Deepfakes in Developing Societies: Handling the Societal Impacts and Cross-Disciplinary Vulnerabilities in Tech-Limited Environments

Rahatun Nesa Priti*, Mahir Absar Khan*, Abdur Rahman*, Azmine Toushik Wasi*†

Shahjalal University of Science and Technology, Sylhet, Bangladesh
Correspondance to: azmine32@student.sust.edu

Abstract

Deepfakes, AI-generated multimedia content that includes images, videos, audio, and text designed to mimic real media, have become increasingly prevalent. Their rise poses substantial risks to political stability, social trust, and economic well-being, especially in developing societies with limited media literacy and technological infrastructure. The motivation for this work stems from the urgent need to understand how these technologies are perceived and how they affect communities with limited resources to combat misinformation. We conducted a detailed survey to assess public awareness, perceptions, and experiences with deepfakes, followed by the development of a comprehensive framework for managing their impact. The framework addresses prevention, detection, and mitigation of deepfakes, providing practical strategies tailored for tech-limited environments. Our findings reveal a critical knowledge gap and a lack of effective detection tools, highlighting the need for targeted public education and accessible verification tools. In conclusion, this work offers actionable insights to support vulnerable populations in managing the challenges posed by deepfakes and calls for further interdisciplinary efforts to tackle these issues.

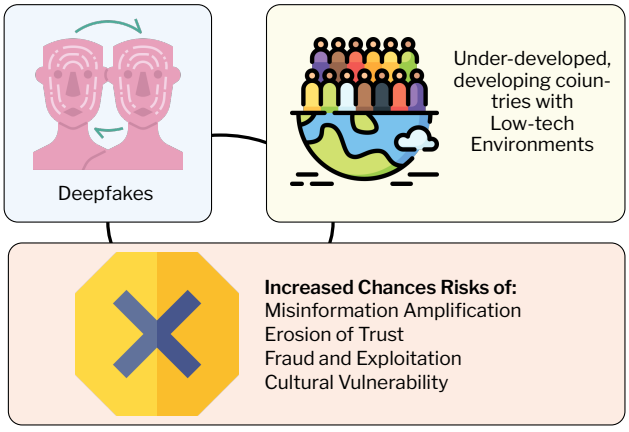


Figure 1: Motivation behind this study is to understand the adverse impact of deepfakes in tech-limited environments, where limited access to digital resources and media literacy exacerbate the risks posed by AI-generated content.

Introduction

Deepfakes, a form of AI-generated multimedia content, include images, videos, audio, and text designed to mimic real media (Fagni et al. 2021). The term combines ”deep learning,” a type of artificial intelligence, with ”fake,” reflecting how advanced algorithms are used to manipulate or create content (Chadha et al. 2021a). This technology excels at producing highly realistic material, often blurring the boundary between reality and fabrication. While deepfakes hold legitimate applications in fields such as entertainment, education, and advertising (Mubarak et al. 2023), their misuse has sparked significant societal concerns. For instance, deepfakes have been implicated in identity theft, fraud, revenge tactics, and even threats to national security. The underlying mechanism relies on AI models trained on vast datasets to generate realistic outputs. A typical example includes a deepfake video portraying someone saying or doing things they never actually did. On the positive side, such technology enables innovations like virtual try-ons in fashion or creating realistic special effects in movies (Chadha et al. 2021b). However, the potential for misuse is equally profound, as deepfakes can deceive audiences or cause personal and collective harm. This duality highlights the critical need to understand both the potential and the risks associated with deepfakes, as well as to develop strategies to mitigate their adverse impacts.

Rise of deepfake technology has raised significant concerns, particularly in underdeveloped and developing countries with low-tech environments (AL-KHAZRAJI et al. 2023). Deepfakes—hyper-realistic AI-generated media—pose risks to political stability, societal trust, and economic conditions, especially in regions with low media literacy and limited capacity to address such threats. A major concern is the potential for deepfakes to spread misinformation and manipulate public opinion. In politically fragile nations, deepfakes can distort reality by creating false narratives that undermine trust in leadership or alter elections (Qureshi 2024). For example, a deepfake video of a leader making controversial statements can lead to public

*These authors contributed equally.
†Corresponding Author

unrest or influence electoral outcomes. In areas with limited reliable media, social media amplifies the reach of these manipulated videos, eroding confidence in governance and media (Christofoletti 2024). Beyond political manipulation, deepfakes contribute to a broader erosion of media trust. As people realize that videos and images can be easily altered, skepticism grows, especially in democratic societies reliant on informed citizens. In environments with limited media literacy and verification tools, the public is more susceptible to manipulation, increasing societal divisions and confusion during key events like elections or protests (Dudka 2023). Economically, deepfakes also pose severe risks. In fragile economies, deepfake-driven scams, including identity theft and fraudulent transactions, can result in significant financial losses (Hummer and J. Rebovich 2023). Combating misinformation—through legal actions, awareness campaigns, and detection technologies—requires resources that many developing countries lack. Culturally, deepfakes can exacerbate existing social tensions, manipulate values, and incite conflict, targeting specific ethnic, religious, or political groups. This deepens divisions and undermines societal harmony (Achyut 2023). The urgency of addressing these challenges is compounded by the rapid spread of digital content and limited capacity in developing countries to regulate emerging technologies. While internet access and social media exposure are growing, these regions lack the infrastructure and legal frameworks to mitigate the impact of deepfakes, leaving them vulnerable to misinformation that can harm social cohesion, political stability, and economic wellbeing.

