Library Progress International *Print version ISSN 0970 1052*
Vol.44 No.3, Jul-Dec 2024: P.4628-4737 *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

Mitigating Human Rights Violations Caused by Deepfake Technology

Aman Gautam[1], Dr. Rupak Kumar Joshi[2], Dr. Aastha Narula[3], Neha Sharma[4]

[1]Research Scholar, School of Law, Sharda University, Greater Noida, U.P.
[2]Assistant Professor of Law, ICFAI University, Dehradun, Uttarakhand
[3]Assistant Professor of Law, Christ University, Bengaluru
[4]Assistant Professor of Law, Delhi Metropolitan Education and Research Scholar, Bundelkhand University, Jhansi, U.P.

How to cite this article: Aman Gautam, Rupak Kumar Joshi, Aastha Narula, Neha Sharma (2024). Mitigating Human Rights Violations Caused by Deepfake Technology. *Library Progress International,* 44(3), 4628-4637.

ABSTRACT
Deepfake based on artificial intelligence is a very effective tool that can generate highly realistic fake media including photos, audio and video which, in most cases, cannot be recognized as fake data. While deepfakes may offer special utility in the realm of entertainment as well as creative works, they are a grave threat to the safety and liberty of people all over the world. Deepfake technology is briefly discussed in this abstract to reflect on the ethics and consequences that they present in society as well as the possibility of using them to offend human rights. Forced credibility, deceit, and slander become more acute with the help of deepfakes, which infringe on people's right to privacy and, therefore, negatively impact their reputation. Therefore, the intentional application of deepfakes for illicit pornography and the dissemination of incorrect data have adverse consequences ranging from a detrimental effect on people's psychological conditions to reduced trust in media and threats to democracy. Again, women, members of racial or ethnic minorities together with political dissenters are the categories of people who are most vulnerable to deep fakes manipulation for harassment and exploitation.
Deepfake technology raises many problems which require an anti-measure approach that considers legal, ethical, and technological factors. Preventing and controlling the spread of deep fake materials is closely related with the development of methods for detecting and verifying deep fake content.
In addition, enhancing the critical thinking skills coupled with media literacy fosters people's capacity to discern between truth and fake information; this reduces the harm that deepfakes pose within the community and the larger society. Legal systems thus need to be extended to encompass guidelines of dealing with deepfakes such as determination of who is legally responsible for producing and circulating fake content and the measures to be taken, punitive and preventive, against those who lose their rights through being victims of deepfakes.

Keywords: Deepfake Technology, AI, human rights, Identity Theft.

Introduction

Deepfake technology emerges as a form shift for the new age of 'seeing and hearing' media products in the digital world. Internationally known as deepfakes well, they are improved by AI, are able to create realistic fake videos and audio that cannot be distinguished from a real one. However, when applied in a negative manner it becomes injurious, evident by the violation of human rights as well as the interruption of the society's cohesiveness witnessed.
Thus, deepfake technology can be an intricate problem for many spheres as it is connected with violation of people's rights to privacy, spreading of fake information, and loss of trust in visual information. While deepfakes are capable of influencing citizens in such a way as to influence electoral decisions, any person depicted here is a link through which others, regardless of their intentions, can violate their privacy and unleash the mentioned vices on them. Besides, deepfakes are utilized for weaponization to distort material that degrades democratic values and

feeds the populace's poor-quality information. The aforementioned form of fake news is even worse in the sense that it is very easy to spread lies on these online platforms and with so much reliability that most people cannot distinguish between the reality and the fake news.

Reciprocally, since negative applications of deepfake technology impact human rights such as the right to privacy, the right to information, and such rights, mitigating it will need the interventional efforts of technology, law, ethics, and human rights. However, technology has provided some of these solutions in justifiable concepts such as; Deep fake detection algorithms and Digital authentication mechanisms which are also not free from weaknesses and ethical issues. In addition, to nail down the misuse of deep fake while not discouraging one's freedom of speech or creativity, the law must improve reckon ability for such heinous acts.

