Examining the Societal Impact and Legislative Requirements of Deepfake Technology: A Comprehensive Study

Sami Alanazi[1,]*, Seemal Asif[1], and Irene Moulitsas[2]

[1] Centre of Robotics and Assembly, Cranfield University, UK
[2] Centre for Computational Engineering Sciences, UK
Email: sami.alanazi@cranfield.ac.uk (S.A.); s.asif@cranfield.ac.uk (S.A.); i.moulitsas @cranfield.ac.uk (I.M.)
*Corresponding author

Manuscript received August 30, 2023; revised October 15, 2023; accepted December 21, 2023; published March 29, 2024.

***Abstract*—Deepfakes, highly realistic fabricated videos, images, or audios created using artificial intelligence algorithms, have gained widespread attention and raised concerns about their social impact and ethical implications. While initially seen as a source of entertainment and utilized in commercial applications, deepfake technology has increasingly been misused for creating adult content, blackmailing individuals, spreading misinformation, and manipulating memories. The negative consequences of deepfakes extend beyond individuals, impacting society as a whole, particularly during sensitive times like elections, where trust can be undermined. The paper looks at the social implications and legislative considerations for deepfakes content, with the goal of cultivating a thorough understanding of their impact on society. By highlighting the importance of enacting laws and regulations, the paper emphasizes the pressing need to control their widespread dissemination. Deepfakes have broad societal implications, especially during critical events like elections, eroding trust. This survey delves into deepfake complexities, aiming to foster understanding and emphasizing the urgency of regulation. The paper also discusses the positive outcomes of deepfake technology for intellectual property protection, highlighting the FORGE system developed to trick attackers who steal company documents. However, it emphasizes the risks posed by easily accessible websites that facilitate the creation of fake identification documents, increasing the likelihood of identity theft, criminal activities, and fraudulent transactions. Implementing restrictions on the use of deepfake technology involving children is also crucial to prevent harm and manipulation targeting minors.**

***Keywords*—deepfakes, social implications, legislation, regulation, authenticity, misinformation, ethical implications**

I. Introduction

Deepfakes, a product of artificial intelligence algorithms, are highly realistic fabricated videos, images or audios. They exploit Artificial Intelligence technologies to convincingly depict individuals engaging in actions or making statements they never actually performed. Furthermore, the accessibility of user-friendly tools, like FaceSwap and Face Swapper [1], allows non-experts to produce deceptively authentic-looking deepfakes. While the technology can have significant positive potential across various sectors, like other new technologies, in the hands of the wrong people, deepfakes can be utilized in negative ways as well.

The emergence of deepfake technology and its ease of content creation have raised concerns about its potential social impact and the need for legislation to address its ethical and legal implications. This paper aims to explore the social implications of deepfakes and the importance of legislation in regulating their use.

Deepfakes have initially been seen as a source of entertainment, with users enjoying sharing animated depictions of historical figures and loved ones using deepfake applications like Deep Nostalgia [2]. Moreover, deepfake technology has found applications in commercial and advertising purposes, such as creating fake versions of well-known works of art and editing scenes in movies, saving time and costs for film production companies.

However, the negative social consequences of deepfakes have become apparent. The technology has been misused for creating adult content, blackmailing individuals, and spreading misinformation. Deepfakes have made it difficult for the public to trust news and visual evidence, leading to uncertainty and the potential to mislead people. Interpersonally, deepfake videos have the power to modify and implant false memories, altering people's attitudes toward each other without valid reasons. Therefore, Recognizing the nature of deepfakes is crucial for effectively tackling the ethical and legal challenges they pose in the modern digital environment [3]. As depicted in Fig. 1, the lifecycle of deepfake content includes its creation, propagation within social media platforms, the detection procedure, and the implementation of mitigation actions where policymaking comes into play. It's crucial to emphasize the importance of establishing a feedback mechanism between detection and mitigation to facilitate the monitoring of deepfake spread.

The social impact of deepfakes extends beyond individuals, affecting society as a whole. The spread of deepfake content during sensitive times, like elections, for example, can undermine trust and have far-reaching and serious implications. If deepfakes continue to proliferate, future generations may find it challenging to rely on any internet-based information, leading to a decline in social trust and impact on various industries.

