

# Special Conference on Telecommunication (SpCT)

*Responsibility of social media companies for content  
hosted on their platform*



## ***Research Report***

Leiden Model United Nations 2023

*“let us sing songs of freedom together”*



**Forum:** Special Conference on Telecommunication (SpCT)

**Issue:** Responsibility of social media companies for content hosted on their platform

**Student Officers:** Medea Beridze & Hidde van der Ligt

## Introduction

Social media, and users of social media have gained a lot of power over the years, as well as the ability to influence many people within the population. Social media companies have different target audiences, and different people using them, however all in all, a very large portion of the population uses social media. They have the ability to lead a lot of consequences to these different audiences that can affect a range of things in a person's life. Consequences can differ majorly, as they can do a lot of good to people, however they can also be threatful to people, major groups, political parties, etc. Additionally, to this, social media makes it very easy for anyone to spread misinformation, which can in extreme cases be very dangerous to different groups, and can negatively affect people's lives. Information on social media, once it is posted, will always remain on the internet due to the digital footprint.

Many ask the question of is it the responsibility of social media companies for the content posted on their platform? To what extent can we blame the corporations and not the people themselves? People have the freedom of expression and freedom of speech, however, whatever is being posted on social media can be difficult to blame on the individuals, as it may be the fault of the companies themselves for not having measures to ensure people posting appropriate content.

## Definition of Key Terms

### Social Media

A platform or service used for a variety of different reasons, including communication, networking, sharing content or education.

### Misinformation

Misinformation is false, inaccurate, incomplete or misleading information which is not spread to deceive people. An example of this would be false and inaccurate information which was not published with the intent to harm or deceive the public.

### Disinformation

Disinformation is false and inaccurate information which was spread with the intent to mislead the public or influence the opinion of people.

### Algorithms

Algorithms are sets of step-by-step instructions to perform a specific task, or solve a specific issue. They are used for anything related to computers and data to solve problems effectively.



## General Overview

One in three people in the world use social media platforms in order to express their thoughts and opinions, to be entertained, to stay connected with people and many other reasons. There are a lot of benefits to social media, and the internet in general. These platforms help people with many different things and are also huge parts of our economy. However, social media does give users a very easy means to spread disinformation, hate speech, child pornography, etc. There have been many complaints about social media companies not having enough regulation and don't take enough action to prevent the dangerous activity on their platforms. Although users are able to report content on social media platforms, so that the person posting the content either gets banned from the platform, or the post gets taken down, over the years people have found many loopholes in order to post anything.

Generally, every social media platform has content policies and community guidelines, which every user is encouraged to read in order to clearly understand what is acceptable to be posted on these platforms. These policies range between platforms, however, usually the policies prohibit people from posting hate speech, any forms of harassment, violent content, illegal content, and pornography. However, due to social media platforms being self-regulated, users are able to share their own content meaning that there is a lot of content to manage. Therefore, it is extremely challenging for the social media companies to keep maintenance of their platforms.

There is also the factor which is related to legal frameworks of the countries in which social media platforms are based, and the users are based. Each country has different laws about speech, and therefore social media companies need to find ways to manage these legal factors as well, whilst giving their users a similar platform.

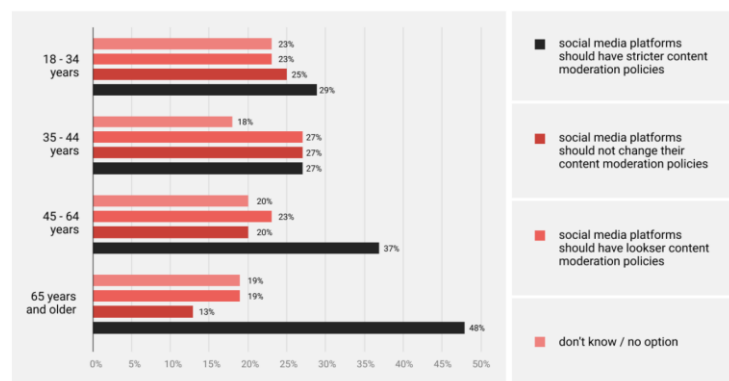
## The use of Artificial intelligence

The content moderation process is an extremely difficult process, as in the past, social media companies have been using human moderators in order to filter out inappropriate content posted by users. This has been problematic and an issue, due to the fact that this content has affected the mental health of users and moderators and it is extremely unethical for humans to watch this content. International organizations have raised concerns about the lack of standardization, subjective decisions, poor working conditions, and the psychological effects of constant exposure to harmful content for human moderators.

