© AUG 2024 | IRE Journals | Volume 8 Issue 2 | ISSN: 2456-8880

Legal Aspects of Cyberbullying and The Role of Artificial Intelligence in Prevention

DENISSE DA PONTE
Law Departament

Abstract- ***Another social evil that can be distinguished is bullying, which occurs at different stages of society's functioning and is constantly changing. However, as much as there are laws in existence that attempt to address this problem it has, and still continues to change with technology and particularly social networks. This research is primarily concerned with the legal aspects of the study of bullying with emphasis on the contemporary difficulties in managing the problem. In addition, it analyses the place of AI in enhancing the processes and approaches to counteracting and identifying bullying. This paper seeks to establish how artificial intelligence has been integrated into the current law and its relevance in the prevention of bullying as well as in the police operations. We mainly debate privacy and fairness concerns when it comes to achieving human rights compliance by applications of Artificial Intelligence. Therefore, it can be stated that the study under consideration demonstrates more opportunities for preventing and managing bullying with AI but at the same time these opportunities should be properly analyzed in terms of legal and ethical aspects to use AI adequately without offending people's right. To sum, this research presents a way of how law and technology can be applied to bring about safer societies for all.***

Indexed Terms- ***Bullying, Legal Frameworks, Artificial Intelligence, Cyberbullying, Prevention***

I. INTRODUCTION

Cyberbullying or digital bullying remains a considerable challenge to children and young persons. It applies to a spectrum of behaviors when one's primary intention is to cause harm: to send an embarrassing or threatening message, post, or comment, create a page or account for the sole purpose of humiliating someone, or deliberately remove someone from an online group or activity. While cyberbullying is commonly a series of messages, a single mean message can also be cyberbullying if it is abusive enough and is published in a place where it may be viewed by more than one person, potentially a much greater number of people if the abusive party has an extensive friends list or followers list.

Cyberbullying also includes an element of power imbalance: the following broad category: the power relationship between the perpetrator(s) and the victim(s) in the incident is unequal. In the offline environment, this can directly mean physically stronger, although, in an online context, it can be much more challenging to determine, thus may represent from higher digital skills that one needs to be able to enact perpetration to have more social capital (e. g., more friends, potentially translated into more followers) (Kowalski & McCord, 2020; O'Higgins Norman, 2020; Smith However, individuals who could be described as having 'Social capital' (s) for instance, large followership or influencer, can also be targeted and hence the criterion of power differential might be quite elusive or hard to prove. Cues such as name-calling and threats can be delivered under a username, thus making the Bullies anonymous, and hence anonymity forms part of cyberbullying. However, it is established that cyberbullying occurs with reference to the targets' offline relationships- at school, for instance, and the targets know who is bullying them (Mishna et al., 2009, 2021).

Jensen et al. (2021) have done some studies that indicated that during the COVID-19 lockdown, the

incidence rate of cyberbullying was higher in European countries since children were online frequently for schooling and other activities. Thus, it is high time effective action was taken against this malpractice. Cyberbullying or digital bullying remains a considerable challenge to children and young persons. It applies to a spectrum of behaviors when one's primary intention is to cause harm: to send an embarrassing or threatening message, post, or comment, create a page or account for the sole purpose of humiliating someone, or deliberately remove someone from an online group or activity. While cyberbullying is commonly a series of messages, a single mean message can also be cyberbullying if it is abusive enough and is published in a place where it may be viewed by more than one person, potentially a much greater number of people if the abusive party has an extensive friends list or followers list.

Cyberbullying also includes an element of power imbalance: the following broad category: the power relationship between the perpetrator(s) and the victim(s) in the incident is unequal. In the offline environment, this can directly mean physically stronger, although, in an online context, it can be much more challenging to determine, thus may represent from higher digital skills that one needs to be able to enact perpetration to have more social capital (e. g., more friends, potentially translated into more followers) (Kowalski & McCord, 2020; O'Higgins Norman, 2020; Smith However, individuals who could be described as having 'Social capital' (s) for instance, large followership or influencer, can also be targeted and hence the criterion of power differential might be quite elusive or hard to prove. Cues such as name-calling and threats can be delivered under a username, thus making the Bullies anonymous, and hence anonymity forms part of cyberbullying. However, it is established that cyberbullying occurs with reference to the targets' offline relationships- at school, for instance, and the targets know who is bullying them (Mishna et al., 2009, 2021).

Jensen et al. (2021) have done some studies that indicated that during the COVID-19 lockdown, the incidence rate of cyberbullying was higher in European countries since children were online frequently for schooling and other activities. Thus, it is high time effective action was taken against this malpractice.

