb.com/news/2021/04/march-2021-coordinated-inauthentic-behavior-report/> accessed 31 January 2023.
[55] 'Association for Progressive Communications Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation' (n 51).
[56] Khan (n 1).
[57] Kaye, Expression and Secretariat (n 24).

consequences. [58] Moreover, account bans or reach limitations may not always be based on clear and transparent criteria, leading to the possibility of arbitrary or discriminatory enforcement. In some cases, these practices may also result in the suppression of dissenting voices, thereby undermining the diversity of perspectives and opinions in public discourse. Another concern is that social media companies don't always make it clear what kind of harm, and what potential harm, could lead to content or account removal. [59]

In addition, the coordinated inauthentic behaviour policy implemented by Facebook may also have a restrictive effect on freedom of expression. This policy is at the heart of the operation of the fake account. The purpose of this policy is to prevent the use of multiple Facebook or Instagram pages, accounts or groups working in concert to engage in behaviour designed to artificially increase the popularity of the content. Facebook has removed thousands of accounts as part of this policy, but this raises concerns about violating freedom of expression as it does not transparently publish the removed accounts and the reasons for removal, and whether it ignores the possibility that these accounts may also act as part of legitimate campaign activity. [60]

The sixth and last potential issue is about collaboration with other organizations. This response can impact freedom of expression in several ways. Firstly, the partnership could result in the sharing of user data between the organizations, which could have implications for user privacy. The organizations involved in the collaboration could use the shared data for purposes that go beyond the original intention of combating disinformation, which could restrict the freedom of expression of users. Secondly, the choice of organizations that a social media company partners with can also have a significant impact on freedom of expression. For example, if a social media company partners with organizations that have a vested interest in suppressing certain voices or narratives, this could result in censorship and restriction of information that is critical of those organizations or their interests. In this case, the partnership could be seen as an infringement on freedom of expression, as it gives undue influence on certain actors

[58] 'Chilling Effects: The Impact of Social Media Platforms' Policies and Practices on Freedom of Expression' (n 50).
[59] Khan (n 1).
[60] 'Inauthentic Behaviour | Transparency Centre' <https://transparency.fb.com/en-gb/policies/community-standards/inauthentic-behavior/> accessed 31 January 2023.

over the content that is available on the platform. Finally, the criteria for determining what constitutes "disinformation" can also impact freedom of expression. If these criteria are not clearly defined and made public, this can lead to arbitrary censorship and restriction of free expression, as different organizations may have different opinions on what constitutes disinformation. [61]

In addition to the last issue, when social media companies work with governments, they can be compelled to restrict access to information or silence opposing voices. This could be seen as a violation of the right to freedom of expression, which includes the right to receive and impart information and ideas without interference from public authorities. In some countries, the government may use this collaboration as a way to suppress political opposition, restrict access to information, or silence critical voices. This can have a chilling effect on free speech and undermine the ability of individuals to engage in meaningful political discourse. [62] Additionally, if social media companies comply with government requests to restrict content or silence accounts, they may be seen as complicit in violating freedom of expression. The extent to which collaboration with the government can be seen as a violation of freedom of expression will depend on the specific circumstances and the laws and regulations in place in each country.

During the 2020 US Presidential elections, social media platforms came under fire for the varied execution of community rules. [63] Following the coup attempt in Myanmar in February 2021, Facebook blocked military-affiliated profiles, but it has not said whether it will do so in future circumstances. [64] Platforms appear to have bowed to demand from the authorities to ban the profiles of reporters and human rights advocates who covered the agricultural' demonstrations in India for a year. [65] In 2020, it was revealed that Facebook had consented to greatly expand adherence with the Authorities

[61] Kaye, Expression and Secretariat (n 24).
[62] 'Chilling Effects: The Impact of Social Media Platforms' Policies and Practices on Freedom of Expression' (n 50).
[63] Khan (n 1).
[64] 'Facebook Restricts Myanmar Military's Accounts for Spreading "misinformation" | CNN Business' <https://edition.cnn.com/2021/02/12/tech/facebook-myanmar-military-intl-hnk/index.html> accessed 31 January 2023.
[65] 'India: Journalists Covering Farmer Protests Charged' (*Human Rights Watch*, 2 February 2021) <https://www.hrw.org/news/2021/02/02/india-journalists-covering-farmer-protests-charged> accessed 31 January 2023.

