

Unmasking The Digital Shadows: Exploring the Ethical Parallels of Cyberbullying and Deepfake Technology

Shahzad Ali^a, Prof. Dr. Nurul Hartini^b, Dr. Nono HeryYoenanto^c, Prof. Dr. Melati Binti Sumari^d

a. Department of Developmental and Educational Psychology, Faculty of Psychology, Universitas Airlangga

b. Department of Educational Psychology and Counselling, Faculty of Education, University of Malaya, Kuala Lumpur, Malaysia

ABSTRACT

This research examines the link between deepfake technology and cyberbullying, their similarities and relation to immorality on social media platforms. Deepfake technology, which uses artificial intelligence, alters media content to spread misinformation and cyberbullying, which uses technology to harass and defame others. To connect these practices, the study employs a multidimensional immorality model, comprising deceit, aggression, utilitarian manipulation, unrighteousness and indulgence, to explain their motivations. The research highlights the ethical aspects of artificial intelligence and the abuse of digital tools. While deepfake technology has ethical uses, its misuse can lead to data integrity issues and social harm. The study also offers key insights for policymakers to establish regulatory measures and intervention approaches to reduce digital harm. The research suggests the importance of detection technologies, legal protections and digital literacy to create a safer and more ethical online environment.

KEYWORDS

Deepfake technology, Disinformation, Cyberbullying, Immorality, Future Aspects, Artificial Intelligence, Digital Literacy

Introduction

In recent years, the phenomenon of fake and sensitive information in the areas of study has become a very dangerous factor in the development of society at the individual and collective level, in the form of influence on businesses, political and democratic systems, (Borges, et al., 2019; Qayyum, et al., 2019). This type of information is a source of a bad environment of conflicting narratives which may mislead people, (Aldwairi & Alwahedi, 2018; Jang & Kim, 2018). The use of social media as a means of spreading fake and misleading information has become prominent because users tend to share information without confirmation because there's insufficient evidence, (Figueira & Oliveira, 2017; Tong, 2017).

In today's world, many people rely heavily on social networking platforms like Facebook, YouTube, Twitter and Instagram for accessing news and information, (Anderson, 2018). These media allow for the quick sharing of text, images and video content without verification, and so the potential for the spread of misinformation is greater, (Anderson, 2017). As explained by (Borges, et al. 2019; Britt, et al., 2019), these platforms can be readily taken advantage of to spread half-truths or propaganda aimed at individuals and groups. As a result, the spread of misleading information can have a profound impact on society, including a regulatory burden on governments and policymakers, (Westerlund, 2019).

According to (Anderson, 2018; Qayyum, et al., 2019), as well as (Zannettou, et al., 2019), there is doubt that it is currently possible to speak about the post-truth period of the world, based on the increased usage of information warfare and digital disinformation. In this environment, anonymous agents control flows by posting fake news through social media networks for the purpose of changing people's perceptions, views, beliefs, and attitudes toward particular topics, (Westerlund, 2019). Likewise, cyberbullying works on similar principles; using techniques like spreading rumors against the adversaries, which form yet another channel for maligning the subject.

According to the definition of cyberbullying presented above it involves deliberate aggressive actions towards other people carried out in an ongoing nature by one or more people using electronic devices and internet facilities such as social networks, often targeting users who do not have the ability to protect themselves, (Smith et al., 2012). However, as compared with other forms of bullying, the relatively easy execution of cyberbullying is particularly remarkable based on its inherent features. Unlike physical bullying, cyberbullying requires a single stroke on an online social platform thereby allowing the transfer of hatred material regarding the subject at hand.

This activity is visible from a potentially large number of social media users thereby communicating without needed restrictions or express permission, (Ali et al., 2022). It is worth labeling cyberbullying as a major form of misuse since it groups all the types of content that can be used for

miscommunication purposes, as well as the specific ways of communication that include emails, phone calls, messaging applications, chat rooms on social networks, and other elements within the domain of social media environments. Based on their behavior during their social media usage, they can be classified into unique roles; these include; cyber victims, cyberbullies, or cyberbullying perpetrators depending on the interactional context of cyberbullying, (Kowalski, & Limber, 2013; Ali, et al., 2024).

While research on misinformation, cyberbullying and other emerging digital threats has been steadily increasing, they are usually considered separately, with little academic attention to their conceptual and operational intersections. Though cyberbullying has already received much attention as a type of online abuse and digital victimisation, deepfake technology has only recently emerged as an advanced form of synthetic media manipulation through artificial intelligence. Recent studies suggest synthetic media is increasingly being used not just for political propaganda but for personal abuse and defamation, and, therefore, overlaps with cyberbullying (Vaccari & Chadwick, 2020; Hancock & Bailenson, 2021).

Moreover, the integration of artificial intelligence technologies with social media affordances has led to hybrid forms of digital abuse, in which deepfakes can enhance the reach, realism and emotional effects of cyberbullying (Chesney & Citron, 2019; Paris & Donovan, 2019). But the current literature has yet to conceptualise this phenomenon as a coherent analytical problem. This is especially important in today's digital climate where disinformation can both deceive and harm.

Hence, there is a pressing academic need to consider deepfake technology and cyberbullying as an intersection, which involves the use of misinformation to harm others through comparable technological, psychological and social processes. Through this integration, this research seeks to offer a more holistic perspective on how novel technologies are transforming online violence and ethics in digital environments.