While the possibilities mentioned above are quite diverse, it is still possible to identify the following challenges that require addressing in the context of the discussed technology: Therefore, this work of research seeks to deliberate on the mitigations to eradicate the human rights abuses brought by Deep fake technology. Therefore, this paper holds the goal of making some recommendations for deepfake threats identification with inevitable prevention and with due consideration for the ethical, legal, and technological aspects and preserving human rights. When understanding the various forms and tactics of deepfake manipulation, the stakeholders will be able to develop the best precaution mechanisms to avoid the adverse effects of manipulation on the credibility of information systems in the current world.

Understanding Deepfake Technology

Deep Learning and AI Algorithms

ANN and AI are two of the most important processes within the contemporary computational environment based on the ground where it has to learn data for the purpose of establishing patterns, coming up with the rational decisions and for solving the tasks that can mainly solve the human mind.

Artificial Intelligence (AI): AI as a prefix of computer science has aimed its objective to design systems which are capable of solving problems in a way akin to theirs. Among them some of the functions are solving, comprehending, perceiving, learning, understanding natural language and deciding. It is important to state that the function of AI systems is to mimic different aspects in the human brain and, therefore, thinking.

1. Machine Learning (ML): , Machine learning is a sub of the AI domain in which computer algorithms give a probability of guessing the output of unknown inputs based on the originals fed into it that are improved over time without being programmed. The use of Modern Learning algorithms can enable a computer to learn how to work with extensive databases to distinguish between the patters, predict the result, and search for associations.

2. Deep Learning: It is an area of machine learning that uses ANN's of at least one hidden layer for the purpose of attaining the result of identifying representations of the data available. The elements of deep learning are multiple models, which can help include particular patterns as input directly. This allows deep learning models to separate several layers and patterns of the given complicated data set.

3. Artificial Neural Networks (ANNs): Artificial Neural Networks, ANN are computing structures developed in terms of their form and functioning resemble with the neuronal network in the brain. ANNs consist of interconnected nodes (neurons) organized into layers: and input layer, one or more hidden layer, and finally the output layer using the neuron based structure. When using the context of the widely accepted connectionist model of the brain, each of the neurons within the aforementioned net would have to feed through the current signals, apply an activation function on those signals, and then pass on the new signal that emerges on the next layer of neurons.

4. Training and Learning Process: The training in deep learning involves inputting the categorized training data to the neural network, adjusting parameters of the instance such as weights and bias, use of optimization algorithms such as gradient descenge and minimizing the error which is the difference of the output of the neural network and label of the training data. This process is known as back-propagation in which the computed error is back propagated through the network to adjust the parameters for the improved result of the model.

5. Common Deep Learning Architectures: Thus, deep learning includes several architectures designed for particular tasks that can be listed as follows:

Most prominently, CNNs have been utilized in tasks such as, object recognition and also in the area of computer vision.

Other types of networks that are very efficient basically when there is manner of sequence as in language natural processing and time series analysis comprise of Recurrent Neural Networks (RNNs).

Autoencoder for lossy image compression and data enhancement or data generation and Super Resolution GANs/SRGANs for Generative Adversarial Network.
Reflection for natural language comprehension and generation, and perhaps a new high in applying transformers to powerful tasks such as translation and abstraction.

How Deepfakes Work?

Deepfake can be defined as, generative adversarial network, commonly referred to as GANs that in combination with autoencoders, make fake images, videos and audio known as synthetic media. To begin with, approximately several thousands of images or video data of the target individual(s) are collected; this data is employed for training the deep learning system. Pre-processing is the second step of data mining process where data is prepared with respect to the specification of the model. In the course of the training, one of the components generative model generally uses GAN learns to generate synthetic samples, and the other component known as discriminator also learns identifying the synthetic data from the actual data. It also facilitates the enhancement of the generator network to come up with better deepfake contents when engaging in the adversarial training. To increase the plausibility of content generation, certain techniques such as style transfer and autoencoder are applied. Cleaning steps are carried out to fine-tune any of the parameters that might be deemed necessary for the specific deepfake back to the environment of the original video. Since deepfake content when created, can be posted to social medial platforms or dispersed through the internet, the issues arising are those to do with privacy, security, and authenticity of videos and audio visuals. As deepfake technology becomes more sophisticated, it also becomes necessary that there exist ways through which people can get to know methods of reversing or preventing deepfake technology from being used in a wrongful manner and or let the public know the repercussions that out come from the misuse of deepfake technology.