of Viet Nam's request to redact "anti-State" information. [66] In addition, many human rights protestors' profiles had been shuttered after the government had shut down the platform's web servers, which caused the website to be sluggish and inaccessible for seven weeks. [67]

Overall, while it is important for social media companies to combat disinformation, it is also important to ensure that their methods do not violate freedom of expression. This may require additional oversight and regulation to ensure that the rights of users are protected. Finally, the perspectives on the balance between preventing disinformation on the platforms of social media companies and protecting freedom of expression are as follows:

1- The first perspective on the issue of how social media companies combat disinformation and its impact on freedom of expression is that social media companies have a responsibility to combat disinformation because it can have serious real-world consequences. It is argued that disinformation can undermine democratic processes, spread misinformation and mistrust among people, and can cause harm to individuals and society at large. From this perspective, it is seen that disinformation on social media platforms is a serious problem that needs to be addressed. Social media companies have a responsibility to protect their users from the harmful effects of disinformation, and they should take measures to prevent its spread on their platforms.

While freedom of expression is important, it should not be used as an excuse to allow the spread of false and harmful information. Social media companies should not allow users to post false or misleading information that can cause harm to others. Moreover, this perspective emphasizes the fact that disinformation can have serious consequences in the real world, such as influencing elections, spreading conspiracy theories, and fomenting violence. Therefore, it is important to take action to combat disinformation, even if it means curtailing some forms of expression.

Social media companies should take a proactive approach to combat disinformation, including fact-checking, removing false content, and suspending or banning accounts

[66] 'Exclusive: Facebook Agreed to Censor Posts after Vietnam Slowed Traffic - Sources | Reuters' <https://www.reuters.com/article/us-vietnam-facebook-exclusive-idUSKCN2232JX> accessed 31 January 2023.
[67] ibid.

that consistently spread disinformation. Additionally, more transparency is needed about how social media companies combat disinformation, and how algorithms used to identify disinformation are designed and tested for bias. There may be challenges and trade-offs in implementing these measures, but the benefit of protecting people from the harms of disinformation outweighs those challenges.

2- The second perspective is that social media companies are private entities and therefore have the right to curate the content on their platforms as they see fit. From this perspective, it is argued that social media companies are not bound by the same rules and regulations that govern freedom of expression in the public sphere. As private entities, they have the right to determine their own policies and guidelines for the content that appears on their platforms.

It is held that the methods used by social media companies to combat disinformation, such as fact-checking and removing false content, are not a violation of freedom of expression because users are still free to express themselves on other platforms. In other words, users are not being censored or suppressed by the government, but by a private company that has the right to decide what content appears on its platform. Additionally, it argues that social media companies have a responsibility to their shareholders and users to provide a safe and reliable environment for communication and information sharing and that combating disinformation is an important aspect of fulfilling this responsibility.

From this point of view, it is also argued that the fact that these companies are private entities means that they are not held to the same standards as the government when it comes to censorship and freedom of expression. They have the right to set their own rules, and users can choose whether to use their platforms or not. However, there may be challenges in implementing these measures, such as the potential for bias in the fact-checking process or the risk of suppressing certain voices or perspectives.

Indeed, it should not be surprising that private entities are collaborating with governments to regulate online speech through a parallel governance exercise. [68] Hedley

[68] 'Silencing the Messenger: Communication Apps Under Pressure' *Freedom House* <https://freedomhouse.org/report/freedom-net/2016/silencing-messenger-communication-apps-under-pressure> accessed 31 January 2023.