In this paper, we propose a comprehensive framework to address the societal impacts and cross-disciplinary vulnerabilities of deepfake technology, particularly in tech-limited environments. Our approach aims to tackle the issue from an end-to-end perspective, beginning with the creation of deepfakes, extending to their spread, and addressing the consequences after they have disseminated. To understand the unique challenges faced by people in low-tech environments, we conducted a detailed survey to gauge their perceptions, fears, and the ways in which they believe deepfakes affect their communities. The survey sought to explore not only individual awareness but also how people perceive their ability to handle deepfake content, including how they think their families and friends are impacted. By gathering these insights, we were able to shape a framework that takes into account their concerns and offers practical solutions for deepfake management. Our framework is structured into three key stages: prevention, detection, and mitigation. In the prevention stage, we address the creation of deepfakes, focusing on the ethical challenges and technological capabilities involved in their development. We propose regulations, public awareness initiatives, and the role of artificial intelligence in identifying and preventing the creation of harmful content. The detection stage deals with identifying deepfakes as they spread, emphasizing the importance of social awareness and AI tools that can quickly identify manipulated content. Finally, the mitigation stage addresses the aftermath of deepfake dissemination, including legal responses, public education efforts, and ethical guidelines to limit the harm caused by deepfake misuse.

Related Works

Several studies have explored various aspects of deepfake technology, examining its societal impact, detection strategies, ethical concerns, and public awareness. Shoaib et al. (2023) provide an overview of how advanced AI technologies contribute to the spread of misinformation and disinformation, particularly in vulnerable societies. In a similar vein, Patel et al. (2023) present a case study on deepfake generation and detection, highlighting the unique challenges faced in developing nations. Gregory (2023) examine incidents where deepfakes have been used to undermine human rights in these contexts, proposing defense strategies. Regionally, Misirlis and Munawar (2023) offer an analysis of the risks and potential benefits of deepfake misuse across various settings. Alanazi and Asif (2024) discuss how deepfakes, originally intended for entertainment, have been misused to produce explicit content and misinformation, leading to societal harm. Additionally, Hancock and Bailenson (2021) explore the psychological, social, and policy implications of deepfakes, particularly their threat to privacy, democracy, and national security. The influence of deepfake videos on public perception, which can result in deception and the erosion of trust in news media, is addressed by Vaccari and Chadwick (2020). Furthermore, Li, An, and Zhang (2021) investigate the ethical dilemmas surrounding deepfake technology, focusing on issues of privacy, consent, and the misuse of misleading content. Detection techniques and their effectiveness are systematically reviewed by Rana et al. (2022), highlighting the need for improvements. Finally, Sippy et al. (2024) examine public awareness and perceptions of deepfakes, emphasizing the critical need for education on the manipulation of AI-generated content.

Our work differentiates itself by focusing specifically on the societal impacts of deepfakes in underdeveloped and developing regions with low-tech environments. While previous studies primarily address the technical aspects, detection methods, and ethical concerns, we emphasize the unique vulnerabilities of these societies, including limited media literacy, fragile political systems, and scarce resources for combating misinformation. We propose a comprehensive framework to address these challenges, incorporating public awareness, legal frameworks, and technological solutions tailored to these contexts.

Methodology

Our methodology involves three key steps: conducting a survey to understand perceptions and concerns about deepfakes in low-tech environments, collecting data on how individuals and their communities are affected, and developing a comprehensive framework that addresses the challenges and provides practical solutions for managing deepfakes.

Designing Survey

The survey was designed to assess the awareness, perceptions, and societal impacts of deepfake technology within a developing society. The primary objective was to understand