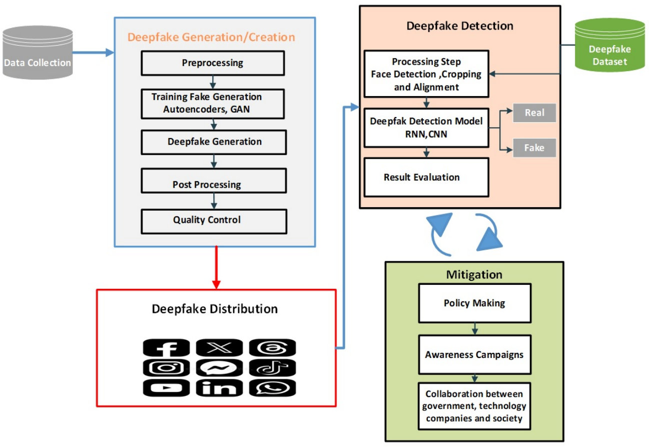
Data Collection
Deepfake Generation/Creation
Preprocessing
Training Fake Generation Autoencoders, GAN
Deepfake Generation
Post Processing
Quality Control
Deepfake Distribution
Deepfake Detection
Processing Step Face Detection ,Cropping and Alignment
Deepfak Detection Model RNN,CNN
Real
Fake
Result Evaluation
Deepfake Dataset
Mitigation
Policy Making
Awareness Campaigns
Collaboration between government, technology companies and society

Fig. 1. Deepfake content workflow.

doi: 10.18178/ijssh.2024.14.2.1194

II. Social Impact of Deepfake

Initially, deepfake videos were an object of fun and it was assumed that both the creator and the subject of the video would derive pleasure from the final product. In addition, social media users have enjoyed sharing animated depictions of historical personalities and their loved ones who have passed away and other objects that evoke deep nostalgia using a deep fake application called Deep Nostalgia [4]. Furthermore, deepfake technology is utilized for commercial and advertising purposes as well. Such technology is being used to create fake versions of well-known works of art, such as movies of iconic Monalisa artwork using the original picture. Moreover, film production companies are increasingly adopting deepfake technology to edit scenes in their movies, thus saving time and reducing costs associated with reshooting [5].

The rapid utilization of the technology led to the creation of adult content and materials with the potential to be exploited for blackmail purposes. Furthermore, one of the major negative social consequences of deepfake video generation technology is that the public finds it difficult to trust the news [6]. This, in response, prompts users of social media platforms to exercise caution and verify information from multiple sources as a means of ensuring the credibility of the news. Not only that, such videos and images also make it easy to mislead people. There is uncertainty among the masses when it comes to believing in videos and images as proof of something. There are also many interpersonal effects of deepfake videos such as their capacity to modify our memories and even implant false memories in people's minds. This can alter a person's attitude towards another without any actual reason [6].

Social media enabled people to connect with friends and family across borders and to save their memories in cameras to cherish in the future. However, deepfake content creation is affecting the confidence of users in technology. On one hand, influential people can be blackmailed if their enemies create their fake content and on the other hand, such content can be created for non-political but pornographic purposes. Deepfake content, whether driven by political motives or intended for explicit purposes, has emerged as a formidable digital menace, comparable in its disconcerting impact to the peril of identity theft.

During specific periods of heightened sensitivity or significant events, the likelihood of the dissemination of deepfake content increases substantially. For instance, during elections, it is highly likely that the opponents will share deepfake videos to bring each other down [7] At such times, the users should be alert of sharing any dubious content. To discourage the creators of deepfake, individuals should refrain from forwarding any content shared by unverified accounts unless it is later acknowledged by a high-credibility organization. This can help prevent the spread of deepfake content and limit its impact. Hence, it can be said that the effect of deepfake on society can be immense, but the role of society is also high in discouraging such content. If deepfakes are continuously shared on the internet, the next generation of users will find it difficult to rely on internet-based information. Therefore, Society needs to recognize that sharing deepfakes can help deepfake creators fulfill their desires to spread misinformation. The social impact of deepfakes is devastating because in the future, people may be hesitant to share their personal content on social media due to trust issues, which could have a widespread impact on industries. Establishing and enforcing laws can play a crucial role in preserving social trust regarding the dissemination of visual content online. Machine learning techniques, including deep learning and Generative Adversarial Networks (GANs) can be used to generate realistic fake images, while NLP techniques can be used to generate convincing text. By combining these techniques, it is possible to create highly realistic and convincing deepfake documents which are difficult to differentiate from authentic ones. There are many techniques that can be used for linguistic characterization and detection of false news, particularly over the internet. Validation of the integrity of information over social media poses a critical and formidable challenge due to the inexpensive and rapid creation, publication, and dissemination of content. The effectiveness of Natural Language Processing (NLP) in addressing this challenge is discussed in reference [8] NLP algorithms help in automatic identification, where automatic proofing of logical statements is carried out using a stylistic-computational approach. NLP solves the scalability problem associated with the manual process of data verification. Through automatic fake news detection, fake news propagation can be significantly controlled. NLP's effectiveness lies in its architecture, which has two parts: Natural Language Understanding and Natural Language Generation [9]. NLP understands as well as generates the text. Just like humans, NLP interprets the lexical meaning of individual words and represents semantics by analyzing the structure of words to verify the authenticity of the content. Creating machine learning techniques that depend on Natural Language Processing (NLP) for fake news detection requires large volumes of data to train the models and produce accurate results. However, this can be a challenging task due to the lack of sufficient forged document datasets [10]. The criminality associated with forging official documents and the imperative to safeguard individuals' privacy, given that fraudulent documents invariably contain personal information, are elucidated as the underlying reasons in the scholarly discourse outlined in this paper. According to TrustID reports, fake ID crimes have increased by 6% in 2020 [11]. Fig. 2 illustrates the common types of ID document forgery crimes.