Bull asserted in 1977 that the international system may transition from one oriented on nation-states to one in that nations could participate in control over their inhabitants with a range of other influential entities, especially transnational businesses. [69] This multilateral system is referred to as "neo-medieval" since there were no nation-states in Europe during the Middle Ages, but rather a number of conflicting prominent players who conducted diverse types of control over citizens. [70] Considering that worldwide social media firms increasingly execute conventional governmental powers by, along with other things, imposing their own speech standards on their networks, it indicates that Professor Bull's vision of a neo-medieval world has become a reality. [71]

Furthermore, the United Nations stated that companies should align their policies and guidelines with international human rights law, especially the ICCPR. [72] A few months before, one of the most prominent worldwide non-governmental organisations advocating for free speech issued a similar demand for businesses to base their own community rules on the international human rights framework. [73] After this call, the Twitter CEO announced that his company's values would be based on international human rights, and Facebook also referred to international human rights when discussing its content moderation policies. [74]

3- The third perspective is that the methods used by social media companies are not enough and that there is a need for government regulation. From this perspective, it is argued that social media companies have too much power and influence over the public discourse and that the government needs to step in to ensure that freedom of expression is protected while also combating disinformation. Social media companies have not done enough to combat disinformation on their own and that government regulation is necessary to ensure that they are held accountable for the content that appears on their

[69] Hedley Bull, *The Anarchical Society* <https://link.springer.com/book/10.1007/978-1-349-24028-9> accessed 31 January 2023.
[70] ibid.
[71] 'Hard Questions: Who Reviews Objectionable Content on Facebook — And Is the Company Doing Enough to Support Them?' (*Meta*, 26 July 2018) <https://about.fb.com/news/2018/07/hard-questions-content-reviewers/> accessed 31 January 2023.
[72] Kaye, Expression and Secretariat (n 24).
[73] 'Side-Stepping Rights: Regulating Speech by Contract' (*ARTICLE 19*, 19 June 2018) <https://www.article19.org/resources/side-stepping-rights-regulating-speech-by-contract/> accessed 31 January 2023.
[74] Aswad (n 27).

platforms. This could include laws and regulations that require social media companies to remove false or misleading information, or to disclose the methods they use to combat disinformation.

Additionally, it is argued that government intervention is necessary to ensure that freedom of expression is protected and that social media companies are not suppressing certain voices or perspectives. Government oversight could help ensure that social media companies' methods for combating disinformation are transparent and fair and that they are not being used to silence dissenting opinions. This perspective is also in favour of holding social media companies liable for spreading disinformation, or for not taking enough action to stop it. It is acknowledged the complexity of the issue and the potential trade-offs that come with government regulation. However, it is argued that government intervention is necessary to ensure that social media companies are held accountable and that freedom of expression is protected.

To criticize this point of view, it can be said that many leading social media platforms have already been highly operative in their efforts to enforce the UN Guidelines Principles when they are confronted with government requests that violate international legal requirements. [75] For instance, the Global Network Initiative (GNI) is a multi-stakeholder partnership between firms (such as Google, Facebook, and Microsoft), entrepreneurs, non-state actors, and academia to give guidelines on upholding free speech in accordance with international principles. [76] Companies with a high GNI are required to comprehend the breadth of global freedom of expression rules and evaluate whether government requests to limit speech comply with Article 19 of the ICCPR and its three-part test. [77] If government legislation or mandates violate the threefold check, GNI firms are supposed to fight to execute the regime's demand prior to cooperating with domestic legislation. [78] GNI organizations can fight by, for example, filing litigation in state courts and requesting aid from foreign countries or the United

[75] Kaye, Expression and Secretariat (n 24).
[76] 'GNI Home' (*Global Network Initiative*) <https://globalnetworkinitiative.org/> accessed 31 January 2023.
[77] 'The GNI Principles' (*Global Network Initiative*) <https://globalnetworkinitiative.org/gni-principles/> accessed 31 January 2023.
[78] 'Implementation Guidelines' (*Global Network Initiative*) <https://globalnetworkinitiative.org/implementation-guidelines/> accessed 31 January 2023.

Nations human rights apparatus. [79]Constantly, the GNI's evaluation methodology has determined that the stakeholders involved have implemented their promises. [80]

In summary, it should be underlined that the corporate responsibility of social media companies to respect human rights while combating disinformation should be an obstacle to governments cooperating in enforcing laws that conflict with relevant international protections and the adoption of platform rules or business models that adversely affect the enjoyment of human rights. [81]

These perspectives are not mutually exclusive, and that different individuals or groups may hold different combinations of these views. Additionally, this is a complex and evolving issue and perspectives may change over time.