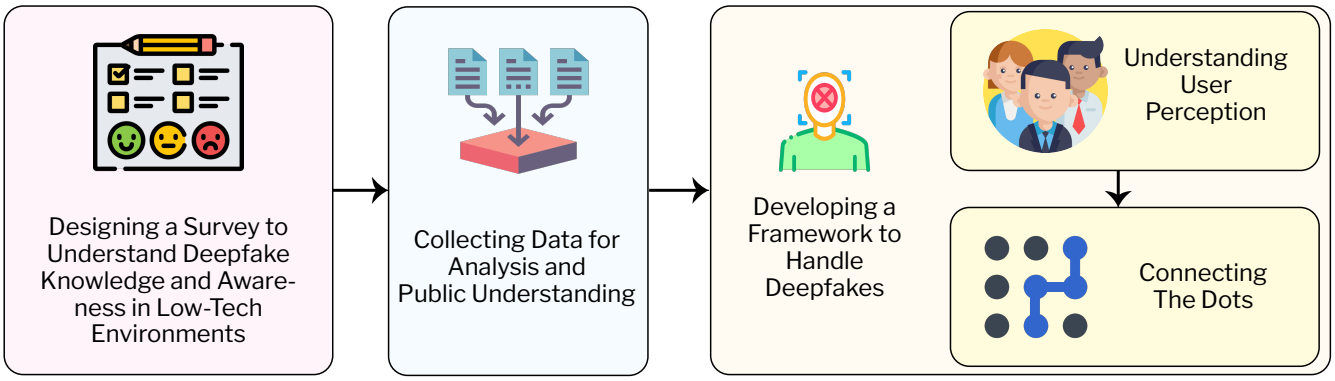


Figure 2: **Research Methodology:** Conducting a survey to assess perceptions and concerns regarding deepfakes in low-tech environments, gathering data on their impact on individuals and communities, and creating a comprehensive framework that tackles these challenges while offering practical solutions for managing deepfake-related issues.

the relationship between individuals and their ability to detect deepfakes, as well as the broader societal consequences of their spread. Specifically, the survey aimed to gauge the extent of familiarity with deepfakes, the frequency of exposure to such content, and the potential consequences participants have experienced or observed. Furthermore, the survey sought to explore public concerns about the misuse of deepfakes, focusing on their capacity to mislead, manipulate, and sway public opinion. By focusing on a developing society with limited technological infrastructure, this survey was intended to uncover the broader societal ripple effects of deepfake technology. It aimed to identify areas where targeted educational initiatives, policy interventions, and the implementation of countermeasures are needed to address the societal challenges posed by deepfake content.

Data Collection

A total of 73 responses were collected from participants in Bangladesh, with a focus on urban populations. The majority (75.3%) of respondents lived in urban areas, 23.3% in suburban regions, and 1.4% in rural areas, reflecting a more modernized and technologically exposed segment. The gender distribution was balanced, with 52.1% identifying as female and 47.9% as male, offering diverse perspectives. Most respondents (83.6%) were between 18 and 24 years old, with 16.4% in the 25-34 age group, indicating a focus on younger, digitally engaged individuals. Educationally, 98.6% had higher education, and 90.4% identified as students, highlighting the intellectual curiosity of the group. A key finding was that 68.5% of respondents had heard of ”deepfake” and understood its meaning, reflecting growing awareness in a young, educated demographic. However, 31.5% were unfamiliar with the term, emphasizing the need for continued education and awareness in developing technological contexts.

Framework Design

After designing the survey and collecting data, we analyze the findings to develop a framework that addresses concerns, incorporates suggestions, and aligns with regulatory changes.

Understanding User Perceptions Deepfakes significantly impact social dynamics in developing societies, particularly where fact-checking resources are limited. These hyperrealistic media can manipulate public perception, incite unrest, and threaten political stability and information integrity, especially in fragile democracies (Noor, Malahat, and Noor 2024). The rapid spread of misinformation through social media exacerbates the issue, with deepfakes often targeting political figures to undermine their credibility (Veerasamy and Pieterse 2022; Ekpang, Iyorza, and Ekpang 2023). For instance, during Nigeria’s recent elections, deepfakes were used for mudslinging and propaganda, distorting political narratives (Ekpang, Iyorza, and Ekpang 2023). As deepfake technology becomes more sophisticated, concerns over trust and social interaction rise, as individuals struggle to differentiate reality from manipulation (Doğan Akkaya 2024; Fehring and Bonaci 2023).

Survey responses reveal growing awareness of deepfakes, particularly in tech-limited societies. 68.5% of respondents were somewhat familiar with deepfakes, and 15.1% were very knowledgeable, but a significant knowledge gap remains, with 16.4% hearing the term but lacking understanding. This highlights the need for public education on deepfakes and their consequences. Notably, 72.1% had encountered deepfake content, even in low-tech environments. Alarmingly, 17.8% experienced harm, while 43.8% were unsure of its impact, suggesting harm may be underreported due to lack of awareness. Emotional distress (41.1%) was the most common consequence, followed by financial (5.5%) and physical distress (8.2Societal ripple effects are significant: 20.5% believed deepfakes had impacted their friends or family, while 39.7% were unsure, indicating potential underestimation of indirect harm. A large majority (91.8%) expressed concern over deepfakes misleading their community. Additionally, 53.4% perceived deepfakes as harmful, and 73.9% were worried about their misuse. Notably, 43.8% believed deepfakes could influence public opinion and decisions, linking exposure to deepfakes with perceived societal consequences. These findings underscore

the urgent need for research and intervention in tech-limited societies, where misinformation can spread unchecked and countermeasures are scarce. A cross-disciplinary approach involving technology, sociology, psychology, and policy-making is essential to mitigate deepfake risks and protect vulnerable populations. Detailed statistics are available in Figure 4 and 5.

Connecting the Dots In this section, we connect the survey findings to actionable insights, bridging perceptions with components to consider for the framework.