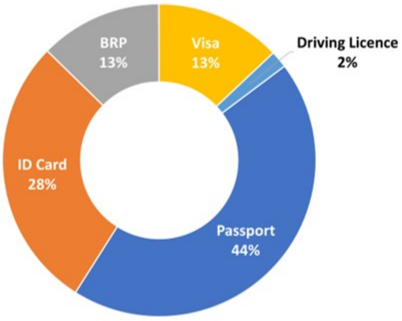

Fig. 2. Types of fake identifications [11].

Regrettably, certain online platforms, such as diplomamarket.com, facilitate the provision of counterfeit diplomas from esteemed universities in the United Kingdom or the United States, thereby perpetuating the illicit practice.

Likewise, within the realm of financial services, the utilization of counterfeit bank statements presents a means to acquire loans or credit for which an individual may not meet the necessary criteria. Moreover, online resources such as fakedocuments.com offer accessible tools for producing fraudulent bank statement documents at nominal costs.

The emergence of deepfake documents has yielded significant positive outcomes in the domain of intellectual property protection [12]. This is because generating fake documents presents potential advantages for companies that struggle with intellectual property theft. As an illustration, in 2020, pharmaceutical companies like Pfizer experienced cyber-attacks perpetrated by hackers who illicitly obtained COVID-19 vaccine data [13]. The leak of the vaccine secret could push other companies to produce the COVID-19 vaccine first, which could cost Pfizer millions of pounds. The FORGE system (Fake Online Repository Engine) has been developed [12], wherein an automated process generates forged samples of authentic documents present within the company's database. The underlying concept of the system revolves around the intention to confound potential attackers subsequent to the theft of documents, as their ability to discern between counterfeit and authentic documents is significantly impeded. This deliberate obfuscation introduces an element of temporal delay that holds intrinsic value for the victim.

The proliferation of easily accessible websites, exemplified by fakedocuments.com and diplomamarket.com, engenders a multitude of risks that extend to both individuals and society at large. For instance, these websites increase the likelihood of identity theft by enabling people to create fake passports, IDs, and other important documents. The use of fake documents opens doors to various criminal activities such as smuggling, drug trafficking, and terrorism. Moreover, the availability of counterfeit financial documents for a small fee contributes to an increase in fraudulent transactions. To effectively tackle these risks, it is crucial to swiftly take down such websites and impose appropriate penalties on individuals involved in offering fraudulent document services.

Furthermore, it is crucial to implement restrictions on the use of deepfake technology involving children. The misuse of deepfake technology, particularly for bullying purposes, can substantially harm minors. The tragic case of Molly Russell, a 14-year-old girl who tragically took her own life due to self-harm, exemplifies the detrimental effects of viewing self-harm content online [14]. This serves as a powerful depiction of the serious outcomes that can emerge from the harmful manipulation of online content concerning underage individuals.