Deepfake technology utilizes artificial intelligence techniques that enable one person's face to be placed or substituted for another's in video or pictures to create an outright fake reality. This technique is often used in

images or videos with world leaders or world celebrities of the film industry and other famous personalities. Such individuals are exposed to the risks of deepfake manipulation of their visual content to create fake and pseudo-real realities, that visually look like real and authentic ones, (Korshunov & Marcel, 2018). According to (Maras & Alexandrou, 2019), deepfake technology systematically uses large databases to synthesize videos, images, and voice. It is a process where one replaces authentic facial or vocal expressions with synthetic ones, creating a transformation.

By so doing, such alterations are made possible through the use of datasets which consist of lengthy arrays of information belonging to a specific targeted person, (Koopman, et al., 2018). Such an action as changing one's personal data, such as videos, or images, depends on artificial intelligence. Based on knowledge of how the minds of human beings work this technology has resulted in the development of deepfake technology. More importantly, a key element of artificial intelligence is machine learning, a scientific technique that facilitates systems to learn from data, making them intelligent, (Maras & Alexandrou, 2019).

The term deepfake subsumes another two ways of functioning. The first strategy entails the use of an individual's set of videos and images, in order to build a new but credible-looking story. The second approach is however a lot more dangerous due to its simplicity, and the frequency with which this plan can be set into action, by various stakeholders. Kantian analysis and utilitarian analysis converge in the idea that the aim of each method is to cause suffering to a particular individual. Different applications have appeared with the aim of letting users know more about such a twist as deepfake technology. Out of these, Reddit is a recognized platform that has striven to help its users reduce the number of deepfakes by demystifying the mathematical computations involved in creating them.

According to (Mattern, et al., 2017) the assumption has been made that after visiting these applications and going through the registration process, the user with just basic programming and machine learning knowledge can get a basic idea of deepfake technology. This acquired knowledge enables the generation of manipulative graphic and audio material in terms of

pictures, motion pictures, and photography that relate to virtually any person. Unfortunately, the capacity for technology and its risks are seen clearly because technology supports and enables violent actions. One such application we will discuss is superimposing unknown person's faces with adult film actors faces, which may result in situations resembling sextortion and blackmail, where the material is used as a form of pressure against the targeted person, (Delfino, 2019).

Furthermore, (Stover, 2018) has identified, this technology does not just end with personal vendors and actually, it has found its way into the political sphere which is just as nefarious. Currently, it is used to create fake videos, voice reproductions, emails and images becoming a powerful instrument with which blackmail political competitors for the purpose of achieving certain goals. As a result, this technology has become an evil negative factor in the modern world. In the same vein as the previous practice, as indicated by (Athanasasou, et al., 2018 & O'Higgins, 2020) cyber bullying constitutes one aspect of the malign use of Information and Communication Technologies (ICTs).

This means that the on-line usage can cause security, safety and welfare of young generation at risk since misuse of these technologies can make a person a candidate for a potentially destructive violation of trust. These are vulnerabilities which are usually exploited through fake and altered images and videos on the social media applications so as to cause what is referred to as digital victimization or cyber bullying. The high availability of these Information and Communication Technologies worsens the problem, given that the offenders use technologies to practice the vice.

Background of Deepfake Technology and Cyberbullying

The practice of steering from one's face is actually dated back to one of the most primal uses using technology and this was done using some figure known as President Abraham Lincoln of the United States in 1865 although it was very crude compared to the current advanced technology. In this historical episode, allusion was made to manipulate a print portrait of President Lincoln and garner an artwork on John Calhoun of 1852. And this

prototype was studied as the preparation of modern deepfakes, and as evidence of the timeless interest of people in changing or distorting images for different purposes, (Chawla, 2019). Over the past few years, an anonymous user with the nickname deepfake published videos with Hollywood female actors such as Scarlett Johansson, Maisie Williams, Taylor Swift, Aubrey Plaza, and Gal Gad Site.

These videos included superimposing their faces in sexually suggestive or even pornographic videos through the aid of deepfake technology algorithms prevalent on the Reddit website. However, these videos were pulled down effectively soon after they were posted online. This episode was a kind of public exposure and as a starting point marks the depth of fake reality which is capable of transforming and twisting facial structures. For this reason, it attracted significant attention from the international media; this stressed the possibility of using such a technology in sharing of unwanted releases of naked material involving celebrities. as noted by (Gardiner, 2019), deep fake technology was created by an anonymous software engineer who released a general kit that enabled new producers of such videos to become effective using it. Deepfakes could be best described as the synthetically created video, image, or audio information depicting a person as doing or saying something they could not.

Thus, the development of high technical applications for the creation of superior software requires the integration of professional competencies and capabilities to enhance interaction with the software products among the users. This imperative is informed by the knowledge that, deepfake technology was first introduced into the public domain by the United States Department of defense through what is referred to as the Defense Advanced Research Projects Agency (DARPA). This technology was considered as a threat to society because of the negative effects it has because of unauthorized alteration of images, videos, and audio by people who lack the skills of a professional photoshop.

Three research scholars at the University of Washington issued a video on July 2017 with an aim of creating awareness of the effects of deep fake technology. Later, another video with the former President of the United

States, Donald Trump, was published by spectators deepfake. This video features President Trump interacting with a Belgian delegation and talking to them about America pulling out of the Paris Accord on Climate Change. These two recent incidents concern the former US. The President was an example of how any other person, including influential personalities, is at risk of the misuse of this technology, (Siekierski, 2019).