Low Confidence in Deepfake Detection. The survey reveals a significant gap in respondents' ability to detect deepfakes, with 16.4% expressing low confidence and 4.1% reporting no confidence at all. This uncertainty extends to the respondents' perceptions of their acquaintances' detection skills, with 45.2% expressing moderate confidence and 26% indicating low confidence. These findings highlight a critical need for awareness and training in deepfake detection. Public education campaigns should be launched to equip individuals with the tools and techniques needed to identify AI-generated content. Furthermore, educational programs could be incorporated into schools, workplaces, and public spaces to promote digital literacy and content verification. By addressing this knowledge gap, communities will be better prepared to protect themselves from the risks posed by deepfakes.

Gaps in Information Verification Practices. Although 69.9% of respondents report cross-checking information with other sources, 21.9% do not engage in verification at all. This indicates a lack of awareness or resources for verifying content, which could be mitigated by providing accessible, AI-powered verification tools. These tools could be integrated into social media platforms, news websites, and digital services, enabling real-time content verification. Additionally, educational programs focusing on the importance of information verification should be prioritized to help individuals understand the significance of skepticism and develop the habit of cross-checking content before sharing. Ensuring that people are equipped with these practices is vital in a world where misinformation spreads rapidly.

Trust Issues in Media and Misinformation. The survey highlights widespread concern about the use of deepfakes to manipulate public perception, with 91.8% of respondents expressing concern about their potential to deceive. This reflects broader mistrust in media and digital content, exacerbated by deepfake technology. To rebuild trust, media outlets must prioritize transparency by clearly labeling AI-generated or manipulated content. This transparency would allow audiences to assess the authenticity of the media they consume. Public awareness campaigns should also be implemented to teach people how to critically evaluate digital content and recognize potential manipulation. By fostering a culture of responsible media consumption, these campaigns can help reduce misinformation and restore trust in media organizations.

Lack of Confidence in AI Detection Tools. While 50% of respondents support the use of AI detection tools, concerns about their effectiveness persist. To address these doubts, it is essential to make AI detection tools widely accessible and user-friendly. These tools should be integrated into popular platforms such as social media, video-sharing sites, and news outlets to facilitate real-time content verification. Public training sessions or online tutorials could also be provided to ensure individuals know how to use these tools effectively. Ongoing development and updates to these tools will be crucial in keeping pace with evolving deepfake technologies. By ensuring accessibility and reliability, these tools can boost public confidence in detecting and mitigating deepfake-related misinformation.

Concerns About Regulation and Accountability. While 63% of respondents support better regulation of deepfakes, there is less enthusiasm for regulating AI technologies themselves. This suggests that while there is recognition of the need for oversight, people may be wary of overregulation that could stifle technological innovation. A balanced approach to regulation is necessary, one that addresses malicious uses of deepfake technology without impeding the progress of beneficial AI advancements. Governments, tech companies, and civil society organizations should collaborate to create clear, targeted regulations that safeguard public safety and ethical standards while promoting innovation. Moreover, regulatory bodies should work with AI developers to ensure that detection tools are regularly updated to counter emerging threats from deepfake content.

Strategies to Restore Trust in Media. The survey indicates that respondents see AI-powered deepfake detection as the most effective strategy for restoring trust in media, with 68.5% supporting its implementation. Transparency from media outlets, improved regulations, and public awareness campaigns also received significant backing. To restore trust, AI-driven fact-checking systems should be deployed across digital platforms to flag deepfakes and misinformation in real-time. Media organizations must also adopt transparent policies, clearly marking AI-generated or altered content to help viewers distinguish between authentic and manipulated media. Combined with public education on media literacy and the importance of verifying information, these strategies will work together to mitigate the effects of deepfakes and rebuild trust in the media.

Framework

In this section, we present the components and stages of the framework we have designed to address the challenges posed by deepfakes. The framework is divided into three key stages: Creating Deepfakes, Spreading Deepfakes, and Responding to Deepfakes. Each stage is discussed in detail, outlining the associated components and strategies necessary to manage the risks and impacts of deepfakes effectively.

Creating Deepfakes

In this stage, we discuss and explore components related to deepfake creation, strategies to control and regulate deepfake generation, and methods to increase public awareness to avoid misuse.