Ideally, social media platforms should establish a robust mechanism that thoroughly scans images and videos prior to their publication, thereby preemptively curbing the dissemination of deepfake content. StopNCII.org has released a freely available scanning tool, which provides support to victims of non-consensual intimate image abuse globally. Victims can produce a hashed fingerprint of the intimate image and send it to participating platforms such as Facebook, Instagram, and TikTok, which will prevent the sharing of any matching images online [15]. Additionally, to effectively address the challenge of deepfake proliferation, it is recommended to establish dedicated resources for monitoring and regulating the spread of such content. For example, Twitter has announced its plans to increase the employment of human moderators and fact-checkers for the purpose of reviewing Twitter posts [16]. These resources can comprise a combination of human moderators and advanced technological solutions, such as AI-based detection algorithms, to identify fake content. Furthermore, forging partnerships with domain experts specializing in deepfake detection and verification can yield invaluable perspectives and foster the advancement of comprehensive monitoring approaches. By allocating resources towards these endeavors, the proactive identification and mitigation of deepfake content can be bolstered, thereby diminishing its potential adverse impact.

III. Legal Consideration of Deepfake Technology

Within the ever-evolving landscape of technology, the emergence of novel crime methodologies is equally relentless. Conventional legal frameworks frequently prove inadequate in effectively addressing these emerging forms of criminal activity. Consequently, there exists a pressing necessity to establish updated and sophisticated legislation that not only comprehends the complexities of cybercrimes but also implements suitable penalties to deter and penalize offenders. There is a huge potential for deep fake technology to cause nationwide distress and disaster, as the 2018 genocide of Rohingya community in Myanmar is believed to be triggered by posts created through deepfake technology [17]. Similarly, during the 2018 elections in Kenya, it was suspected that deepfake content of a sick presidential candidate was shared to encourage people to stand for him, when it was doubted that the candidate was either bed-ridden or already dead [17]. While certain actions, such as murder or theft, are inherently criminal and with direct impact to the lives of others, there are also actions that may not necessarily be classed as crimes but can lead to criminal activity and harm. For instance, driving while in a state of intoxication is prohibited under the law because a drunk driver can cause accidents on the road that affect their and other people's lives. Similarly, the deepfake content can affect the reputation of a person which is sometimes irreparable. The proliferation of deepfakes not only spreads disinformation on matters of public interest, but such content can cripple public discourse and undermine democracy. Public hysteria caused by fake videos can lead people to distrust their government and law enforcement agencies [18].

Developed countries have recognized the gravity of risks posed by deepfake content and its profound impact on social and legal realms, leading to the adoption of both legal and non-legal measures aimed at mitigating such activities [19]. In light of the growing concern among courts regarding the increasing use of deepfake content as evidence, the necessity for countermeasures to evaluate the credibility of such proofs becomes evident. While the European Union currently lacks specific legislation focusing on deepfake regulation, ongoing developments include the introduction of an AI regulation mandating transparency of AI systems [19]. Moreover, the proposition of legislation aimed at addressing the utilization of deepfakes in identity theft, age fraud, illegal immigration, and espionage, as highlighted by the argument put forth by

Ref. [20], is essential. The repercussions of these activities extend beyond mere individuals, encompassing substantial implications for businesses, social trust, and political processes. Presently, the General Data Protection Regulation (GDPR) closely aligns with the realm of deepfakes, offering a foundation for regulatory oversight. Currently, the GDPR bears relevance to deepfakes as it pertains to their legal framework. However, it is important to note that the GDPR does not encompass the dissemination of fabricated information through deepfakes if an individual can be readily identified within the content [19]. As a result, the applicability of the GDPR extends to both the underlying data employed in the creation of deepfakes and the deepfakes themselves.

In the context of data protection, the European Data Protection Board (EDPB) assumes a critical role in both establishing and enforcing relevant laws. One of its key functions involves providing guidance on the application of the GDPR to organizations operating outside the European Union (EU). Particularly, these guidelines prove invaluable for Middle East organizations, aiding them in assessing their adherence to GDPR regulations. As demonstrated by the findings presented in Ref. [21], which highlight the potential of implementing modern technologies and processes to fulfill the statutory role of maintaining data integrity. To illustrate the implementation of modern data protection measures, a notable example can be observed in the Kingdom of Saudi Arabia, which introduced its Personal Data Protection System in September 2021 [22]. This initiative exemplifies the commitment of countries to safeguard data integrity by adopting comprehensive frameworks in line with contemporary technological advancements, as elucidated in Ref. [23]. In affluent nations such as the United Kingdom, there exists a robust dedication and abundant resources to establish laws that ensure accountability for deepfake content creators. However, the effectiveness of such legislation relies on jurists possessing a comprehensive understanding of deepfake technology. Recognizing the urgency of the matter, the UK government acknowledges the necessity of enacting laws that specifically target various forms of deepfakes, such as face reenactment, face generation, and speech synthesis [17]. As deepfake technology advances, the detection and penalization of such content pose increasing challenges. Hence, legislation is being developed to discourage the creation of deepfake content for political and social manipulation, recognizing the potential harm it can cause to individuals, organizations, and political entities [17].