Types of Deepfake Media

Deepfake technology enables the creation of various types of synthetic media, including:As for the definition, it is possible to state that synthetic media refer to various types of fakes, created with the help of deepfake.

1. Face Swapping: Regarding deepfakes, it translate to the swapping of one image or a video with the figure or face of another personality. It is somewhat possible to describe it as a method in which it is possible to transfer a person's image to someone else and more often than not one will get relatively good impressions at that.
2. Voice Cloning: The mode of fake speech with the voice representation of the targeted is included in voice cloning deepfakes. There is actually some suggestion that deep neural networks [can] get down to the level where one can actually make realistic recordings that sound like the intended speaker.
3. Body Manipulation: On the same note, in the same genre of the technique called deep fake, movement and poses of the human being can also be synthesized in a video. Therefore, to receive perverted visual information, deepfake authors make alterations to such parameters as position, movement, and pose of the subject.
4. Text Generation: Recoding to describe such applicability, It also applies in such areas where deep learning models can write text that looks like has been written by a certain person. The kind of deepfake media used in such a method is to respond to an article, a social media update, or a specific message that was generated from the individual.
5. Audio-Visual Deepfakes: The AV deepfakes are produced using the face swap idea and further an addition of an authentic voice cloning version to match the person's lip and fake spoken content as it can be performed to make a person utter something he or she never said. For this reason, deepfakes of such types might be astonishing and result in moments when people get confused.

Human Rights Implications of Deepfake

Privacy violations and Surveillance Concerns

Deepfakes are complete anathema to privacy and lead to high levels of surveillance and this causes a myriad of ethical and human rights issues. Deepfake technology poses a major threat for the right to privacy because it enables the creation of authentic synthetic media of people without their permission. The facial features and voices of people can be replaced and grafted onto pornographic scenes, without consent and to the detriment of the person's character and well-being. This manipulation takes away individuals' agency over their matters, thus reducing consent and self-governance in the digital age's affairs. In addition, deepfake videos are very effective in identity theft and impersonation, hence increasing the vulnerability of the people's privacy and security. Self,

it is not only personal privacy infringements but also surveillance that is promoted by the deepfake technology at hand since the created content can be employed for any purposes ranging from political spying, corporate espionage, and social engineering,. Deepfake surveillance is a clear example of how technology is used for expansion of authoritarianism and terrorism of democracy , freedom of speech and of people's trust with reference to such values as privacy and autonomy. Meeting these privacies and surveillance issues needs effective legal frameworks, technology measures, and information and awareness campaigns to manage the consequences of deep fake manipulation and to promote the proper human rights in the info society.

Impact on Reputation and Defamation

Self-generated deepfakes are highly dangerous as they allow people to risk others' reputation and expose them to defamation. Due to the tactic of blending the visuals and sound, deepfake videos make it possible to generate fake stories, portray a person doing something immoral or unlawful or making denigrating statements. The involved videos can therefore easily be uploaded and shared on social media and other online applications and forums, thus leaving the affected people with little to no influence over the dissemination of the fake information and the debunking of nasty rumors. Furthermore, deepfake technology distorts the perception of reality and disrupts the citizens' trust in the genuine visual information, as well as complicating the opportunities to control the accounts of the credibility of digital . The losses of deepfake-based defamation could be severe emotional and social disorientation, social exclusion and job loss due to the defamation of personalities by malignant manipulation. Litigation against deepfake defamation is a question in light of current defamation legislation because it appears that the modern approach does not allow for coverage of such problems and the distribution of fake materials . In order to bring prevention of circulating deepfake-based defamation, it is necessary to imply both technical measures to recognize and check content and legal measures to strengthen responsibility and liability of offenders and legal reforms, as well as the informatization of the general public and critical thinking in media consumption. Recognizing the reputational and defamation risks posed by deepfake technology nowadays, it is possible to prevent the attacks on individuals' reputation and respect the principles of truthfulness and freedom from false accusations in digital society.