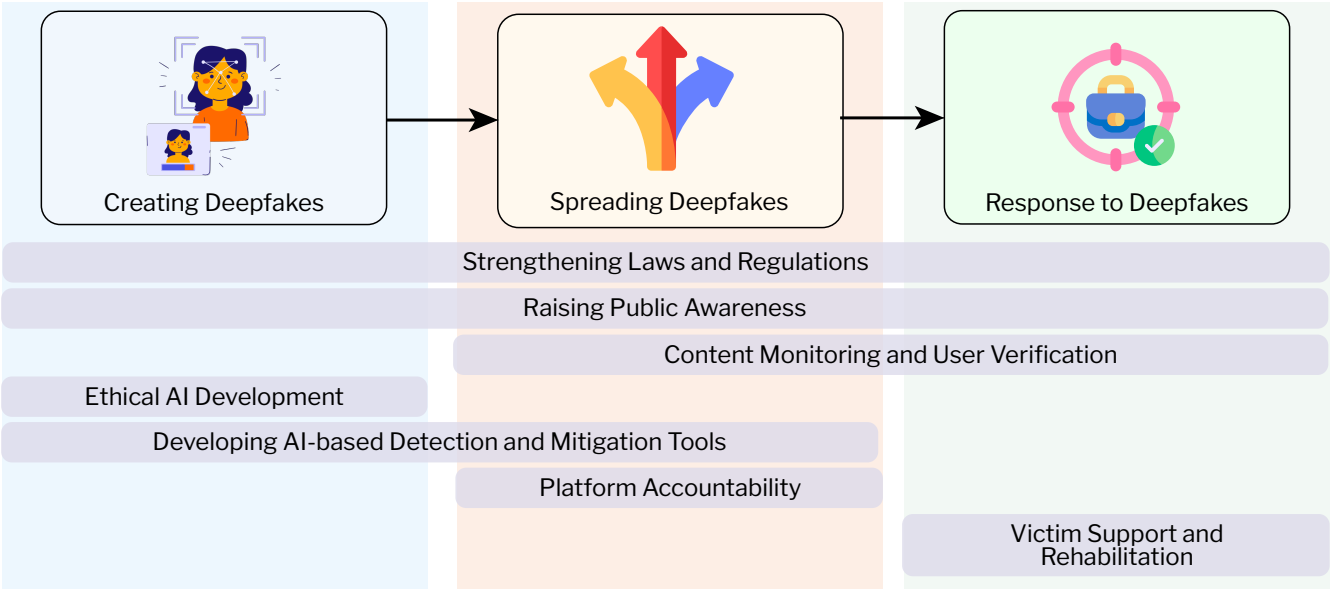


Figure 3: Our framework for handling deepfakes in low-tech environments consists of three key stages: prevention, detection, and mitigation. In the prevention stage, we focus on the ethical challenges and technological factors involved in the creation of deepfakes, aiming to reduce their emergence and impact.

Strengthening Laws and Regulations Existing laws and regulations surrounding deepfakes are still evolving and vary significantly across jurisdictions. In the European Union, the proposed AI Act is a key regulatory effort, focusing on transparency obligations for creators rather than banning deepfakes outright, as seen in Article 52(3), which requires creators to disclose when content is AI-generated or manipulated (de Almeida, dos Santos, and Farias 2021). However, while this establishes a model for regulation, it mainly addresses the use of deepfakes rather than their creation. In the U.S., tort laws like ”appropriation of name or likeness” and ”false light publicity” can provide avenues for addressing harm caused by deepfakes, but challenges persist, such as the difficulty in proving harm or the balance with First Amendment rights (Mania 2022). Some states have taken more specific actions, like Indiana’s HB 1133 requiring disclaimers on fabricated media (Mania 2022) and Florida’s SB 1798 criminalizing deepfakes involving minors (Mania 2022). Internationally, Australia has introduced legislation targeting deepfake sexual content without consent. However, gaps remain in addressing the creation of deepfakes, particularly in the context of privacy, intellectual property, and ethical considerations. What is needed is a comprehensive global framework that regulates both the creation and distribution of deepfakes, ensuring creators are held accountable, while also protecting freedom of speech. Strengthening the alignment between existing laws, such as copyright and privacy protections (Widder et al. 2022), and new AI-focused legislation will be essential to address the multifaceted challenges posed by deepfake technology. Legal clarity, better enforcement, and cross-border cooperation will be necessary to curb the creation and misuse of deepfakes effectively.

Raising Public Awareness Raising public awareness during the deepfake creation stage is crucial to preventing misuse and mitigating its harmful effects. Educational campaigns should focus on informing the public about the existence of deepfakes and their potential consequences, such as disinformation or privacy violations (Tan et al. 2024). Platforms and policymakers can collaborate with media organizations to highlight ethical guidelines and the dangers of creating harmful deepfakes. Additionally, promoting awareness of detection technologies and their role in identifying manipulated content can empower users to be more critical of what they encounter online (Jiang et al. 2024). Legal frameworks like the Deepfake Accountability Act can also raise awareness by holding creators accountable and emphasizing the importance of ethical practices (Masood et al. 2022).

Ethical AI Development We need a multidimensional approach to develop ethical AI systems that cannot be used to create harmful deepfakes, combining technical and social frameworks. Developers should establish clear guidelines for responsible AI creation, emphasizing the prohibition of malicious applications like non-consensual pornography and misinformation campaigns. This includes incorporating built-in safeguards to limit unethical uses, robust access controls on open-source platforms such as GitHub, and auditing mechanisms to track AI model usage in real time (Masood et al. 2022). Collaborating with ethicists, legal experts, and affected communities ensures AI systems align with societal norms and values. Transparency, while vital for accountability, must pair with ethical constraints to prevent exploitation by malicious actors (Li, An, and Zhang

2021). Technical measures, such as watermarking and automated detection tools, can further discourage misuse (Verdoliva 2020). Raising public awareness, supported by journalistic efforts and regulatory bodies, reinforces the importance of ethical AI practices in combating digital deception (Widder et al. 2022). By promoting responsible development, involving diverse stakeholders, and enforcing accountability, AI systems can prioritize societal well-being over harmful applications.