New laws are frequently constructed based on preexisting legal structures, and one of the primary justifications for imposing legal consequences on deepfake content is its significant contribution to the dissemination of misinformation [17]. In line with this objective, the government is collaborating with various stakeholders such as think tanks, media experts, parliamentary select committees, and technical professionals to develop legal measures that effectively deter the misuse of deepfake technology. Notably, the UK government is actively working towards criminalizing the sharing of non-consensual pornographic deepfakes [24], and mechanisms are being devised to facilitate complainants and ensure that perpetrators face appropriate penalties, including potential imprisonment [23]. Furthermore, the EU's AI Act aims to guarantee user transparency and awareness when engaging with AI systems including those capable of generating or manipulating media content like deepfakes [25]. The specific measures will depend on the risk level of the AI system, with lower-risk systems subject to minimal transparency requirements. The goal is to protect users and enable them to make well-informed decisions when utilizing AI applications.

In the United States, the development of laws concerning deepfake content is guided by the principles of freedom and privacy protection for individuals. The citizens have the right to privacy, but deepfake content can jeopardize the personal lives and privacy of common people, celebrities, or politicians by faking their videos to create an immoral image in society. On one hand, the First Amendment in the US constitution protects freedom of speech, but on the other hand, the privacy of people also prevails which means that no citizen has the right to violate the privacy of individuals or to invade their legal rights [18]. Two Tort laws in the US that can be weakly linked to the prevention of deepfake are Prosser Torts namely "intrusion upon seclusion" and "publicity given to private life". Two other Tort Laws that are relatively more applicable to the deepfake technology are "appropriation of name or likeness" and "false light publicity" on victims' faces, which affects the privacy of victims. However, the limitation of Tort Laws is that when bringing deepfake case into a court of law, the victims need to prove that either harm was caused to them through deepfake or there was potential harm that could be caused because of the public spread of deepfake regarding them. Consequently, while the First Amendment safeguards freedom of speech, it is essential to balance it with the preservation of personal privacy and the prevention of harm caused by deepfake content, as it can potentially tarnish the reputations and violate the privacy rights of ordinary citizens, public figures, and politicians [18].

Deepfake content creation can fall within the purview of Tort Law, particularly in cases where face-swapped videos have the potential to cause severe emotional distress, leading to claims of Intentional Infliction of Emotional Distress (IIED) [26]. If the defamation claim is accompanied by an IIED claim, then the deepfake content creator can be penalized. It is important to note that the defense of "parody content" cannot be invoked in such cases, as parody should not be misleading or detrimental to the mental well-being of the subject.

However, the use of Tort Law as a means of protection against deepfake content is not without criticism. As previously mentioned, it is vulnerable to potential challenges based on the protections afforded by the First Amendment, which safeguards the freedom of speech. The First Amendment states that "Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof, or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances" [18].

In the process of creating laws for deepfake content, the existing copyright laws are also considered. However, the copyright law also lacks the ability to offer a remedy for victims of deepfake content essentially because deepfake creators argue that the individuals own the right to their original photos and videos, not to the parody or edited forms of original content. According to Article I, Section 8, Clause 8 of the Constitution, there is a limitation to "authors and inventors the exclusive right to their respective writings and discoveries". Hence, Shannon [18] argues that if the plaintiff captured photos or video that are then used to create deepfake, then the plaintiff owns the copyright of the original media, not the deepfake created using original content. The transformative work or parody is covered under Section 107 of the Copyright Act which means the plaintiff will lose the case. The problem with Tort law and Copyright laws, discussed so far, is that these consider deepfake content creation as standalone action irrespective of the consequences caused by those actions. As with the example of alcohol use while driving discussed above, there is a need to make laws regarding deepfake content creation that focus on the drastic impact of such content, not on how much content was produced and who owns the right to that content.

The dissemination of deepfake content can be effectively addressed by leveraging existing laws such as the Revenge Porn and Non-consensual Pornography Law. In many instances, individuals resort to spreading pornographic videos or images of others, often with malicious intent to tarnish their reputation. Such acts, commonly known as "Revenge porn" or non-consensual pornography, are prohibited under specific revenge porn laws that forbid the distribution of sexually explicit material without the consent of the individuals involved. While this offense is typically committed by former partners or spouses, the law extends its scope to encompass the sharing of any explicit content without consent. Presently, the law applies to the sharing of original content, making it reasonable and logical to extend its application to the dissemination of fabricated or fake content by adversaries without the consent of the individuals depicted.