Psychological Harm and Emotional Distress

The employment of fake content or deep fake has psychological effects and causes emotional pain to the victims involved. The presence of deepfakes in person's daily life can easily make the person feel insecure and violated because deepfakes can be really convincing, which results in severe emotions like anxiety, fear and even paranoia. Database: Deepfake-based harmonic experiences and impersonation lead to severe psychological harm to the victims because the former is malicious exploitation of the later to harm others or disseminate false information. Ironically, through the means of deepfake content, the sense of power and control is lost as people cannot prevent the spread of fake content regarding them in mainstream media. In addition, it is very difficult to identify real or fake (generated) content, which can result in permanent distrust and doubt in other people and the interactions in social networks, media, and everyday life experiences. Those who have been targets of deepfake leading to emotional distress are also limited in terms of remedy since it is difficult to find support for the harm caused by digital manipulation. Prevention of the deepfake technology's psychological effects calls for extensive approaches that are deemed legal, psychological, informative, and transformative. That is the reason why the psychological impact of deepfake technology should be analyzed as well as the emotional suffering of people; by doing so, stakeholders will be able to build a significant level of success in terms of victims' support, prevention of mental health deterioration, and ultimately the overall well-being of individuals navigating social media platforms within the Information Age.

Threat to Vulnerable Communities

Deepfake is now being used to much more vulnerable groups worsening the existing disparities and actually presenting a real threat to person's wellbeing and security. This is because such revenues are gained through deceit and manipulation of vulnerable individuals, especially from the minority, women, and low-income earners as they hardly have access to enough resources, social support, and legal help. Deepfakes have been used in Cases of sexual harassment, revenge porn, and Identity theft against women which in turn increases gender-based violence and violation of their rights and dignity. Sophisticated biases are postulated and sustained by deep fake technology and escalate social stigma against the minority groups. Also, deepfake vulnerabilities may be intensified depending on one's disabilities and the mental problems they may be suffering from, as these are taken advantage of for ill motives. Deepfake content that is created to target a specific sensitive group means that these minority

groups will continue receiving this content creating a feeling of powerlessness, lack of trust and social isolation. To prevent deepfake targeting on vulnerable groups, an interconnected perspective that targets on the social, economical, and systematic factors that enshrine vulnerability and exploitation of such groups should be formulated. This involves increasing the legal protection of these individuals, increasing their access to supporting services and facilities, educating the media and the public on appropriate portrayal of profiles with such disabilities and developing positive online communalities and online skills. To solve the problems that occur to vulnerable populations under the influence of deepfake manipulation, stakeholders would be able to advance population protection, people's rights, and fair treatment in the digital realm.

Undermining trust in media and the democratic process

Deepfake technology harms media and democratic processes since such manipulated content introduces dystopian elements of the postmodern media landscape, escalates the circulation of mis- and disinformation and weakens the population's trust in the accuracy and reliability of the information environment. The availability of deepfake videos and audio threatens the recognition of true and fake information, which makes it problematic for a person to decide whether the given information is real. This blurring of the line between reality and fiction is the second problem that reduces the trust in media sources as people begin doubting the credibility of the video and audio materials that are being used in the news, posts, and other types of content. Deepfake technology also contributes to the proliferation of fake news and deceitful information since artificially generated material can be utilized to create fake stories and ideologies, twist the words of politicians, and influence people's perceptions. Due to the realistic look of the fake content that deepfake videos present, the new type of fake news confusing the society and sowing discord in democratic countries. Furthermore, the use of deepfake for developing political slogans, propaganda, political attacks and influencing elections and its impact on the society and the democracy is dangerous because the manipulated heavy content destroys the faith of the people on the systems, the election results. Preventing the continued use of deepfake manipulation and rejuvenating faith in media as well as democratic activities cannot be possible through a single approach but through collaboration of innovation, policies, schools and society. Thus, by combating the sources of distrust and falsehood, stakeholders may contribute to improving the level of information credibility and the quality of democratic processes in the country.

Technological Solutions for Detection and Authentication of deepfakes

It is, therefore, worthy a noble cause to seek ways and means of how to discover the deep fake and any gadget or system that can recognize the deep fake and content authentication system is useful and important in preventing manipulations and safeguard information domain. Combating the fake fake news deepfakes, the real deepfakes along with the techniques to find the real deepfake videos and real deepfake audios Smart algorithms using Artificial Intelligence, Computer Vision, and Digital Forensic methodology have been developed.