Developing AI-based Detection and Mitigation Tools Developing AI-based detection and mitigation tools is crucial for preventing deepfake creation by proactively identifying manipulations during content creation and distribution. AI-driven models, such as convolutional and recurrent neural networks, can detect subtle discrepancies in videos, including unnatural facial expressions, lighting inconsistencies, or pixel-level changes (Jiang et al. 2024). Integrating these tools into content-sharing platforms can screen media before wide distribution, reducing the risk of malicious content going viral (Tan et al. 2024). Additionally, monitoring physiological cues like irregular eye blinking and lip-syncing inconsistencies provides an extra layer of protection (Dolhansky et al. 2020). Metadata analysis using AI models helps spot post-creation anomalies, ensuring media authenticity (Whyte 2020). Watermarking technologies, including blockchain-based solutions, trace media origins, complicating manipulation without detection (Solaiman and Rana 2024). Compression artifact detection tools, such as GANalyzer, expose flaws often present in deepfake-generated content (Akhtar, Pendyala, and Athmakuri 2024).

Spreading Deepfakes

In this stage, we discuss and explore components designed to prevent the spread of deepfakes after they have been created. This includes identifying technologies and mechanisms that can detect deepfakes in real-time, developing systems to flag and remove such content from social media and other platforms, and establishing collaboration with tech companies to enforce these measures.

Strengthening Laws and Regulations Existing laws and regulations provide some frameworks for addressing deepfakes but remain fragmented and insufficient for comprehensive deterrence. In the U.S., laws like Texas' amended Election Code[1] and California's Deceptive Audio or Visual Media Act [2] criminalize deepfake content aimed at deceiving voters (Farish 2019). Similarly, the EU's proposed AI Act mandates transparency, requiring creators to disclose AI manipulation in media, which could limit deepfake spread if effectively enforced (de Almeida, dos Santos, and Farias 2021). However, these measures focus more on consequences after dissemination than on prevention or rapid detection during viral spread. Privacy laws like the GDPR protect individuals from unauthorized use of their likeness but do not fully address the rapid proliferation of deepfakes on social media (Hoofnagle, van der Sloot, and Borgesius 2019). A holistic approach is needed, integrating content moderation, stricter penalties for platforms failing to remove harmful deepfakes, and real-time detection systems. Regulations should hold platforms accountable for rapid removal, enforce transparency in AI-generated content, and encourage the adoption of advanced detection tools. International cooperation is also essential to ensure consistent enforcement across borders.

Raising Public Awareness To address the risks of deepfakes, enhancing public awareness before their spread is vital. A comprehensive strategy includes promoting digital literacy, ethical guidelines, and educational initiatives to help the public recognize manipulated content. Social media platforms, such as Instagram with its educational campaigns and reporting features, demonstrate how proactive measures can inform users (Cochran and Napshin 2021). Integrating media literacy into curricula, like Google's Digital Literacy and Citizenship program, enhances critical thinking and helps students evaluate online content (Whyte 2020). Schools can further support this through workshops and seminars on identifying fake media. Accessible detection tools, such as Deepware Scanner and Sensity AI, empower users to verify media authenticity (Wang et al. 2024). Public awareness campaigns, such as the EU's #StopFakeNews initiative, educate wider audiences about deepfake risks through diverse media channels (Samuel-Okon et al. 2024). Platforms must also invest in AI-powered detection systems, collaborating with initiatives like the Deepfake Detection Challenge to flag manipulated content in real-time (Wang et al. 2024). Governments can contribute by enacting laws like the Deepfake Accountability Act and the EU's Digital Services Act to hold platforms accountable and mandate pre-upload AI checks for deepfakes (Whittaker et al. 2023). Combining education, technology, legislation, and proactive platform policies creates a multifaceted approach to reducing deepfake creation and spread, fostering a more informed society.

Content Monitoring and User Verification Content monitoring and user verification are crucial measures for preventing the spread of deepfakes on platforms like YouTube, Facebook, and Twitch. By utilizing advanced detection technologies, these platforms can identify and remove harmful content before it reaches a large audience. For example, GAN Fingerprinting helps identify deepfakes by analyzing patterns unique to Generative Adversarial Networks (GANs), allowing platforms to trace the source of manipulations (Frank and Schönherr 2021). Furthermore, AI-based models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) analyze video frames to detect inconsistencies in facial expressions or skin texture (Jiang et al. 2024). Physiological analysis tools like DeepFake-o-meter[3] can also spot unnatural eye-blinking patterns (Dolhansky et al. 2020), while metadata analysis can detect discrepancies in timestamps or camera details, helping verify the authenticity of media (Solaiman and Rana 2024). Platforms can integrate watermark-