6- Alternative Solutions for Combating Disinformation While Respecting Freedom of Expression and Conclusion

Alternative solutions and recommendations on the subject can be listed as follows by summing up all the answers and perspectives.

The significance that digital media brings to democratic governance, sustainable growth, and human rights, as well as the critical role that the right to free speech plays in that balance, should not be overlooked by social media businesses. The purpose and method for battling misinformation are the right to freedom of expression; they are not a component of the issue. Disinformation is an issue, but corporate answers have been ad hoc, subpar, and ambiguous. The leading companies shouldn't be concentrating on enhancing content filtering while disregarding issues with human rights related to their marketing strategies, inadequate transparency, and users' insufficient due process rights. The main difficulty facing governments, businesses, and the media is regaining the public's confidence in the accuracy of the information system. Disinformation must be combated through multifaceted, sub-holders' methods that are firmly rooted in the whole broad spectrum of human rights and actively involve governments, businesses, international institutions, non-state actors, and social media. It is impossible to stress the

[79] 'Company Assessments – Global Network Initiative' <https://globalnetworkinitiative.org/company-assessments/> accessed 31 January 2023.
[80] *Citizens United v FEC, 558 US 310 (2010)*.
[81] Evelyn Mary Aswad, 'IN A WORLD OF "FAKE NEWS," WHAT'S A SOCIAL MEDIA PLATFORM TO DO?' UTAH LAW REVIEW.

importance of multi-stakeholder cooperation and engagement. They should abstain from doing so unless it is necessary to uphold the stringent and limited interpretation of articles 19 (3) and other provisions of the ICCPR.

Governments have a responsibility to make sure businesses uphold human rights. They shouldn't ask firms to decide whether the content is legal under domestic legislation, which should be decided by the courts or to delete or ban it if it is protected by international human rights legislation. Companies should avoid negotiating a secret deal and be clear about this kind of demand by Governments.

The national education system should include media literacy and knowledge to interest both the world and his wife. Digital inclusion must receive more focus in addition to digital literacy in order to provide valuable, independent, transparent, coherent, dependable, and safe Internet access to individuals in developing nations who are currently completely dependent on platforms of social media and text messaging services for interconnection (via zero rating).

International human rights law requires businesses to uphold human rights. Despite being independent, social media platforms have a significant effect on human rights in public life. As a result, they are answerable to society as well as its consumers. Companies should consciously address these issues by evaluating their marketing strategies, recognising the entity and independence of account holders as right holders, and encouraging them by enhancing transparency, regulation, and preference as well as by guaranteeing judicial process. This goes beyond simply improving content moderation. Social media firms should make sure that their data collecting, and processing procedures comply with applicable national consumer protection laws, data protection principles, and international human rights standards, such as Article 19 of the ICCPR. Additionally, they ought to evaluate how their goods affect human rights, especially regarding how algorithms and rank mechanisms contribute to the spread of false or misleading information. Such analyses must be carried out on a regular basis, including before and after important occasions like general elections or catastrophic emergencies such as the COVID-19 epidemic. Businesses should assess their advertising strategies to check that they do not negatively affect the variety of perspectives and views and that the standards for targeted ads are understood.

After consulting with all pertinent parties and adhering to international human rights law, businesses should develop clear, concise advertisement and contents rules that address disinformation. Additionally, they should implement them uniformly throughout all regions. They should make sure that every policy is uniformly applied, clearly accessible, and intelligible by users while considering the unique settings in which it is used. Companies should make sure that users may choose the parameters that will determine their online experience while also giving them access to precise and useful knowledge about the specifications of company algorithms or recommendation systems. Companies should release thorough, in-depth, and contextualised transparency reports, as well as distinct findings to confront exceptional situations like the COVID-19 pandemic, that detail the measures taken to combat content that spreads misinformation or disinformation as well as the responses received to those measures, including the number of shares, viewpoints, concerns, and demands for abolishment. Users must have legal options. Businesses ought to set up internal appeals processes for a broader variety of content moderation choices and kinds of content, like coordinated impersonation. They ought to think about setting up outside supervision organisations like social media councils.