1. Forensic Analysis: Forensic encompasses the detection of shifts in the elements of the appearance and sound of content, which point to the fact that the content has been tampered with. This may include comparing the locations of the facial features, comparing the types or sources of illumination used in the source and generated images, or comparing modulations of the inherent shadow in images to modulations made artificially, during the deepfake generation phase. Some of the forensic techniques can be rather useful in the identification of the particular content that might be involved in the production of deepfake.
2. Machine Learning Algorithms: To fight frauds, deepfakes and their patterns, as well as the characteristics of the manipulated aspects are described by machine learning algorithms. The majority of the machine learning methods, including supervised ones, use sets of legal and fraudulent samples in an attempt to identify deepfakes. CNN and RNN architectures can also be utilized in deepfake detection since they have a deep learning capacity of locating features from videos and sounds.
3. Behavioral Analysis: Behavioral approaches to analysis are based on deviations in people's behavior or their physiological response that might contain the information about the fake manipulation. This may include trying to examine features such as the patient's face, eyes, or the sounds emitted in a pursuit of attempting to see or find any peculiarities or disparities. Some of the behavioral analysis techniques may afford technical measures, enabling to determine their admissibility to the digital documents or artifacts, based on people's feeling's consideration.
4. Blockchain and Digital Watermarking: Based on the possibilities of eliminating the falsification of content and determining its source, the cryptographic instruments in blockchain technology and methods of digital watermarking. the integration of blockchain and water marking techniques provides the ability

of hashing media files and linking it with a unique digital signature that avails an evidence or a trace of its authenticity. They are technologies that possess aspects of identifying and preserving the unlawful alterations and total wellness of electronic assets.

5. Multi-Modal Fusion: Higher integration of different modalities involves the data from different modes such as video, audio and context to boost and rely on the deepfake detection approaches. The multi-modal fusion strategies complement the operating other modalities to enhance the performances of the detection algorithms and overcome the demerits that may arise from the specificity of the detection strategy. The following are some of the specific ways by which fusion approaches could assist in enhancing the DF detection mechanisms' resilience against adversarial actions and complicated methodologies of signal manipulation:

Legal Framework in India

Information Technology Act, 2000: Some of the legislation from the Information Technology Act 2000 and amendments are related to the aspect of Digital Communication which includes The Electronic Authentication Act, Digital Signature Act, and Cybercrime Act. The sections 66 and 66 D of the Act relates to offences done with the help of computer, relate to identity theft; cheat by personation which means using computer resource for this purpose and punishment for cheat of impersonation.

Indian Penal Code (IPC): Although, there are other IPC provisions that may be applicable to deepfake manipulation such as Libel (S. 499 IPC), Forgery (Ss. 463-470 IPC) and Personation(s. 419 IPC). These sections may be required to address the application of deepfake in provoking or contributing to affection or incidences of fraud.

Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021: These rules known as Intermediary Guidelines and Digital Media Ethics Rules have been notifies by Ministry of Electronics and Information Technology. They also require that, the intermediaries, block sections of the content which is considered unlawful; such as invasion of privacy or identity theft.

Although Deep fake manipulation is not completely unregulated in India, the following legal instruments could be of some help: Currently the laws governing social media include the Information Technology Act, the Indian Penal code, Intermediary Guidelines, and the Digital media ethical code; there are however, some laws that are still awaiting approval and they include the Data protection bill of India and the election laws of India. However, their adequacy in effectively tackling the challenges posed by deep fakes remains limited for several reasons: Still, in some sense, judges are as far as it is enough for deep fakes by the following reasons;

1. Undefined Lack of Specificity: Notwithstanding, since there are no provisions from free software ,there is no way to prevent fake technologies including deep fake technology or manipulation; this paper has been unable to find any law that can penalize any of the above said technologies. After all, when adhering to the rules of the traditional crimes, the provisions of defamation, on forgery, and on impersonation could be useful while dealing with computer frauds even though it lacks connection with several of the elements or features of the deepfake manipulation. That is enforcement it is particularly so for the statutes because they are slightly less refined and elaborate than other legal instruments utilized in legal procedures.
2. Undefined Technological Complexity: Basically, this technology is easily used by deepfake and the extent that this technology does not allow one to distinguish between a real person and a fake person increases exponentially. Thus, it can mean that legislation, as well as Indian laws and the respective legislation, might be less developed to fit the changes in some of the rapidly growing technological industries which, in its turn, can lead to the emergence of the stated problem and the appearance of the problem of defining the concrete deepfake manipulation.
3. Undefined Enforcement Challenges: Therefore, although the modern laws do not have deepfake manipulation as such incorporated into the concepts of theoretical thinking in the given aspect, one might attempt to stop similar sins in real life – and this might turn into a real mission. Some of the challenges that police encounter include: This is for instance becomes difficult for police which were at first addressing completely opposite sorts of threats and did not have exposure or tools, concerning the type of threat of the present sort, can the police track the origin of deep fake or the people behind it or offer enough evidence to show that the offenders wanted to it.