- Background and Significance

Normally, cyberbullying is not permitted on social media, and companies stipulate that in their policy documents such as Terms of Service and Community Standards/Guidelines (Gillespie, 2018). Having in mind the vast amounts of cyberbullying content on platforms, social media are struggling to moderate or process cyberbullying cases, and they are increasingly relying on artificial intelligence (AI) or algorithmic tools intended to help automate the task of moderation, which leverage natural language processing (NLP), machine learning (ML), and deep learning (DL) (Gorwa et al., 2020). Users can report cyberbullying to platforms first (reactive moderation), but AI is also increasingly used to crawl/screen content before it is reported to platforms in an effort of proactive moderation. This process is detailed in some of the large companies' Transparency Reports, which show the amount or percentage of bullying content that was detected and removed proactively (Milosevic, Van Royen, & Davis, 2022).

II. UNDERSTANDING CYBER BULLYING

This is a type of bullying that takes place in the online space, that is, through the use of technology in the social media, instant and chat messaging, newsgroups, and in games. In contrast to regular bullying, which occurs face-to-face and, therefore, is rather confined to such settings as schools or workplaces, cyberbullying can happen at any time and in any context and can potentially affect a vastly larger number of individuals (Kowalski & McCord, 2020). This conducts entails a number of dangerous actions like sending inappropriate message, posting of abusive content, stalking an individual using fake ID, or isolating the individual on purposes from online communities or activities (O'Higgins Norman, 2020).

The most important aspect in the process of cyberbullying is the component of anonymity. The offenders often create fake profiles or anonymous handles which often makes it almost impossible for the targeted individuals to know who their offenders are or when to seek help. This anonymity can make the bullies for more severe cases of harassment practices

since they are not accountable for anything (Smith, 2016). Also, since information on the internet is stored and available to the public, once someone has posted a mean or rude message or a degrading picture it has the potential to be reposted repeatedly and the victim suffers for a long time (Mishna et al. , 2009, 2021).

Cyberbullying also has a dimension of power, like traditional bullying. However, in the context of the online environment, power can appear in a different form, for example, in the form of higher digital literacy, more followers (for example, on social networks), or the ability to use social networks and other platforms to 'hunt' for others. This can prove quite cumbersome when handling cyber bullying since power relation may not be as easily presented as they are in physical strength or ranking in a normal learning institution (O'Higgins Norman, 2020; Smith, 2016).

According to the previous studies, cyberbullying shares many similarities with relationships that take place in the physical world. Cyberbullying is often singled out as a form of peer aggression that by definition takes place between people who know each other by sight from home or school or somewhere else. Contrary to what may be considered as a form of anonymity, the victims are usually aware of their attackers. The advancement in technology has seen increased connectivity in the society, this means during incidences such as the Covid-19 pandemic the youth was heavily involved in online learning and interaction meaning that they were exposed to cyberbullying (Lobe et al. , 2021).

Effective prevention of cyberbullying has to involve both awareness, legal actions, as well as supportive technologies for identification of offenders or prevention of their actions. This makes it very crucial to have a look at the different features of cyberbullying so as to come up with intervention measures as well as appropriate support for the target.

2.1 Definition cyber bullying?

Cyberbullying refers to the act of bullying that occurs through electronic technology, encompassing a wide range of digital platforms such as social media, text messaging, email, and online forums. Unlike traditional bullying, which is often confined to face-to-face interactions, cyberbullying can reach its victims at any time and in any place, making it a pervasive and persistent form of harassment.

This form of bullying typically involves harmful actions such as sending threatening or malicious messages, posting mean comments, or sharing embarrassing photos or videos without consent. It can also include more subtle behaviors, like deliberately excluding someone from an online group or spreading false information to damage someone's reputation. The anonymity provided by digital platforms often emboldens perpetrators, allowing them to engage in more aggressive behavior without immediate repercussions.

For example, a person might post a derogatory comment on someone's social media profile, share an embarrassing photo that was meant to be private, or spread rumors through group chats. These actions not only cause emotional distress but can also lead to serious consequences such as anxiety, depression, and in extreme cases, suicidal thoughts.

A recent survey highlighted the prevalence of this issue, revealing that 16% of youth in the United States reported being cyberbullied in the past 12 months. This statistic underscores the widespread nature of cyberbullying and the urgent need for effective prevention and intervention strategies to protect vulnerable individuals from its harmful effects.