At first, deepfake content received some protection under claims of being created for parody or satirical intentions. However, US courts have begun recognizing the potential for deepfake content to facilitate various forms of criminal activity. Consequently, several states have taken the initiative to establish specific laws addressing the issue of deepfakes. An example of this can be seen in Texas, where in 2019, Section 255.004 of the Election Code was amended to include provisions regulating the creation and dissemination of deepfake videos during state elections [27]. Such an offense can result in a punishment of up to one year of imprisonment in the county jail and a fine amounting to $4,000 [27]. This legal provision highlights the seriousness with which the state addresses the potential threats posed by deepfake content in the context of elections.

Similarly, in California, there are specific laws in place to address the dissemination of deepfake content during elections. These laws are outlined in the Deceptive Audio or Visual Media Act, which is codified in Section 20010 of the California Election Code. This legislation specifically prohibits the sharing or distribution of images, videos, and audio that may appear genuine to the public but are, in fact, fabricated and not original. However, it's important to note that the legislation does not restrict the production and dissemination of videos, audio, or content that constitute satire or parodies, as long as they do not mislead the audience. Additionally, California Assembly Bill 1280 has been introduced, which specifically prohibits the distribution of deepfakes during Elections, further emphasizing the state's commitment to combatting the negative effects of deepfake technology [28].

Legal frameworks are increasingly considering the establishment of legal penalties for the forgery of visual data. To effectively address this issue, legislative bodies are adopting a proactive approach by advocating for the development and utilization of forgery detection tools specifically designed to identify deepfake content. Similar to the use of diagnostic tests in the medical field, the detection of deepfakes necessitates the reliance on specialized tests and tools for visual and audio content. Given the prevalence of digital evidence in the modern era, it is crucial to determine proper protocols for handling and utilizing such evidence. In Israeli law, particular emphasis is placed on ensuring the admissibility and evidentiary value of deepfake proof, underscoring the significance attributed to addressing the challenges posed by deepfake technology in the legal context [29]. Israeli law is based on Anglo-American evidence law which is highly strict in terms of formalistic rules of admissibility of evidence. However, Israeli law is slowly becoming more flexible resulting in rules related to the probative weight of evidence that shapes the rules of admissibility.

The relaxation of admissibility rules in Israeli and international laws coincides with the departure from the best evidence practice [29]. New flexible laws prioritize the presentation of primary evidence, which is considered the most reliable, over the need for extensive quantities of evidence. Loopholes in the best evidence doctrine can impact court rulings, making it crucial to provide both original and tampered data to demonstrate how the authentic content was manipulated using technological tools. It is not baseless to assume that the law requires providing a suitable reason for not presenting the original document. Failure to provide the original file in court can make it challenging to establish where changes have been made through deepfake technology [29]. Conviction on the basis of proof can have long-term implications for individuals and organizations, hence it is important that strong evidence is presented in the court of law to convict the criminals. In this regard, deepfake detection tools can provide the best evidence when real and fake content is compared, and forgery is detected. AI systems with human-level intelligence have the potential to present significant dangers to society and humanity. In response to these concerns, an open letter titled "Pause Giant AI Experiments" has been issued, calling for a temporary halting of training AI systems that exceed the capabilities of GPT-4 for a minimum duration of six months [30]. The letter emphasizes the importance of collaborative efforts between AI labs and independent experts to establish and enforce shared safety protocols for the design and development of advanced AI. These protocols should undergo rigorous auditing and oversight by external experts to ensure the safety and dependability of AI systems. The letter highlights the risks associated with AI systems capable of generating deepfake content for political manipulation. Furthermore, the letter puts several recommendations for policymakers to mitigate the risks associated with advanced AI systems, including regulations on access to computational power and the establishment of standards for identifying and managing AI-generated content and recommendations. Signed by influential figures from academia, industry and other domains.

Taking all of these factors into consideration, it becomes clear that existing laws often fail to address the potential risks associated with deepfake technology. It is crucial for policymakers to take swift action in developing comprehensive legislation that addresses the creation of deepfake content and imposes penalties on offenders, not only based on their actions but also considering the harm inflicted upon victims. A notable example is the tragic case of Molly Russell, where the UK government announced its intention to criminalize online self-harm content in late 2022 [6]. However, it is concerning that it took the government five years following the teenager’s death, which likely had adverse effects on her family and friends and increased the possibility of similar cases occurring among minors during the gap between the incident and the government’s action.