4. Undefined Intermediary Liability: The Intermediary Guidelines and the Digital Media Ethics Code amend provisions concerning the mechanisms for eradicating unlawful information, for example, deepfakes. However, the problems are hit when it comes to the intermediary and liability and censorship for advice to self-regulation interpretation there are chances that the platform will avoid the concern altogether and instead over-compensate on the censorship of the real content.
5. Undefined Privacy Concerns: Since Deepfake infringes on privacy in the sense that the creation of the fake deepfake video involves the individual's image and other data, the Pull of Privacy is usually present. It has been mentioned earlier that the current laws are not entirely void of provision for the protection of privacy; However, for the same reason, it becomes pertinent to state that there might also be even more legal loopholes that are tolerated by laws and are not against legal provisions of countries that have laws governing the protection of different types of data in absence of enforcement agencies and thus citizens are in very unfavorable conditions or are in adverse position that can be more uncon
6. Undefined Need for Comprehensive Legislation: Now they discuss more of the future existence of sufficient laws which as in the current instances Will only help in combating deepfake malice. It may encompass the nodes of how one may ban deepfakes, the legal words that could be said to apprehend, charge or punish a person, steps that are typical in identifying deepfake criminals and thus prevent them in the first instance.

Measures to be taken

To improve the legal and regulatory response to deepfakes in India, as well as globally, several measures can be considered:Based on the research, the following steps could be suggested to enhance the legal and regulatory response to deepfakes in India and globally.

1. Legal Frameworks: Pass a separate law dealing with manipulation relating to deep fake; to provide measures regarding the qualities of deep fake, the kinds and definition of deep fake, and civil and penal justice systems regarding the defense against deepfake offense and methods of recognizing and managing them.
2. Enhanced Enforcement: Improve the level of knowledge of the LEA and prosecutors about the deep fakes, sources of information on deep fakes and possibilities to get convictions for the deep fakes creators. The activities shall include co-ordination and organization of meetings and conferences with the heads of the top managers of government bodies and enterprises, head of NGO's to deliberate on matters of mutual concern, sharing of experiences as well as the progress achieved in enforcement.
3. Public Awareness and Education: Participate in public deliberate communication which will entail; carry out campaigns and education programs to explain to people the risks of falling a victim of deep fake; secondly raise the people's media literacy and competency level to between the original fakes and the more manufactured ones.
4. Technological Innovation: Suggest funding toward the sort of research that constitutes a credible threat to the deep fake problem through raising the technological advancement in identification and safe authentication of deep fakes by modifying the existing AI, Digital sciences, Cryptography. It is possible to co-ordinate the application of the given setting of the academic industry alongside governmental organisations in a manner that ensures the solutions offered are practical in the most profound sense of the word and in a scalable manner.
5. International Cooperation: Promote the depth of the international cooperation because deepfake manipulation is global in other to share information and thus improve the capacity and spread the knowledge of the best practice among countries or regions.
6. Stakeholder Engagement: Survey tech workers, journalists, nonprofits, and university students/staff on the best practices/strategies and caveats of deepfake manipulation and misinformation consumption/contribution.

Ethical Guidelines and best practices for responsible AI development

The following principles can guide developers, policymakers, and other stakeholders. As before, some of them would be useful to developers, policymakers and other before presenting them below.

1. Transparency and Accountability: Introduce the idea of presenting why and how the AI systems are in existence and their functions as well as the efficiency of the systems and the processes that went into the creation of such systems and or the approach towards addressing deep fake and its effects on people and

the community. In conclusion shift the responsibilities /blame towards the developers and or the users of such AI systems as the moral agents for the actions /outcomes of the system.