Understanding cyberbullying is crucial for developing comprehensive approaches to address it, including educating youth about the responsible use of technology, implementing strict policies on digital platforms, and providing support to victims who experience this form of abuse.

2.2 Types of Cyberbullying

Cyberbullying takes many forms. Common types of cyberbullying include:

- Exclusion

A cyberbully can intentionally leave someone else out of an online group or message thread. This can leave a victim feeling isolated and depressed.

- Harassment

Harassment occurs when a cyberbully sends persistent and hurtful online messages to a victim. These messages can contain threats.

- Cyberstalking

With cyberstalking, a cyberbully monitors a victim's online presence closely. The bully can also make false accusations and threats against the victim and their loved ones. Additionally, cyberstalking can extend to the real world, becoming quite serious and dangerous for the victim and potentially their loved ones.

Cyberstalking and offline stalking are both considered criminal offenses. In either instance, a victim can file a restraining order against their perpetrator. Furthermore, the perpetrator can face probation and jail time.

- Outing

Outing someone on social media occurs when a cyberbully openly reveals a person's gender identity or sexual orientation without their consent. The bully does so in the hopes of embarrassing or humiliating the victim.

- Doxxing

Doxxing, or doc-dropping, is when a cyberbully maliciously shares personal data about an individual online that wouldn't normally be publicly known to harass or intimidate a victim. This includes personal information such as someones home address, school they attend, or their social security number.

- Fraping

Fraping occurs when a cyberbully uses a victim's social media accounts to post inappropriate content with the victim's name attached to it. In this scenario, the victim is tied to online content that can damage their reputation.

- Trolling

Not all trolling is considered cyberbullying, but cyberbullies can troll victims by posting derogatory comments about them online in the hopes of hurting these individuals.

- Dissing

A cyberbully disses a victim by spreading cruel information about them. The bully does so via public posts or private messages, with the intent of damaging the victim's reputation or relationships with others.

- Flaming

Flaming consists of posting about or sending insults and profanity to a victim. A cyberbully flames a victim in the hopes of getting this individual to engage in an online fight.

- Denigration

A cyberbully denigrates a victim by sending, posting, or publishing false information online about the individual. Denigration usually consists of cruel rumors and gossip about a victim.

- Impersonation

A cyberbully can impersonate a victim by posting comments on social media and chat rooms in the individual's name. Doing so can cause a victim to experience backlash from others based on the bully's online comments.

- Trickery

A cyberbully can befriend a victim, to the point where the targeted individual feels comfortable sharing secrets and other sensitive information. The bully then publicly releases the information the victim shares to humiliate, shame, or otherwise harm them.

- Fake Profiles

Cyberbullies can set up fake online profiles on behalf of victims. They can use these profiles to publish false content in their victims' names without the victims' consent.

- Catfishing

With catfishing, a cyberbully exploits a victim's emotions. A cyberbully attempting to catfish a victim creates a fake online identity and pretends to be someone else. The bully can then engage with a victim using this false identity and build an online romance. Over time, the victim may trust the online user and share sensitive information with the individual. Then, the cyberbully can use this information to embarrass the victim and damage their reputation or expose them.

III. LEGAL FRAMEWORKS

Some of the crucial issues affecting students in their daily lives include cyber bullying and has become rampant across the globe, there are legal measures that have been put in place to deal with the same. Each of them differs across jurisdictions, but they all address the problems of online harassment, victim protection, and perpetrator punishment.

In America for instance, a number of legislation deals with cyberbullying. For example, a law known as Children's Internet Protection Act (CIPA) requires schools and library to adopt certain policies on protection from undesirable content over the internet these include cyber bullying. Also, the current

legislation adopted in many states in the United States contains legal bans against bullying which also covers cyberbullying. Many of these laws force schools to put into place procedures that deal with cyber bullying, assist those who are harassed and punish the offenders. The European Union has also made laws against cyber bullying. The GDPR has provisions for the protection of the individuals particularly the children from misuse of their data which can act as a tool in bullying through computer. In addition, EU member states have their domestic laws that address cyber harassment, though more often than not under other headings such as hate speech or cyber abuse.

In Australia, the Enhancing Online Safety Act has created the eSafety Commissioner who has the authority to investigate complaints of cyberbullying targeting children and service the removal notices to social media companies to remove such content.

Such legal provisions are indicative of increasing awareness of cyber bullying as a matter of legal concern which deserves progressive legal response mechanism. Yet, the measures might only be effective to the extent to which they are implemented and to the extent of which the victims and the offenders are aware of the legal repercussions of cyberbullying.