The word bully was introduced a long time ago and people used it starting in 1693 if the records are to be believed. However, fore, documented examples of this form of aggression can be comprehensively dated back even earlier than this, (Maree, 2005). For over three centuries since the creation of the term bullying cases of this phenomenon were considered to be part of the behavior of people, and given the specifics of individual cases, it was difficult to predict the undesirable consequences of bullying, (Campbell, 2005). In the early era of bullying, the different acts included theft, confusion in lunch money, disruptive separation, verbal intimidation, threats, and mimicking someone's clothes or trend of dressing. Bullying in the present age has manifested into many styles and one of them is cyberbullying. With the complete advent of modern advanced technologies and friendly social sites, it is quite possible for anyone to quelched bullying behavior towards individuals, (Miriam, 2010).

Cyberbullying involves acts performed on social media platforms with the use of technology discrete front-end interfaces and with the main objective of posting malicious rumors about one's rivals. These technologies and activities are carried out over the internet more often with the spectators' purpose of embarrassing others in a cruel and aggressive way leaving the victims vulnerable. Other subtypes of cyberbullies include those who engage in the repetitive harassment of their targets, via, sending threats, meanness, and interference with the psychological health of targets. The images of the victims can also be isolated or altered, and combined with abusive language or a humorous statement to bring a feeling of shame, (Rifauddin, 2016). Similarly, finding out the people behind deep fake technology, such as cyberbullies is tasked since their operations are attributed to the World Wide

Web. This leads to higher levels of online and the rewards of heinous acts because the perpetrators can never be traced, (Hanish, et al., 2004).

Such sites as YouTube, Twitter, Facebook, Snapchat, and Instagram have poorly established policies when it comes to creating an account for an individual thus encouraging malicious persons to bully other people online. Since reregistration is not a rigid protocol, various age groups of users can easily create accounts and feel that their activities are not policed online by certain authorities. This perception can make people use these platforms to be bullies, (Milosevic, 2018).

In a general way, we can define cyberbullying as being the act of using special social applications to spread rumors, threats, and jokes to adversaries. As pointed out by (Nasrullah, 2014), the Director of the Center for Safe and Responsible Internet Use in America, it occupies a wider band. It includes discrimination, humiliating someone, defamation and making remarks that are towardward, and revealing one's personal and private details with the clear knowledge and intention of bringing down persons on social media and the internet.

According to (Coloroso, 2006;2015) and (Nuha, 2014) there are four types of cyberbullying:

1) Cyberbullying involves the use of technologies where there is a power difference between the two or more people involved. This power imbalance is also a unique characteristic that illustrates that bullying behaviors do not stem from a contest or parity, be it race, race rivalry, or sibship. Nevertheless, the bullies themselves are older and more powerful, as well as more professional in some or all of the aspects. It allows professions of their verbal ability, bodily power, size, standing in society, and depending on the situation, even their social rank within a particular race or nationality.

2) Cyberbullying is intentional and premeditated aggression on the part of the perpetrator. It is not usually described as impulsive, but as enjoying the suffering of the victims. Evidently, the victim in cyberbullying also suffers emotional violence in addition to a physical one. Criminals enjoy it when they are inflicting horns on other people, they get a kick out of such a feeling.

3) The pattern of a cyberbully engaging in aggression by threatening the targets is another characteristic of cyberbully-victim encounters.

4) This systematic act of putting force and brutality to power by using force and instilling fear to other persons is also termed as terror bullying because its main goal is to instil fear and terror to victims.

Cyberbullying and Deepfake as the Retification of a Lie

The authors have described Cyberbullying as a deliberate and continuous action where in a person employs present day technologies such as Mobile phones and computers to harm, another person, (Hinduja & Patchin, 2009; Patchin & Hinduja, 2015). Such behavior patterns mostly occur covertly and are initiated by the offender with the purpose of sabotaging the victim's relationship. This is not a favorable situation for the victim, but one has to face it through the help of widely available electronic devices from the side of the cyberbully. Before proceeding, it should be understood that today, millions of people use these devices, (Grant, 2012). Although deepfake technology is used in a series of manipulative technologies to smear an individual's reputation and relationships through algorithmic software, these manipulated media files are later shared on available social media platforms such as Facebook, Instagram, Twitter, and YouTube, thus increasing their availability, (Maras & Alexandrou, 2019).

At one time or another, deepfake content including videos, images and voices gets to the traditional print media or online newspapers, (Stover, 2018). More recent examples include Hollywood celebrities such as Scarlett Johansson and Taylor Swift, whose images were posted on pornographic videos, despite being previously taken down from reddit. Also, in politics, people experienced fake media in which political leaders such as former presidents of the USA, Barack Obama and Donald Trump to mention but a few, fell prey to fake audioclips and in this case, their voices were edited, (Stover, 2018). As has been seen both cyber bullying and deep fake technology pose the contemporary challenge due to their downstream social impacts. According to (Mills & Robson, 2019) misinformation was described as the act of passing wrong information with a view to misleading people.