For this, social media companies are trying to create solutions with the use of artificial intelligence, in order to filter and monitor the content posted on their platforms. They are innovating automated practices to make social media safe, with the use of artificial intelligence based algorithms.

Content moderation done by artificial intelligence is an extremely powerful tool for removing inappropriate content from social media platforms. It has been proven to be mostly effective, and it is being used by many large social media companies, such as YouTube, Facebook, Instagram and TikTok. However, it has negatives and drawbacks, just like human moderation.

Artificial intelligence does not have a 100% accuracy rate, as algorithms can be biased, and can make mistakes. If artificial intelligence is implemented in a successful and effective manner, it would be the ideal solution to the issue of harmful content on social media platforms. It is important for social media companies to not eliminate the human factor of content filtering, as human oversight is proven to be the current most accurate moderator.





## Major Parties Involved

### Social Media Companies

These are the corporations that own and operate the platforms where users create and share content. They establish and enforce content policies, provide the infrastructure, and are responsible for user experience and safety.

### Social Media Users







Individuals who create, share, and interact with content on social media platforms. They have rights to free expression, but their actions may also be subject to platform guidelines.

### Governments and Regulatory Bodies

Governments and regulatory agencies play a crucial role in setting legal frameworks and regulations that dictate how social media companies operate within their jurisdictions. They may establish laws regarding content moderation, data privacy, and more.

### Content Moderators

These are employees or contractors hired by social media companies to review and moderate user-generated content. They are tasked with enforcing platform policies and ensuring user safety.

		 Jigsaw			 GIFCT <small>Global Internet Forum to Counter Terrorism</small>	 Microsoft
<b>system</b>	content ID	perspective API	quality filter	toxic speech classifiers	shared-industry hash database	photoDNA
<b>issue area</b>	copyright	hate speech	spam, harassment	hate speech, bullying	terrorism	child safety
<b>target content</b>	audio, video	text	text, accounts	text	images, video	images, video
<b>core tech</b>	hash-matching	prediction (NLP)	prediction (NLP)	prediction (NLP), deep-learning	hash-matching	hash-matching
<b>human role</b>	trusted partners upload copyrighted content	label training data and set parameters for predictive model	label training data and set parameters for predictive model	label training data and set parameters for predictive model; make takedown decisions based on flags	trusted partners suggest content, add content to database	civil society groups add content to database

### Commercial companies

A lot of commercial companies use social media platforms as advertising channels. They may have a stake in ensuring their content is presented in a brand-safe environment and may advocate for specific content policies. Furthermore, technological advancements like the use of AI for content moderation are enabled by major tech companies.