2. Fairness and Equity: Rights of a person in environment created using AI systems have to be secured to be predispersed specific data, as well as processes of setting up of the algorithms and fine-tuning of selected models to be free from discriminator. Pursue more manifest discrimination at the company level to work on the improvement of the status of diversity on the more organisational level of the development teams of Al systems.
3. Privacy and Data Protection: The final aim of the integrated schema presented in the work is associated with the topic of privacy; while in the case of data linked to the creation of AI systems; users' consent is significant in the data collection and usage relevant to the establishment of AI systems. Their measures and techniques include; High levels of securing the information and passwords, anonymized so that the developing data cannot be exposed to the wrong personnel.
4. Human-Centered Design: That AI has attributes of what have been described as ''AI done right''; which refers to the user need, the user preference and the user welfare. Stress that the technologies supplemented with AI contribute to people's competency and well-being, do not interfere with personality rights, and ardently follow the principles of dignitarianism.
5. Accountability Mechanisms: Support the strategies that can be maintained while assessing and interacting with the moral and social consequences of AI systems and the action that should be taken regarding to complaints, feedback or any concern which may be obtained from individuals or societies.
6. Continuous Learning and Improvement: Most of them are internment learning on the ethical allows, social, and governance issues, and AI introduction and laws. Involve the different factions because they also give feedback and/or participate in the numerous processes in the AI system.

Promoting media literacy and public awareness

With deepfakes informative manipulation takes place, and it harms the people and the society, therefore, becoming media literate is the best protection. The awareness or rather set of competencies used in the process of reception and decoding different types of media, different meanings, outcomes, and orientations and utilizing it for the decision-making process in the world dominated by fake news.

1. Detection and Identification: When media literate, the people will be in a good stander to perceive the deep fake manipulation, and distinguish between reality and fake reality. Citizens must therefore ensure that they are always able to comprehend how deep fakes are made as well as the indications that one should lookout for in order to be effectively critical users of the materials which are produced and shared within the social media.
2. Prevention of Spread: A safer form of naive sharing of deepfake content is guaranteeing that there is created media literacy that will stop people from sharing such information. Therefore, it is even more critical to disseminate information about such abilities and the need to check the sources as well as the methodology for evaluation of fact-checking and critical thinking whenever concerning media content in order to reduce the effects of deepfake manipulation.
3. Protection of Privacy and Security: Some of the actions that can be done under media literacy campaign also have potentiality to raise the consciousness regarding the growing possibility of deepfake manipulation for identity theft, impersonation, and to harm someone's reputation. Hence, privacy and security are inculcated in media literacy as it is good to know the risks that come with sharing one's details or interacting with a poss/ect feared substance/deal.
4. Empowerment and Resilience: The media literacy assists the people to know that in the society today whenever something is done by a person that was his/her own decision to do that and should suffer the consequences. Media literacy benefits a person directly to be a wise participant of the social media by presenting certain analytical instruments which would assist in discarding such undesirable values as doubtful information, fake or even deliberately misleading information.
5. Civic Engagement and Democratic Participation: In the given context media literacy is resembled with a major concept of commitment towards democracy as it will ensure that people have acquired requisite knowledge and dexterity in how they can acquire information, analyze it and utilize it taking into account a fact that they have become active voters. Hence, raising its media sensitivity, people educating the

public, people work on fortifying the principles of the democracy, and the other structures associated with it for the required clients, making it sufficiently powerful.

6. Collaborative Efforts: A change that goes in a positive direction and enhances the media competence to increase the general public is hardly constructed through the development of government departments, schools and universities, media, civil society organizations or technology company alone. Media literacies cop as with other stakeholders, when appended, one can also extend and pass on the message or spread awareness and distribution of materials in the fastest way possible.

Conclusion

Hence, the present research paper seeks to stress the need to call for sufficient measures for addressing the manipulations of the given deepfake technology as its concerns rise. As it can be seen from the above examples deep fake technology transformed into the efficient instrument for manipulation as well as for spreading fake news and destabilizing the society, media and democratic principles. This paper has discussed that the ramifications of deepfake manipulation are not just in the technicalities of an arrangement but in ethical, legal and social implications which cannot say no to.