IV. Conclusion

The deepfake technology presents both benefits and risks to society. While it has the potential for entertainment and creativity, its misuse can lead to significant harm, including political propaganda, misinformation campaigns, harassment, and erosion of trust in the media.

Addressing the social implications of deepfakes requires a multifaceted and inextricably intertwined approach involving legislation, regulation, and awareness. Existing laws often fall short in dealing with these emerging forms of criminal activity. Therefore, there should be legislation to deal with people and software who make and/or publicly share deepfake content for the purposes of defamation or blackmailing. Doing so will reinforce the public perception that creating and sharing deepfake content of other people is both socially unacceptable and legally impermissible. Countries like the United Kingdom and the United States have recognized the need for specific laws targeting different forms of deepfakes and have made efforts to penalize the creation and dissemination of deepfake content for political and social misuse.

The legal considerations surrounding deepfake technology demand urgent attention and robust legislation. The existing legal frameworks are often inadequate in addressing the emerging forms of cybercrime associated with deepfakes. The potential for deepfake technology to cause widespread distress and disasters, as seen in past incidents, highlights the need for updated and sophisticated laws that comprehend the complexities of deepfake crimes and impose suitable penalties on offenders. Developed countries have recognized the gravity of these risks and have taken steps to mitigate deepfake activities through legal and non-legal measures. However, challenges remain in evaluating the credibility of deepfake evidence and ensuring accountability for deepfake content creators. Existing laws, such as Tort Law and Copyright laws, have limitations in addressing the full scope and consequences of deepfake content. Therefore, it is crucial for lawmakers to establish comprehensive legislation that focuses on the potential harm caused by deepfake content, balancing freedom of speech with the preservation of personal privacy and the prevention of harm. The use of existing laws, such as revenge porn laws, and the development of specific laws targeting deepfakes in elections demonstrate the recognition of deepfake threats and the importance of legal measures. Furthermore, the relaxation of admissibility rules and the development of deepfake detection tools can strengthen the evidentiary value of deepfake proof in court proceedings. In light of the growing risks associated with deepfake technology, it is imperative for policymakers to act swiftly and decisively in establishing comprehensive legislation that effectively addresses the creation and dissemination of deepfake content, imposing appropriate penalties on offenders, and safeguarding individuals, organizations, and democratic processes from the harmful impacts of deepfakes.

In addition to legislation, non-legal measures such as technological solutions such as social media platform scanning tools and moderators resources can contribute to mitigating the impact of deepfake content. These tools aim to prevent the spread of deepfakes, particularly on social media platforms. However, addressing the challenges of deepfakes requires a comprehensive understanding of the technology and continuous adaptation of laws and regulations.

Furthermore, awareness campaigns should be run by media and government agencies whereby the consequences of making and sharing such content should be clearly communicated so that people abstain from generation deepfake content with the intention of blackmailing, bullying or embarrassing people. Like many other domains of technology, this issue can only be effectively addressed through the collaboration of government and society. A cooperative approach involving researchers, policymakers, and technology platforms will play a crucial role in tackling the evolving landscape of deepfake threats and ensuring a more secure digital environment.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Sami Alanazi’s research analyzed the societal impact of deepfake technology across ethics, law, and technology, and proposed recommendations for policymakers to mitigate risks. This work was done under Seemal Asif’s guidance and with contributions to ideas and writing from Irene Moulitsas. All authors had approved the final version.

References

[1] C. Wilpert. (2022). 7 best deepfake software apps of 2022 (50 Tools Reviewed). [Online]. Available: https://contentmavericks.com/best-deepfake-software/

[2] M. Zhang. (2021). ‘Deep Nostalgia’ brings people in old photos back to life. [Online]. Available: https://petapixel.com/2021/03/01/deep-nostalgia-brings-people-in-old-photos-back-to-life-with-movement/

[3] S. Alanazi and S. Asif, “Understanding deepfakes: A comprehensive analysis of creation, generation, and detection,” 2023.

[4] N. El-Hadi. (2021). Faces of histories the ‘Deep Nostalgia’ face animator conflates the desire to honor the past with an impulse to appropriate it. [Online]. Available: https://reallifemag.com/faces-of-histories/

[5] B. U. Mahmud and A. Sharmin, "Deep insights of deepfake technology: A review," 2020.