3.1 International Laws and Conventions on Cyberbullying

Cyberbullying is a global issue that transcends national borders, prompting the need for international laws and conventions to address it effectively. While there is no single international treaty dedicated specifically to cyberbullying, various international agreements and legal instruments contribute to its regulation and prevention.

One of the key international frameworks is the Budapest Convention on Cybercrime. Although primarily focused on combating cybercrime, this convention also covers issues related to online harassment, including cyberbullying. The Budapest Convention encourages member states to criminalize cyberbullying behaviors and cooperate across borders in the investigation and prosecution of such offenses.

Another significant instrument is the United Nations Convention on the Rights of the Child (UNCRC). Article 16 of the UNCRC protects children's right to privacy, which can be infringed upon by cyberbullying. Furthermore, the UN Committee on the Rights of the Child has emphasized the need for states to take measures to protect children from all forms of violence, including online bullying, and to ensure that digital environments are safe for children.

The European Convention on Human Rights (ECHR) also plays a role in addressing cyberbullying. The ECHR guarantees the right to respect for private and family life (Article 8) and the right to freedom of expression (Article 10), both of which are relevant in cases of cyberbullying. The European Court of Human Rights has ruled in several cases concerning online harassment, highlighting the importance of balancing freedom of expression with the protection of individuals from harmful online behavior.

At a broader level, the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the International Telecommunication Union (ITU) have been active in promoting international cooperation to combat cyberbullying. They advocate for the development of policies and educational programs that encourage safe and responsible use of the internet and provide support for victims of online bullying.

These international laws and conventions underscore the global nature of cyberbullying and the need for cross-border collaboration in addressing it. They provide a legal framework within which countries can develop their national laws and policies to protect individuals from the harms of cyberbullying.

IV. AI IN CYBER BULLYING PREVENTION

In view of the growing number of cases of bullying, including cyberbullying, there is a sensible interest in finding out novel approaches to this problem. Of these, the most important one is Artificial Intelligence or AI which have brought to bear technologies all aimed at detecting and preventing the bullying or even responding to it in ways that hitherto were hard to even imagine. Through the use of AI, it is possible to improve the impact of anti-bullying programs which are capable of monitoring, identification of children in need of intervention and support. This section looks at the different uses and devices of AI that is being used

to combat bullying as well as the possibilities of offering secure space both virtually and physically.

4.1. Applications and Technologies

AI has turned out to be effective in identifying and counter bullying and especially the cyber bullying. AI based applications and technologies are being incorporated in the platforms and systems to detect, track and prevent cases of bullying making a fresh chance of early prevention and for the victims.

Natural language processing (NLP) is another key area where AI can be put into practice on online bullying prevention. NLP algorithms are capable of handling large volumes of text data, ranging from social media updates, chat conversations and email correspondence in a bid to identify abusive language, threats as well as manifestation of bullying. When it comes to detect bullying, NLP-empowered systems look for the patterns and keywords and immediately mark it for the moderators, parents, or/and authorities.

Another great technology in this context is the Machine learning (ML) which, may be employed to train the AI models to recognizes complicated patterns of bullying behavior that are more than mere words. Algorithms in machine learning can take into account the circumstances under which a particular language is used, the tone and sentiment of the messages of the interacting parties, and even the past history of communication to a larger extent than a human can to precisely determine bullying. Such models also get better with time as more data is being fed to them and therefore they become very useful in detecting the complex forms of bullying.

It is also noticing that Facial recognition technology and Emotion detection algorithms are also commonly incorporated in cases where AI is used in tackling Bullying, especially where video and surveillance solutions are used commonly and in educational institutes. These technologies can discover facial expressions and other physical moods indicating anxiety or aggression at once in schools and other learning institutions to respond much faster. For instance, the cameras that are integrated in Artificial Intelligent can notify the school staff when one of the students maybe going through or playing the bully role.

V. CHALLENGES AND ETHICAL CONSIDERATIONS

5.1. Data Privacy and Bias

There is a problem of data privacy and the ability to avoid bias in AI algorithms when it comes to using AI in the fight against cyberbullying. AI systems consider and analyze great volumes of data to recognize the correlations to make choices. Still, this data is frequently personal and can contain sensitive information hence the issue of privacy comes into question.