This paper has analysed the current legal and regulatory frameworks pertain to deep fake manipulations and the realization that none of the technological interventions are sufficient in mitigating the risks of deep fake. Nevertheless, there are some rules such as the Information Technology Rules 2021 and Intermediary Guidelines that try to fight cybercrimes and modulate social media but those laws do not pay sufficient regard to the issues escalated by deep fake technology. Likewise, the measures derived from the approaches utilizing technologies for detection and authentication of deepfakes still has issues with the question of efficiency in the augmenting significance of technologies in the falsification proceedings and generation of fake news.

Thus, this paper postulates that increasing the levels of media literacy and awareness of this type of manipulation represents the two primary basic approaches to tackling deepfake manipulation. Thus, media literacy education as the whole-bearer approach to develop individuals' competencies necessary for the media interaction is widely recognized as essential for the nonescalation of deepfakes and protection of society from manipulations, fake, and misleading information. Moreover, to initiate and develop the provision of effective media literacy programmes and the public awareness campaigns; the solutions with the support of the governmental institutions, schools & universities, media houses and civil society companies together with other technology-based companies are of pivotal importance.

Notably, creating awareness of deepfake manipulation as a new type of manipulation activity, this paper calls for the formulation of legal and institutional means for combating deepfake manipulation in particular. Regarding elements necessary for the classification of deepfakes, rules of liability and responsibility for them, as well as mechanisms for the detection, prevention, and counteraction to the Deepfake-related crimes. Furthermore, there is a need for additional technological development and cooperation for improving the methods to combat deepfake and regarding the interstate character of deepfake manipulation.

Thus, it is safe to conclude that the risks and consequences' reduction links directly to the utilization of deepfake technology is achievable in cooperation of those actors who act in different fields and industries. Thus, by supporting media literacy, enhancing the technologies' threat recognition, promoting better legal and regulatory conditions, and encouraging ethical work, it is possible to create a safer world with ethical work, protection, and laws based on truthful technologies. Thus, society should ensure that only collective action results in emergent technologies providing benefits to people's welfare and overall quality of life in the information age.

References

1. Bradshaw, S., & Howard, P. N. (2019). The global disinformation order: 2019 global inventory of organized social media manipulation. Computational Propaganda Research Project, 26.
2. Citron, D. K. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. Foreign Aff., 98, 147.
3. Citron, D. K., & Chesney, R. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. Calif. L. Rev., 107, 1753.
4. Farid, H. (2020). Deepfake detection: Current challenges and future directions. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 16(4s), 1-19

5. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).
6. Hassner, T., Harel, S., Paz, E., & Enbar, R. (2019). The unreasonable effectiveness of deep features as a perceptual metric. In IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 2112-2120).
7. Hosseini, H., Rahmati, A., & Rahmati, H. (2020). Deepfake Detection Using Attention Mechanism and Capsule Network. arXiv preprint arXiv:2003.12218.
8. Li, Y., Chang, M. C., Kuo, C. C. J., & Chang, S. F. (2019). In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking. In European Conference on Computer Vision (ECCV) (pp. 736-753). Springer, Cham.
9. Marra, F., Gragnaniello, D., & Verdoliva, L. (2019). A large-scale analysis of the impact of face recognition-based deepfake attacks. In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS) (pp. 1-6). IEEE.
10. Nguyen, T., Kim, J., & Lee, S. (2020). Multi-modal Fusion for Detecting DeepFake Videos. arXiv preprint arXiv:2003.06713.
11. Roberts, T. L. (2020). Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics. Foreign Aff., 98, 147.
12. Rosenblatt, J. (2018). Law and the future of truth. Harv. JL & Tech., 31, 423.
13. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1-14).
14. Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2Face: Real-time face capture and reenactment of RGB videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2387-2395).
15. Wang, Y., Ma, Z., & Jiang, C. (2019). DeepFake detection based on frame aggregation convolutional neural network. In 2019 3rd International Conference on Image, Vision and Computing (ICIVC) (pp. 903-907). IEEE.
16. Zhang, Y., Ye, X., Xu, W., & Li, S. (2020). U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation. IEEE Transactions on Image Processing, 29, 5893-5906.