In the interest of letting the general public know and the general non-technical Joe and Jane, both these terms fall under one category known as misinformation. Both synthetic media and deepfakes, by their nature, are capable of aggravating the problem of dis-information and misinformation, primarily through the process called infocalypse, (Schick, 2020). Infocalypse is defined as information, which is incredibly difficult to differentiate as being either accurate or fake, (Helmus, 2022). These techniques have been widely used by different political actors, both individual politicians, as well as political parties. A useful historical event that deserves mention is the explosion on the battleship USS Maine that took place in Havana harbor in 1898. This event was presented in such a way that people were meant positively to influence the population base since a newspaper in the United States spread the information. This erroneous reporting, in turn, led to war between the USA and Spain, (Chesney & Citron, 2019).

As (Kedar, 2020) noted that specific misinformation was spread in the 1930s as part of Hitler's plan to stage an attack on Poland. He said that Poland was actually invading Germany, which he said, caused the Second World War. Similarly, the head of the United States in the period of 1941; Franklin D. Roosevelt fed the population with misinformation contained in the German documents dubbed to be planning on eliminating religious practices across the globe. However, it was later found out that this statement by President Roosevelt was either a lie or misleading information. Although it was one of the misleading postures, this announcement was of great importance in the United States to join the Second World War against Germany and all Axis powers. No such documents were found in the investigations conducted after the war. However, understanding the U.S.'s decision to enter the war is a challenging endeavor, (Hemming, 2019).

Several factors have significantly contributed to the increased risk of misinformation dissemination through Synthetic Media and Deepfakes in the past: A) The ability to create pictures and or videos with this technology making it easy to manipulate. B) The ability to distort images and videos based on one's subconscious notion of people. C) The capacity and speed at which fake news travels due to social media walls. According to behavioral science

studies, users of these platforms are inclined to put a lot of faith in information quickly. As stated in these studies, the social media environment is more likely to utilize information that is easier to process, and is considered credible and persuasive. Such information seems to create a major focus in people's minds compared to information that is published in regular newspapers. Regardless of whether such information is genuine or false and not officially substantiated, people are quick and eager to believe it, (Greifeneder, et al., 2021).

Similar to deepfake technology, cyberbullying has also thrived in social media due to people's internet connection after the evolution of present-day technology. It is now possible for this technology to be in the hands of a large population of people within the so-called 'technological society'. The speedy advancement in development especially in the technology segment is a daunting issue for parents, families, schools, and societies when particularly addressing issues about control of children and young people on the internet, (Grant, 2012). These advances in technology have created a convenient approach to such kinds of networks, being instrumental in the development of distinctive old-style bullying into a more advanced form of bullying known as cyberbullying, (Shariff, 2009). Children, adolescents, and students in particular have revealed new ways of being cruel to each other through the use of new technologies, the World Wide Web, and especially through social networks. This change has revolutionized the techniques of teasing or bullying classmates in school and college.

In the pre-social media period, students may have opted for physical aggression, for example, stealing from fellow students or even using humiliating language on fellow students. But, with the help of social media applications which include the likes of Twitter, Facebook, Instagram, and others, bullying classmates found a tool, (Calvert, 2001). Thus, the use of SMT has become a new hidden tool, with which individuals can cyberbully others. Such behavior can also be nonphysical, at any time, and does not require the use of conventional weapons; hence the effectiveness of such behavior transcends and is not limited to physical violence, (Connors, 2018).

Cyberbullying encompasses various forms, as described by (Willard, 2007), who coined multiple types of this behavior:

1. Flaming; is the act of forwarding messages motivated by anger and aggression that involves furious texting or messaging.
2. Harassment; means developing and corresponding with a person using messages that can cause anxiety and nuisance through a computer system including emails or sending messages on social and instant channels.
3. Cyberbullying; is where one unworthy character of another, through the development of images and names that offend the victim's reputation, honor, and respect on Facebook.
4. Impersonation; means sending immoral messages with the aim of lowering the status of fellow social media users.
5. Exclusion; is a process where one has to delete a person from their group of social media chat.
6. Cyber-stalking; is defined as launching an attack on the subject's name on social networks to tarnish their reputation on the social circuit.
7. Outing & Trickery; according to (Satalina, 2014), involves posting on social media platforms other's secret pictures and information f forwarding texts driven by anger and aggression, typically involving furious texting or messaging.

This final type, described by (Willard, 2007) in the book *Cyberbullying and Cyberthreats*: One of these is a reaction strategy for Outing & Trickery Symptom, also known as 'Online Social Aggression, Threats and Distress' show that the issue is complex in the digital world.

Purpose of Current Study and Research Question

The current study is grounded on a synthesis of existing literature on deep fake technology of cyberbullying. Through integrating these lines of literature, the study intends to generate useful knowledge for scholars so as to conduct comprehensive analysis on these variables. Although the cybersecurity issue and particularly the focused topic of cyberbullying has been researched extensively and is discussed daily, deepfake technology is more novel and has received growing interest among scholars.

Deep fake and cyber bullying are both defined as significant threats to a person's prestige, honor and esteem. Falsification through deepfake affects an individual through providing him/her with fake multimedia content while cyberbullying is the use of social media platforms to cause harm to an individual's reputation. Drawing from the existing literature on deepfake technology and cyberbullying, the following research questions have been formulated:

1. What are the similarities between deepfake technology and cyberbullying?
2. What is a relationship between immorality or ethical issues and deepfake technology and cyberbullying on social media?

These research questions will form a basis for further exploration and will be designed to look for similarities and examine the ethical aspects of two massive phenomena.