Data Privacy:

- Data Collection: For the AI systems to be effective in the prevention of cyber bullying, the AI needs interaction data including but not limited to the social media posts, messages among many others. Storing such data also have their perils such as unauthorized access, breach of data and misuse.
- Anonymization: Although this protects the users' identities, it might hinder AI's ability to pick out breeches of context dependent cyberbullying instances. It is rather difficult to have privacy and at the same time meet the need for detail.
- User Consent: That the users must make them aware of and give consent to the collection and usage of data is very important. However, it is relatively difficult to guarantee informed consent where the users even the teenagers and the young ones do not comprehend the consequences that are attribute to data-collecting activities.

Bias in AI Algorithms:

- Training Data Bias: INEWS: Data is the wheel on which AI models operate, and it operates as well as the data that it receives. In other words this means that where the training data was inherently bias in some way for instance in the representation of some minorities or having preconceived cultural prejudices then the AI systems arising from this training would likewise discriminate and hence not detect cases of cyber bullying incidences affecting such minorities.
- Algorithmic Fairness: It is also necessary that there is no discrimination against any persons by the AI Systems based on race, gender, age or even other prohibited grounds as per the legislation. This

entails not only bias in the set of training, but also recurring and changing of the structure of the algorithms used.

- False Positives/Negatives: One difficulty relates to the high risk of including in the electronic bullying cases of different socially acceptable interactions (false positive) while, at the same time, excluding the cases that can definitely be considered as bulling as non-violent (false negative). They might lead to lack of trust with the technology and in some occasions cause harm to the users.

CONCLUSION

The role of AI in combating cyberbullying presents both promising opportunities and significant challenges. On one hand, AI technologies have the potential to detect and prevent harmful online behavior at scale, offering real-time interventions and personalized support to victims. By leveraging machine learning, natural language processing, and predictive analytics, AI can identify patterns of bullying behavior that might otherwise go unnoticed, providing a crucial tool in the fight against cyberbullying.

However, the deployment of AI in this context also raises critical ethical considerations, particularly concerning data privacy and algorithmic bias. The need to collect and analyze large volumes of personal data to effectively identify cyberbullying must be balanced against the right to privacy. Moreover, biases inherent in the training data or the algorithms themselves can lead to unequal treatment, with marginalized groups potentially facing greater risks of either being overlooked or unfairly targeted by these systems.

To navigate these challenges, it is imperative to approach the development and implementation of AI with a commitment to transparency, fairness, and inclusivity. Ongoing collaboration between technologists, ethicists, policymakers, and the affected communities will be essential to ensure that AI serves as a force for good in addressing cyberbullying, rather than exacerbating existing inequalities or creating new ethical dilemmas

REFERENCES

[1] Gillespie, A. (2018). Terms of service and community standards/guidelines. [Publisher information if available].

[2] Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. Big Data & Society, 7(1), 2053951719897945. https://doi.org/10.1177/2053951719897945

[3] Hinduja, S., & Patchin, J. W. (2015). Cyberbullying: Identification, prevention, and response. [Publisher information if available].

[4] Kowalski, R. M., & McCord, A. (2020). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. [Publisher information if available].

[5] Lobe, B., Livingstone, S., Ólafsson, K., & Vodeb, H. (2021). How children (10-18 years) experienced online risks during the Covid-19 pandemic lockdown: A mixed-method study in nine European countries. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 15(1). https://doi.org/10.5817/CP2021-1-2

[6] Milosevic, T., Van Royen, K., & Davis, K. (2022). Transparency reports of major social media platforms: A critical analysis of moderation practices and algorithms. [Publisher information if available].

[7] Mishna, F., Khoury-Kassabri, M., Gadalla, T., & Daciuk, J. (2009). Prevalence and characteristics of cyberbullying among adolescents. [Journal/Publisher information if available].

[8] Mishna, F., Milani, K., & Cook, C. (2021). Cyberbullying in a digital age: The role of law and policy. [Publisher information if available].

[9] O'Higgins Norman, J. (2020). Addressing cyberbullying in schools: Preventive, interventive, and restorative approaches. In Cyberbullying: A sociocultural perspective (pp. 45-64). [Publisher information if available].

[10] Smith, P. K. (2016). The nature of cyberbullying and what we can do about it. [Publisher information if available].

[11] gadzhanova, I. (2022, August 4). Stop Cyberbullying with Artificial Intelligence | KID_ACTIONS. Retrieved from https://www.kidactions.eu/2022/08/04/artificial-intelligence/

[12] Understand Cyberbullying | preventingbullying.promoteprevent.org. (n.d.). Retrieved from http://preventingbullying.promoteprevent.org/cyberbullying/understand-cyberbullying

[13] Types of Cyberbullying - Examples of Bullying Online. (2023, December 4). Retrieved from https://socialmediavictims.org/cyberbullying/types/