## Timeline of Events

- 2000s** Emergence of Social Media: Platforms like Facebook, YouTube, and Twitter begin to gain popularity, creating new spaces for user-generated content.
- 2005** YouTube Launches: YouTube is founded, allowing users to upload and share videos. This marks a significant milestone in user-generated content.
- 2006** Facebook Opens to Public: Facebook opens its doors to the public, rapidly growing its user base and becoming one of the largest social media platforms.
- 2007** Introduction of Content Moderation: Social media companies start implementing content moderation policies to address issues like hate speech, explicit content, and harassment.
- 2010s** Rise of Influencers and Brand Partnerships: Influencers gain prominence, and social media becomes an essential marketing tool for businesses.
- 2010** Arab Spring and social media: Platforms like Twitter and Facebook play a significant role in the Arab Spring movements, highlighting the potential of social media for political activism.
- 2012** Facebook's IPO: Facebook goes public, marking a new phase of scrutiny and accountability as a publicly traded company.
- 2016** Election Interference and Fake News: The U.S. presidential election sees concerns over fake news and foreign interference on social media platforms, leading to increased scrutiny.
- 2017** #MeToo Movement: The #MeToo movement gains traction on social media, demonstrating the power of online platforms in driving social change.
- 2018** Cambridge Analytica Scandal: Facebook faces a major scandal involving the unauthorized access of user data, leading to increased calls for privacy and accountability.
- 2019** Regulatory Scrutiny Increases: Governments around the world begin considering and implementing stricter regulations on social media platforms regarding content moderation, privacy, and data security.
- 2020** COVID-19 Misinformation: The COVID-19 pandemic leads to a surge in misinformation and disinformation on social media platforms, prompting increased efforts to combat false information.
- 2020-2021** Section 230 Debate: The U.S. debates potential reforms to Section 230 of the Communications Decency Act, which provides legal immunity to online platforms for content posted by users.
- 2021** Capitol Riot and Platform Bans: The storming of the U.S. Capitol prompts social media companies to ban then-President Donald Trump from their platforms, sparking a renewed debate over content moderation and free speech.
- 2021** AI in Content Moderation: AI and machine learning technologies play an increasing role in automating content moderation, helping platforms identify and remove prohibited content more efficiently.
- 2021** Global Regulatory Efforts: Thus far, various countries, including the EU, propose or enact legislation to hold social media companies more accountable for content hosted on their platforms, with considerations for AI-based moderation.



## Possible Solutions

### **Establishing a Global Content Moderation Standards and Oversight Committee**

The proposed solution involves the formation of an international committee composed of representatives from governments, social media companies, civil society, and academia. This committee's main task would be to develop standardized content moderation guidelines and set up oversight mechanisms.

To put this plan into action, a comprehensive framework of universally accepted content moderation standards would be established, specifically targeting issues like hate speech, misinformation, and explicit content. Additionally, social media companies would be required to provide regular reports detailing their content moderation practices and their effectiveness. This step aims to ensure transparency and accountability in the actions of these platforms.

Furthermore, measures would be implemented to enforce compliance. Penalties, such as fines or temporary suspensions of platforms, would be introduced for cases of severe non-compliance with the prescribed content moderation standards.

### **Ethical AI and Algorithmic Accountability Act**

The proposed solution, the Ethical AI, and Algorithmic Accountability Act, calls for the enactment of legislation that would compel social media companies to adopt ethical AI models for content moderation, and put in place mechanisms for ensuring algorithmic accountability.

### **Collaborative Verification Networks**

The proposed solution, Collaborative Verification Networks, envisions the establishment of a cooperative alliance involving fact-checking organizations, independent researchers, and social media platforms. Together, they would work collectively to verify the accuracy of information and counteract misinformation.

## Sources

Anand, S., & Ganguly, S. (2023). Should social media platforms be held responsible for the content posted on their platforms *Lexology*.

<https://www.lexology.com/library/detail.aspx?g=c5690692-7abe-4624-8447-1a301b251122>

Cusumano, M. A. (2021, January 19). *Social Media Companies Should Self-Regulate*. now. Harvard Business Review. <https://hbr.org/2021/01/social-media-companies-should-self-regulate-now>

Data, L. Y. (2023). AI content moderation for responsible social media practices. *Label Your Data*. <https://labeyourdata.com/articles/ai-content-moderation>

Forrester. (2023, February 28). Fundamental problems on social media platforms. *Forbes*. <https://www.forbes.com/sites/forrester/2023/02/28/fundamental-problems-on-social-media-platforms/?sh=723ec2587068>



Linck, A. (2021). The Digital Services Act and the role of (social media) platforms. *European DIGITAL SME Alliance*. <https://www.digitalsme.eu/the-digital-services-act-and-the-role-of-social-media-platforms/>

Stephen, K., & Stephen, K. (2021). The Social Responsibility of Social media Platforms | The Regulatory Review. *The Regulatory Review*. <https://www.theregreview.org/2021/12/21/stephen-social-responsibility-social-media-platforms/>