Methods

The current study employs a qualitative literature review method to explore the theoretical and practical link between deepfake and cyberbullying. Due to the exploratory nature of the research questions, a narrative synthesis approach was used to synthesise the evidence from interdisciplinary literature sources, including research from the areas of information technology, media studies, psychology, and cybersecurity.

To enhance transparency and rigour in the review process, a search strategy was defined. We searched the literature using key databases such as Google Scholar, Scopus, Web of Science and ScienceDirect. The search terms employed combinations of the following keywords: deepfake technology, cyberbullying, synthetic media, misinformation, digital harassment and online victimization. The search strategy primarily targeted peer-reviewed journal articles, conference proceedings, and reputable reports from 2015 to 2024 to capture key developments and advances in the field.

The selection criteria for the studies were: (a) articles that focused on deepfake technology, cyberbullying, or both related to digital or social media environments; (b) studies that mentioned misinformation, cyber harm or

ethical considerations of new technologies; and (c) full text publications in English. The exclusion criteria were as follows: (a) non-academic sources like blogs or opinion articles with no empirical or theoretical foundation; (b) studies not related to digital spaces; and (c) overlapping sources.

After selection and screening of the literature, a qualitative synthesis method was used. The papers were thematically synthesized to determine common themes, parallels and ethical considerations of deepfake technology and cyberbullying. This allowed classification of the literature under various themes, such as misinformation techniques, types of online victimization, and ethical considerations of technology abuse.

This approach provides a holistic and synthesis of the interaction between deepfake technology and cyberbullying, and identifies gaps and areas for future research.

Results and Discussion

The first research question of the current study seeks to identify the what are the similarities between deepfake technology and cyberbullying? According to the literature review carried out on this topic, deepfake technology entails the use of rumors with the assistance of technology, including artificial intelligence particularly at the technological front in the internet, (Albahar & Almalki, 2019). In the same manner, cyberbully technology is used to egg some person or a group on social media, by changing their pictures or adding shameful comments that reduces that persons or groups respect and esteem, (Baier, et, al., 2019). Cyberbullying and deepfake are applied to besmirch the image and reputation of the targeted person or a group, and, at times, deepfake technology is utilized to blackmail, including political and other public figures, (Harris, 2018). Similarly, the study found that cyberbullying has engaged students especially in blackmailing through naked images, (Tanrikulu & Erdur-Baker, 2021).

The first hypothesis of this work will involve asserting that deepfake technology is very much akin to cyberbullying whereby both methods convey false information with the sole intent to cause harm to an individual or a particular group of folks. New studies hold that social media and internet are

responsible for publishing fake news with the aim of tarnishing the reputation, enhancing Cyber victimization. Thus, it might be expected that subsequent investigations using both indexes as the parameters within the same system will reveal much more specific similarities and analyze the impact of these factors on the adult's perceptions and ways of thinking. Such research may explain how directly or indirectly, these phenomena are responsible for influencing the culture of the society.

The second and final research question of this study focuses on the What is a relationship between immorality or ethical issues and deepfake technology and cyberbullying on social media? Prior research has examined the relationship between immorality and personality characteristics and found a good correlation with participants of cyberbullying. Morality is a basic element of personality, while ethics and immorality are the systematic of the emotionally, cognitively, and behaviorally established moral standards learned from societies. This framework establishes the person's interior character and directs him or her actions according to the degree of morality or immorality. Immorality, as a dimension of behavior, encompasses five key elements: These he sums up with deceit, aggression, utilitarianism, unrighteousness and indulgence.

In turn, any of these elements in some or other variation can explain an individual's propensity towards evil. It has been found that this aspect of character is a reliable predictor of negative behaviors, and it can be reasoned that cyberbullying is one form of this characteristic, which makes an individual likely to become a cyberbully on social media such as YouTube, Twitter, Instagram, and Facebook a social media that has been identified to have played a significant role in cyberbullying acts according to earlier studies by, (Sun, et al., 2020).

On the other hand, (Menesini, et al, 2013) asserts that morality has an insignificant relationship with cyberbullying. This view is supporting the observation made by (Gini, 2006; Menesini & Camodeca, 2008). However, the analysis of the synergy between immorality and cyberbullying is more complex and will be discussed in a separate article that is dedicated to the tendencies of personality change. Contrary, the technological development

contributed to the fast-paced change and brought out technical issues as a negative phase of the modernization. This technology has a prospect of immense positive functionality for the development of artificial intelligence but the majority of users have taken it as an opportunity of data deception with certain communities, groups or ethnicities in mind. For example, deepfake has been applied in multimedia sectors of television, movies, films, and other social media such as Facebook, Twitter, Instagram, and Snap Chat. This popularity stems from its advantageous in that data alteration can be done for the purposes of certain communities, societies, individuals and ethnic groups, (Gardiner, 2019).

But it is important to stress that sharing deepfake videos is unlawful and subpar since the use of data is immoral to create negative consequences and interfere with people's lives. It also has implications for the validity of other documents that have been forged or altered since vouchers for fake documents, (Gong, et al., 2020). A list of subjects concerning deepfakes, which has been compiled by (Zannettou, et al. 2019) consists of criminals, companies, industries, governments and other groups that use this technology with the ultimate motive of harming the targeted individuals, providing fake information and projecting their aspirations. It is necessary to note here that, to some extent, this technology can be considered as entertainment or a source of fun, (MacKenzie & Bhatt, 2020).