[1] https://statutes.capitol.texas.gov/Docs/EL/htm/EL.274.htm

[2] https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill$_id = 202320240AB2839$

[3] https://zinc.cse.buffalo.edu/ubmdfl/deep-o-meter/landing$_page$

ing techniques, including blockchain-based digital watermarks, to ensure content integrity (Whyte 2020). Audio-visual synchronization tools, like Microsoft's Video Authenticator, help detect mismatches between speech and lip movements (Akhtar, Pendyala, and Athmakuri 2024). By combining these technologies with user verification measures, such as account authentication and behavior monitoring, platforms can create a safer digital environment, reduce the likelihood of deepfake spread, and uphold legal and ethical standards in content distribution. Fact-checkers like PolitiFact[4] and FactCheck.org[5] are essential in combating deepfakes by using expert analysis and AI tools to verify content authenticity. They cross-reference manipulated material with trusted sources, such as public appearances or reputable news outlets, to identify inconsistencies. For example, when a deepfake portrays a politician making false statements, fact-checkers compare it with existing transcripts and reports to expose discrepancies. Platforms like Facebook collaborate with these organizations to label deepfake content, offering context to users. Additionally, crowd-sourcing through user reports helps platforms like YouTube and Twitter quickly verify suspicious content with input from experts and AI systems, improving the fight against deepfakes (Masood et al. 2022; Yadlin-Segal and Oppenheim 2020; Hoofnagle, van der Sloot, and Borgesius 2019; Widder et al. 2022).

Developing AI-based Detection and Mitigation Tools AI-based detection and mitigation tools are essential in preventing the spread of deepfakes, especially on platforms like YouTube, Facebook, and Twitch. By leveraging machine learning, AI can identify deepfake content before it gains widespread attention. GAN Fingerprinting detects patterns unique to Generative Adversarial Networks (GANs), such as pixel-level inconsistencies, enabling early identification and source tracing (Frank and Schönherr 2021). Blockchain technology secures digital content by embedding cryptographic signatures to verify authenticity and ensure media integrity post-distribution (Whyte 2020). AI systems like Microsoft's Video Authenticator use neural networks to analyze facial and motion inconsistencies, assigning confidence scores to manipulation likelihood (Heidari et al. 2024). Multi-modal detection systems analyze both visual and audio components, improving detection accuracy by cross-verifying inconsistencies (Ki Chan et al. 2020). Real-time detection tools, such as TrueMedia, monitor live-streamed content for deepfake signs, crucial during events like elections (Tan et al. 2024). Audio forensic analysis tools identify synthetic audio manipulations, offering a comprehensive approach to media verification. These technologies help detect and prevent deepfakes from spreading, contributing to a safer digital environment as they evolve.

Platform Accountability Ensuring platform accountability for managing deepfakes involves a combination of legislative actions, technical measures, and corporate responsibility. Laws like the U.S. Deepfake Accountability Act[6] and the EU's Digital Services Act[7] hold platforms accountable for hosting harmful content, while transparency initiatives, such as labeling AI-generated content, enhance trust. Major tech companies, including Meta and IBM, are investing in AI tools for real-time deepfake detection and collaboration with cybersecurity experts. Platforms use advanced detection methods like GAN fingerprinting and multi-modal systems to identify manipulations, while combining automated systems with human oversight to prevent the spread of harmful content (Tan et al. 2024).

Response to Deepfakes

In this stage, we focus on strategies to respond to deepfakes after their spread. This includes deploying detection tools to verify authenticity, clearly labeling or debunking false media, and issuing corrections through trusted channels like official statements.

Strengthening Laws and Regulations After a deepfake is spread, existing laws and regulations can help mitigate or reduce the risks and harms, though challenges remain in their effectiveness. Laws like the U.S. tort laws, including "false light publicity" and "appropriation of name or likeness," provide victims with avenues to seek redress if deepfakes harm their reputation or emotional well-being (Mania 2022). For instance, individuals can claim damages for the emotional distress caused by manipulated content. However, these laws require victims to prove harm, which can be difficult when the content has already gone viral. In the EU, the AI Act's transparency obligations could help mitigate the spread by forcing creators to disclose AI manipulation, making users more aware of the content's authenticity (de Almeida, dos Santos, and Farias 2021). In addition, GDPR can be instrumental when deepfakes violate personal data protections, as manipulated content that features identifiable individuals falls under GDPR's[8] jurisdiction, enabling individuals to request the removal of such content from online platforms (Hoofnagle, van der Sloot, and Borgesius 2019). Despite these existing laws, they often lack the speed and scale required to address the widespread, real-time nature of deepfakes.