[6] J. T. Hancock and J. N. Bailenson, "The social impact of deepfakes," *Cyberpsychol. Behav. Soc. Netw*., vol. 24, no. 3, pp. 149–152, 2021.

[7] A. Jaiman, "Debating the ethics of deepfakes," *Tackling Insurgent Ideologies in a Pandemic World, ORF and Global Policy Journal, New Delhi*, pp. 75–79, 2020.

[8] N. P. D. Oliveira, P. S. Pisa, M. A. Lopez, *et al.*, "Identifying fake news on social networks based on natural language processing: Trends and challenges," *Information (Switzerland)*, vol. 12, no. 1, 2021.

[9] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends and challenges," *Multimed. Tools Appl.*, vol. 82, no. 3, pp. 3713–3744, 2023. https://doi.org/10.1007/s11042-022-13428-4

[10] N. Sidere, F. Cruz, M. Coustaty, and J. M. Ogier, "A dataset for forgery detection and spotting in document images," in *Proc. 2017 Seventh International Conference on Emerging Security Technologies (EST)*, 2017.

[11] Inscribe. (2021). Identifying fake documents: A complete overview. [Online]. Available: https://www.inscribe.ai/document-processing/fake-documents

[12] Q. Han, C. Molinaro, A. Picariello, *et al.*, "Generating fake documents using probabilistic logic graphs," *IEEE Trans. Dependable Secure Comput.*, vol. 19, no. 4, pp. 2428–2441, 2022. https://doi.org/10.1109/TDSC.2021.3058994

[13] S. Porter, "Pfizer COVID-19 vaccine data leaked by hackers," *HealthCareitNews*., 2021.

[14] North London Coroner's Service, *Regulation 28 Report to Prevent Future Deaths*, London, 2022.

[15] Swgfl, "New industry partners join StopNCII.org to prevent the sharing of non-consensual intimate images online," 2022.

[16] K. Deka. (2023). EU tells Elon Musk to hire more staff to moderate Twitter – FT. [Online]. Available: https://www.reuters.com/technology/eu-tells-elon-musk-hire-more-staff-moderate-twitter-ft-2023-03-07/

[17] GOV.UK. (2019). Snapshot paper – Deepfakes and audiovisual disinformation. [Online]. Available: https://www.gov.uk/government/publications/cdei-publishes-its-first-series-of-three-snapshot-papers-ethical-issues-in-ai/snapshot-paper-deepfakes-and-audiovisual-disinformation

[18] R. Shannon, "The deepfake dilemma: Reconciling privacy and first amendment protections," 2020.

[19] B. V. D. Sloot and Y. Wagensveld, "Deepfakes: Regulatory challenges for the synthetic society," *Computer Law and Security Review*, vol. 46, 2022. https://doi.org/10.1016/j.clsr.2022.105716

[20] R. Roy, I. Joshi, A. Das, A. Dantcheva, and B. Pilani, "3D CNN architectures and attention mechanisms for deepfake detection," 2022.

[21] PWC Middle East, "EU general data protection regulation applicability to the Middle East," 2021.

[22] Boe.gov.sa., "Personal data protection system in Saudi Arabia," Personal Data Protection System in September 2021, 2021.

[23] GOV.UK. (2022). New laws to better protect victims from abuse of intimate images. [Online]. Available: https://www.gov.uk/government/news/new-laws-to-better-protect-victims-from-abuse-of-intimate-images

[24] M. Plaha and J. Lee. Sharing pornographic deepfakes to be illegal in England and Wales. [Online]. Available: https://www.bbc.co.uk/news/technology-63669711

[25] Europarl. (2023). EU AI act: first regulation on artificial intelligence. [Online]. Available: https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

[26] E. Gerstner, "Face/Off: 'DeepFake' face swaps and privacy laws," *Defense Counsel J.*, vol. 87, 2020.

[27] V. K. Kigwiru, "Deepfake technology and elections in Kenya: Can legislation combat the harm posed by deepfakes?" 2022.

[28] M. F. Ferraro, "Deepfake legislation: A nationwide survey—State and federal lawmakers consider legislation to regulate manipulated media," WilmerHale Report, 2019.

[29] G. Alon, A. Haider, and H. Hel-Or, "Judicial errors: Fake imaging and the modern law of evidence," *UIC Review of Intellectual Property Law*, vol. 21, no. 1, 2022.

[30] Futureoflife. (2023). Pause giant AI experiments: An open letter. [Online]. Available: https://futureoflife.org/open-letter/pause-giant-ai-experiments/