Conclusion

Thus, the present research, drawing on the literature, confirms a rapid advancement in data science, machine learning and wide-spectrum networking. This advancement is used to enhance content creation agility, particularly multimedia content for social networks. As a result of this revolution, the so-called deepfake technology has been created.

On the other hand, it is important to acknowledge that in recent times technology has been employed to defame people's images and ideas and to spread disinformation through social networks like Facebook, Twitter and Instagram. These have provided a convenient means to bully or edit a person's image and video using unfair, unethical and sometimes illegal

practices. This is concerning in terms of societal safety, especially for children and teenagers, and calls for better legislative and regulatory measures around the world.

Being aware of when the audience is exposed to deepfake content, however, is not the solution. In particular, social media platforms should take proactive measures to address this phenomenon, in the form of platform policies and technological solutions to identify and block manipulated media content. On the other hand, these technologies can also be employed for constructive purposes, such as identifying criminal activities and preventing illegal activities. Social media providers have to utilise scientific and technological methods to prevent the spread of disinformation and content that promotes cyberbullying.

A study by (Westerlund, 2019) has proposed solutions to deal with the challenges posed by deepfake disinformation, such as regulatory policies, law enforcement, training experts, countermeasures and raising public awareness.

Lastly, it is important for individuals to report cyberbullying and deepfake technology misuse. Joint efforts between individuals, organisations and governments are needed to create a safer digital environment for current and future generations.

Suggestions for prospective scholars

Here are some suggestions for combating the issues of deepfake technology and cyberbullying in society, which can serve as valuable research areas for future scholars and academia:

Enhance the Humanities Digital Learning

Investment in large-scale multi-layered learning programs to promote awareness and facilitate skill training to users, particularly among young people, in perceiving and evaluating online content. Improving digital literacy helps users to discern true and false information.

Future studies should adopt experimental approaches to assess how digital literacy programs can enhance users' capability to cope with deepfake and cyberbullying. Furthermore, longitudinal research is also encouraged to

explore the impact of digital literacy on online behaviour and susceptibility over time, especially among young people and university students.

They must also learn to design and apply effective detection instruments.

Researchers should consider adopting integrated research methods that combine deepfake computational detection systems with user-focused qualitative methods (such as interviews and focus groups), to understand the effectiveness and users' perceptions of deepfake detection technologies.

More and more promoting the development of deepfake detection. This can aid in the recognition and removal of such content with the goal of reducing the circulation of fake content.

Future research could also employ machine learning and experimental validation methods to evaluate the performance, effectiveness and real-world feasibility of deepfake detection models.

Enact Clear Legal Frameworks

Tightly cooperate with other lawyers and law makers in coming up with exacerbated tangible and enforceable laws and policies regarding deepfake creation and dissemination. Boundaries must be put in place on the individuals who will be found guilty of deploying this technology in the wrong manner.

Future studies should employ comparative legal research designs to study legal approaches to deepfake technology and cyberbullying in different countries, and to assess the impact of legal responses to these problems. Moreover, empirical research employing survey and experimental designs should be conducted to determine awareness of these laws, as well as their effects on online user behavior.

Spread the word on Ethical AI and Deep Learning

Promote more development and networking explorations that fill the goal of working on the principles of ethical practice in artificial intelligence and deep learning. Further research into the positive and negative possible consequences of using these technologies, then how the technology can be developed and applied in an ethical fashion.

Researchers should also adopt interdisciplinary research designs, drawing on computer science, psychology and communication. These approaches can be used to address ethical issues and experimental methods may be employed to understand the impact of AI-generated content on user perceptions, particularly in relation to cyberbullying and misinformation.

These four suggestions relate to major areas for further study and intervention, that might help reduce the adverse impact of deep fake technology and cyber bullying in society.

Overall, these approaches require rigorous and multidisciplinary research to inform individuals about the challenges and opportunities of digital technologies.

Implications for Policymakers

With the growing intersection of deepfake and cyberbullying as forms of online abuse, policymakers are key to addressing their negative impacts. There is a clear need to build robust and flexible policies that specifically target the production, sharing and misuse of synthetic media in the digital sphere.

Future policy-oriented research should take an evidence-based approach, such as policy evaluation studies and impact assessments, to assess the efficacy of existing legal frameworks in preventing and managing deepfake misuse and cyberbullying. Comparative international research can also inform other countries of best practices and regulatory challenges.

Policymakers should also engage with digital developers, teachers and researchers to develop policy interventions that combine legal measures with online literacy programs and AI-driven detection tools. This holistic approach can enhance preventative measures and promote a robust digital environment.

Contribution by Author(s)

This study's author(s) equally contributed their participation in working and responsible for confirming as the true authors put their efforts into the contribution of searching data, analysis of rephrasing articles, selection, screening, and drafting of study manuscript.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement:

This research is supported by the Indonesian Endowment Fund for Education (LPDP) on behalf of the Indonesian Ministry of Higher Education, Science and Technology and managed under the EQUITY Program (Contract No. 5494/B/UN3.LPPM/PT.01.03/2025).