Companies, as actors on a worldwide scale, should devote more sources to improving their comprehension of the regional contexts that influence disinformation and addressing the gaps in research, lingua franca, guidelines, and services related to developing nations, minorities, and other disadvantaged populations, taking into consideration the viewpoints of regional society and groups that are the objective of disinformation. The last but not least, the UN system for protecting human rights, and in particular the Human Rights Council, has a significant responsibility to play in ensuring that all initiatives to combat false information are strongly entrenched in human rights law, together with reverence for free speech. The Council should think about launching efforts on the topic of defending and advancing human rights in the online world and convening frequent multi-stakeholder discussions with States, businesses, civil society organisations, and pertinent global and local partners.

In conclusion, the fight against disinformation is a complex issue that requires a nuanced and multi-faceted approach. Social media companies have been working to combat disinformation, but their methods have been criticized for potentially violating

freedom of expression. It's important to note that these methods are not fool proof and disinformation continues to evolve and adapt to evade detection. Also, companies may have different policies and strategies to tackle disinformation and it might vary by country as well. In summary, the fight against disinformation is complex, but it is possible to effectively combat it while also protecting freedom of expression. By taking a multi-faceted approach and prioritizing transparency, accountability, and collaboration, social media companies can play a key role in combating disinformation while also protecting the rights of individuals and groups to freely express themselves.

Bibliography

- admin, ‘GNI Home’ (*Global Network Initiative*) <https://globalnetworkinitiative.org/> accessed 31 January 2023
- ——, ‘Implementation Guidelines’ (*Global Network Initiative*) <https://globalnetworkinitiative.org/implementation-guidelines/> accessed 31 January 2023
- ——, ‘The GNI Principles’ (*Global Network Initiative*) <https://globalnetworkinitiative.org/gni-principles/> accessed 31 January 2023
- Alimardani M and Elswah M, ‘Trust, Religion, and Politics: Coronavirus Misinformation in Iran’ <https://papers.ssrn.com/abstract=3634677> accessed 31 January 2023
- ‘Association for Progressive Communications Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation’
- Aswad EM, ‘IN A WORLD OF “FAKE NEWS,” WHAT’S A SOCIAL MEDIA PLATFORM TO DO?’ UTAH LAW REVIEW
- ——, ‘The Future of Freedom of Expression Online’ 17 TECHNOLOGY REVIEW
- B. Whitton J, ‘Propaganda and International Law (Volume 72)’, *Collected Courses of the Hague Academy of International Law* (Brill 1948) <https://referenceworks.brillonline.com/entries/the-hague-academy-collected-courses/propaganda-and-international-law-volume-72-A9789028610927_06> accessed 31 January 2023
- Broadband Commission, ‘Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression’ (UNESCO 2020)

- Bull H, *The Anarchical Society* <https://link.springer.com/book/10.1007/978-1-349-24028-9> accessed 31 January 2023

- 'Center for Law and Democracy Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation' <https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/UN-SR-on-FOE-CLD-Submission-Disinformation-Mar21-final.pdf>

- 'Chilling Effects: The Impact of Social Media Platforms' Policies and Practices on Freedom of Expression'

- 'Company Assessments – Global Network Initiative' <https://globalnetworkinitiative.org/company-assessments/> accessed 31 January 2023

- 'Derechos Digitales Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation' <https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/Derechos-Digitales.pdf>

- Dixit CS Ryan Mac, Pranav, '"I Have Blood On My Hands": A Whistleblower Says Facebook Ignored Global Political Manipulation' *BuzzFeed News* (14 September 2020) <https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo> accessed 31 January 2023

- European Commission and Directorate-General for Communication Networks C and T, *A Multi-Dimensional Approach to Disinformation: Report of the Independent High Level Group on Fake News and Online Disinformation.* (2018) <https://data.europa.eu/doi/10.2759/739290> accessed 30 November 2021

- 'Exclusive: Facebook Agreed to Censor Posts after Vietnam Slowed Traffic - Sources | Reuters' <https://www.reuters.com/article/us-vietnam-facebook-exclusive-idUSKCN2232JX> accessed 31 January 2023