References

- Albahar, M., & Almalki, J. (2019). Deepfakes: Threats and countermeasures systematic review. *Journal of Theoretical and Applied Information Technology*, 97(22), 3242-3250.
- Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. *Procedia Computer Science*, 141, 215-222.
- Ali, S., Hartini, N., Yoenanto, N. H., & Paramita, P. P. (2024). Exploring the genetic underpinnings of bullying: a contemporary analysis of scholarly investigations. *Society*, 12(2), 614-630.
- Ali, S., Hartini, N., & Yoenanto, N. H. (2022). Characteristics Of Bullying Perpetrators And Bullying Victimization At The Indonesians Schools; A-Review. *Journal of Positive School Psychology*, 3392-3404.
- Athanasiou, K., Melegkovits, E., Andrie, E. K., Magoulas, C., Tzavara, C. K., Richardson, C., & Tsitsika, A. K. (2018). Cross-national aspects of cyberbullying victimization among 14–17-year-old adolescents across seven European countries. *BMC public health*, 18(1), 1-15.
- Anderson, K. E. (2017). Getting acquainted with social networks and apps: Social Media in 2017. *Library Hi Tech News*.
- Baier, D., Hong, J. S., Kliem, S., & Bergmann, M. C. (2019). Consequences of bullying on adolescents' mental health in Germany: Comparing face-to-face bullying and cyberbullying. *Journal of Child and Family Studies*, 28(9), 2347-2357.
- Borges, L., Martins, B., & Calado, P. (2019). Combining similarity features and deep representation learning for stance detection in the context of checking fake news. *Journal of Data and Information Quality (JDIQ)*, 11(3), 1-26.
- Britt, M. A., Rouet, J. F., Blaum, D., & Millis, K. (2019). A reasoned approach to dealing with fake news. *Policy Insights from the Behavioral and Brain Sciences*, 6(1), 94-101.
- Calvert, C. (2001). Off-Campus Speech, On-Campus Punishment: Censorship of the Emerging Internet Underground. *BUJ SCi. & teCh. I.*, 7, 243.

Campbell, M. A. (2005). Cyber bullying: An old problem in a new guise?. *Journal of Psychologists and Counsellors in Schools*, 15(1), 68-76.

Chawla, R. (2019). Deepfakes: How a pervert shook the world. *International Journal of Advance Research and Development*, 4(6), 4-8.

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, 107, 1753.

Coloroso, B. (2015). the bully, the bullied, and the not-so-innocent bystander. Akses pada, 25.

Coloroso, B. (2006). *Penindas, Tertindas, dan Penonton; Resep Memutus Rantai Kekerasan Anak dari Prasekolah hingga SMU*. Jakarta: Serambi Ilmu Pustaka, 29-31.

Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Aff.*, 98, 147.

Connors, J. (2018). Turning Suicide into Homicide: Can You Be Bound by Your Text Messages. *J. High Tech. L.*, 19, 477.

Delfino, R. (2019). Pornographic deepfakes—revenge porn’s next tragic act—the case for federal criminalization. Available at SSRN.

Figueira, Á., & Oliveira, L. (2017). The current state of fake news: challenges and opportunities. *Procedia Computer Science*, 121, 817-825.

Gardiner, N. (2019). Facial re-enactment, speech synthesis and the rise of the Deepfake. Edith Cowan University, Theses 2019.

Gini, G. (2006). Social cognition and moral cognition in bullying: What's wrong?. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, 32(6), 528-539.

Gong, D., Goh, O. S., Kumar, Y. J., Ye, Z., & Chi, W. (2020). Deepfake forensics, an ai-synthesized detection with deep convolutional generative adversarial networks. *International Journal*, 9(3).

Grant, N. P. (2012). Mean girls and boys: The intersection of cyberbullying and privacy law and its social-political implications. *Howard LJ*, 56, 169.

Greifeneder, R., Jaffe, M., Newman, E., & Schwarz, N. (2021). The psychology of fake news: Accepting, sharing, and correcting misinformation (p. 252).

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, behavior, and social networking*, 24(3), 149-152.

Harris, D. (2018). Deepfakes: False pornography is here and the law cannot protect you. *Duke L. & Tech. Rev.*, 17, 99.

Hanish, L. D., Kochenderfer-Ladd, B., Fabes, R. A., Martin, C. L., & Denning, D. (2004). Bullying among young children: The influence of peers and teachers. In *Bullying in American schools* (pp. 163-182). Routledge.

Helmus, T. C. (2022). Artificial intelligence, deepfakes, and disinformation: A primer (p. 24). RAND.

Hemming, H. (2019). *Agents of Influence: A British Campaign, a Canadian Spy, and the Secret Plot to Bring America Into World War II*. PublicAffairs.

Hinduja, S., & Patchin, J. W. (2009). Cyberbullying fact sheet: A brief review of relevant legal and policy issues. Cyberbullying Research Center. Available at http://www.cyberbullying.us/cyberbullying_legal_issues.pdf.

Hui, J. (2018). How deep learning fakes videos (Deepfake) and how to detect it?. Medium Corporation, 28.

Jang, S. M., & Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions. *Computers in human behavior*, 80, 295-302.

Kedar, H. E. (2020). Fake news in media art: fake news as a media art practice vs. fake news in politics. *Postdigital Science and Education*, 2(1), 132-146.

Koopman, M., Rodriguez, A. M., & Geradts, Z. (2018, August). Detection of deepfake video manipulation. In *The 20th Irish machine vision and image processing conference (IMVIP)* (pp. 133-136).

Korshunov, P., & Marcel, S. (2018). Deepfakes: a new threat to face recognition? assessment and detection. *arXiv preprint arXiv:1812.08685*.

Kowalski, R. M., & Limber, S. P. (2013). Psychological, physical, and academic correlates of cyberbullying and traditional bullying. *Journal of adolescent health*, 53(1), S13-S20.

MacKenzie, A., & Bhatt, I. (2020). Lies, bullshit and fake news: Some epistemological concerns. *Postdigital Science and Education*, 2(1), 9-13.

Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, 23(3), 255-262.

Maree, K. (2005). Bending the neck to the yoke or getting up on one's hind legs? Getting to grips with bullying. *Acta Criminologica: African Journal of Criminology & Victimology*, 18(2), 15-33.

Matern, F., Riess, C., & Stamminger, M. (2019, January). Exploiting visual artifacts to expose deepfakes and face manipulations. In *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)* (pp. 83-92). IEEE.

Menesini, E., Nocentini, A., & Camodeca, M. (2013). Morality, values, traditional bullying, and cyberbullying in adolescence. *British Journal of Developmental Psychology*, 31(1), 1-14.

Menesini, E., & Camodeca, M. (2008). Shame and guilt as behaviour regulators: Relationships with bullying, victimization and prosocial behaviour. *British Journal of Developmental Psychology*, 26(2), 183-196.

Milosevic, T. (2018). Protecting children online?: Cyberbullying policies of social media companies. MIT Press.

Mills, A. J., & Robson, K. (2019). Brand management in the era of fake news: narrative response as a strategy to insulate brand value. *Journal of Product & Brand Management*.

Miriam D. Martin, Suicide and the Cyberbully, *EzINE ARTICLES* (May 21, 2010), <http://ezinearticles.com/?Suicide-and-the-Cyberbully&id=4328660>.

Nasrullah, R. (2014). *Teori dan Riset Media Siber (Cybermedia)*. Jakarta: Kencana Prenadamedia Group.

Nuha, I. S. (2014). Hubungan Perilaku Bullying dengan Perilaku Asertif pada Santriwati (Doctoral dissertation, UIN Sunan Ampel Surabaya).

O'Higgins Norman, J. (2020). Tackling bullying from the inside out: Shifting paradigms in bullying research and interventions. *International journal of bullying prevention*, 2(3), 161-169.

Paris, B., & Donovan, J. (2019). Deepfakes and cheap fakes. *Data & Society*, 50.

Patchin, J. W., & Hinduja, S. (2015). Measuring cyberbullying: Implications for research. *Aggression and Violent Behavior*, 23, 69-74.

Qayyum, A., Qadir, J., Janjua, M. U., & Sher, F. (2019). Using blockchain to rein in the new post-truth world and check the spread of fake news. *IT Professional*, 21(4), 16-24.

Rifauddin, M. (2016). Fenomena cyberbullying pada remaja. *Jurnal Ilmu Perpustakaan, Informasi, dan Kearsipan Khizanah Al-Hikmah*, (4)1, 25-44.

Satalina, D. (2014). Kecenderungan perilaku cyberbullying ditinjau dari tipe kepribadian ekstrovert dan introvert. *Jurnal Ilmiah Psikologi Terapan*, 2(2), 294-310.

Shariff, S. (2009). *Confronting Cyber-Bullying: What Schools Need to Know to Control Misconduct and Avoid Legal Consequences*. Cambridge University Press. 32 Avenue of the Americas, New York, NY 10013.

Schick, N. (2020). Deep fakes and the infocalypse: What you urgently need to know. Hachette UK.

Siekierski, B. J. (2019). Deep Fakes: What Can Be Done About Synthetic Audio and Video?. Library of Parliament= Bibliothèque du Parlement.

Smith, P. K., Del Barrio, C., & Tokunaga, R. S. (2012). Definitions of bullying and cyberbullying: How useful are the terms?. In *Principles of cyberbullying research* (pp. 26-40). Routledge.

Stover, D. (2018). Garlin Gilchrist: Fighting fake news and the information apocalypse. *Bulletin of the Atomic Scientists*, 74(4), 283-288.

Sun, X., Chen, L., Wang, Y., & Li, Y. (2020). The link between childhood psychological maltreatment and cyberbullying perpetration attitudes among undergraduates: Testing the risk and protective factors. *PLoS one*, 15(9), e0236792.

Tanrikulu, I., & Erdur-Baker, Ö. (2021). Psychometric properties of a Cyberbullying Inventory for university students. *British Journal of Guidance & Counselling*, 49(3), 494-507.

Tong, C., Gill, H., Li, J., Valenzuela, S., & Rojas, H. (2020). "Fake news is anything they say!"—Conceptualization and weaponization of fake news among the American public. *Mass Communication and Society*, 23(5), 755-778.

Turbert, K. (2008). Faceless bullies: Legislative and judicial responses to cyberbullying. *Seton Hall Legis. J.*, 33, 651.

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social media+ society*, 6(1), 2056305120903408.

Westerlund, M. (2019). The emergence of deepfake technology: A

review. *Technology Innovation Management Review*, 9(11).

Willard, N. E. (2007). *Cyberbullying and cyberthreats: Responding to the challenge of online social aggression, threats, and distress*. Research press.

Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019). The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *Journal of Data and Information Quality (JDIQ)*, 11(3), 1-37.