- ‘Facebook Restricts Myanmar Military’s Accounts for Spreading “misinformation” | CNN Business’ <https://edition.cnn.com/2021/02/12/tech/facebook-myanmar-military-intl-hnk/index.html> accessed 31 January 2023

- ‘Facebook Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation’ (Facebook) <https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/4-Companies/Facebook.pdf>

- Farquhar M, *A Treasury of Deception: Liars, Misleaders, Hoodwinkers, and the Extraordinary True Stories of History’s Greatest Hoaxes, Fakes and Frauds* (1st Printing Edition, Penguin Books 2005)

- Fetzer J, ‘Disinformation: The Use of False Information’ (2004) 14 Minds and Machines 231

- Hancock JT, ‘Digital Deception: Why, When and How People Lie Online’

- ‘Hard Questions: Who Reviews Objectionable Content on Facebook — And Is the Company Doing Enough to Support Them?’ (*Meta*, 26 July 2018) <https://about.fb.com/news/2018/07/hard-questions-content-reviewers/> accessed 31 January 2023

- ‘Inauthentic Behaviour | Transparency Centre’ <https://transparency.fb.com/en-gb/policies/community-standards/inauthentic-behavior/> accessed 31 January 2023

- ‘India: Journalists Covering Farmer Protests Charged’ (*Human Rights Watch*, 2 February 2021) <https://www.hrw.org/news/2021/02/02/india-journalists-covering-farmer-protests-charged> accessed 31 January 2023

- ‘Intervozes Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation’

<https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/Intervozes-Coletivo-Brasil.pdf>

- Jones K, 'Online Disinformation and Political DiscourseApplying a Human Rights Framework'

- Kaye D, Expression UHRCSR on the P and P of the R to F of O and and Secretariat UHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression ': <https://digitallibrary.un.org/record/1631686> accessed 22 August 2022

- Khan I, 'Disinformation and Freedom of Opinion and Expression Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (UN Human Rights Council 2021) A/HRC/47/25

- 'March 2021 Coordinated Inauthentic Behavior Report' (*Meta*, 6 April 2021) <https://about.fb.com/news/2021/04/march-2021-coordinated-inauthentic-behavior-report/> accessed 31 January 2023

- OHCHR, 'General Comment No.34 on Article 19: Freedoms of Opinion and Expression' (2011) <https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no34-article-19-freedoms-opinion-and> accessed 22 August 2022

- 'Oversight Board | Independent Judgement. Transparency. Legitimacy.' <https://oversightboard.com/> accessed 31 January 2023

- Piper PS, *Web Hoaxes, Counterfeit Sites, and Other Spurious Information on the Internet* (2002)

- 'Privacy International Submission to UN Special Rapporteur on Freedom of Opinion and Expression for Report on Disinformation' <https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/Privacy-International.pdf>

- Publishers H, 'The American Heritage Dictionary Entry': <https://www.ahdictionary.com/> accessed 31 January 2023
- 'Side-Stepping Rights: Regulating Speech by Contract' (*ARTICLE 19*, 19 June 2018) <https://www.article19.org/resources/side-stepping-rights-regulating-speech-by-contract/> accessed 31 January 2023
- 'Silencing the Messenger: Communication Apps Under Pressure' *Freedom House* <https://freedomhouse.org/report/freedom-net/2016/silencing-messenger-communication-apps-under-pressure> accessed 31 January 2023
- Wardle C and Derakhshan H, 'Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making' <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html> accessed 31 January 2023
- 'Working Groups - Forum on Information & Democracy' (*Forum Information & Democracy*) <https://informationdemocracy.org/working-groups/concrete-solutions-against-the-infodemic/> accessed 31 January 2023
- *Citizens United v FEC, 558 US 310 (2010)*
- *Salov v Ukraine* [2005] ECtHR 65518/01
- European Convention on Human Rights
- International Covenant on Civil and Political Rights
- The African Charter on Human and Peoples' Rights
- The American Convention on Human Rights
- The Arab Charter on Human Rights
- The ASEAN Human Rights Declaration

- The OAS Charter of the Organization of American States
- The SAARC Human Rights Charter
- Universal Declaration of Human Rights