Raising Public Awareness After a deepfake spreads, raising public awareness is crucial to minimize harm and prevent victim-blaming. Campaigns should educate the public about the potential for manipulation, highlighting that anyone can be targeted by deepfake technology, thus discouraging victim-blaming (Yadlin-Segal and Oppenheim 2020). Additionally, raising awareness about legal avenues for redress can empower victims, ensuring they know how to report and seek justice for harm caused by deepfakes (Hoofnagle, van der Sloot, and Borgesius 2019). Promoting media literacy and fact-checking through trusted platforms like PolitiFact (Masood et al. 2022) is essential for encouraging

[4]https://www.politifact.com/

[5]https://www.factcheck.org/

[6]https://www.congress.gov/bill/118th-congress/house-bill/5586/text

[7]https://www.eu-digital-services-act.com/

[8]https://gdpr-info.eu/

users to verify suspicious content. Platforms like Facebook collaborate with fact-checking organizations to label manipulated content and provide context, helping users identify deepfakes and understand the risks (Tan et al. 2024). Additionally, legal frameworks, such as the EU Digital Services Act[9], enable swift content removal and accountability for perpetrators, while blockchain-based content provenance systems ensure transparency and traceability (Masood et al. 2022). Platforms can track digital footprints and metadata to trace the origins of manipulated content, aiding in identifying creators or distributors (Whyte 2020). AI-driven tools like Sensity AI[10] and Deepware Scanner[11] help detect deepfakes in real-time, preventing further spread (Samuel-Okon et al. 2024). These combined efforts protect victims and empower users to combat the impact of deepfakes.

Content Monitoring and User Verification Content monitoring and user verification can play a crucial role in identifying offenders and supporting legal cases after a deepfake has spread. Platforms can track the origins of deepfakes by analyzing the metadata and digital footprints left by the content's creator, helping identify the individual responsible for its creation or distribution (Whyte 2020). User verification methods, such as requiring verified accounts for content upload, can deter malicious actors and provide a clear chain of accountability. When deepfakes are flagged, the platform's monitoring tools can cross-check the content against known sources and user reports to assess its authenticity, aiding law enforcement in identifying perpetrators. Additionally, ensuring the authenticity and integrity of media content involves advanced technologies like digital certificates, cryptographic hashes, and AI-based authentication tools. For example, Microsoft Azure uses digital certificates and cryptographic hashes to verify content authenticity (Samuel-Okon et al. 2024), while AI tools like Microsoft Video Authenticator detect deepfake artifacts and provide confidence scores to alert users to potential manipulations (Farish 2019). Platforms also integrate real-time content verification systems, flagging suspicious media upon upload to proactively prevent the spread of manipulated content. Blockchain-based systems, like Project Origin from the BBC, establish secure media provenance, offering irrefutable evidence of manipulation and origin, supporting legal claims and ensuring that victims have the necessary tools to seek justice

Victim Support and Rehabilitation After a deepfake spreads, providing comprehensive support and rehabilitation to victims is essential, as they often face significant emotional and psychological distress. Legislative frameworks must be updated to criminalize the malicious creation and distribution of deepfakes, offering victims clearer legal recourse (Romero Moreno 2024). Legal aid services are crucial to help victims navigate the complex legal landscape. Mental health support, including specialized counseling and therapy, is vital for recovery (Mania 2022). Platforms should establish rapid response systems for removing harmful content, enabling victims to flag deepfakes for quick removal, while automated detection tools can proactively identify them. Collaboration among legal teams, tech companies, and mental health professionals can streamline this process. NGOs should work with governments, law enforcement, and tech firms to provide immediate assistance through hotlines, helping to remove victims' images from platforms without delay (Haimson et al. 2021), and offer educational resources on victims' rights and available support services (Jiang et al. 2024). These efforts ensure a holistic recovery approach addressing legal, psychological, and technical needs.

[9]https://eur-lex.europa.eu/eli/reg/2022/2065/oj

[10]https://sensity.ai/

[11]https://scanner.deepware.ai/

Discussion

This research highlights the impact of deepfake technology in underdeveloped and developing regions, where access to technology and digital literacy is limited. By focusing on the experiences of individuals in these areas, we emphasize the need for public education campaigns and legal frameworks to address the risks posed by deepfakes. In regions with fragile trust in media, deepfakes threaten political stability and social cohesion, making it crucial to offer tools for better decision-making and protection. Our work extends beyond theory by providing concrete use cases, such as political manipulation and financial fraud, and offering solutions that can guide policymakers in addressing deepfake issues. In urban areas with high digital media use, public education can help citizens identify and report deepfakes, while international collaboration is crucial to tackling this global issue. We also examine ethical dilemmas, particularly in regions with limited digital literacy, proposing a balanced approach to mitigate harms while acknowledging AI's legitimate uses. We aim to inform policymakers, educators, and tech developers about deepfake risks in developing countries, emphasizing collaborative efforts for effective solutions. A challenge was varying participant awareness levels, with some recognizing risks while others were unaware. Our urban focus also limited rural perspectives, where digital literacy challenges and deepfake exposure may differ. As deepfake technology advances, the potential for misuse grows, highlighting the need for proactive measures by governments, tech companies, and educators. Our research provides a foundation for future studies exploring rural perspectives and societal impacts to develop better solutions.

Conclusion

In conclusion, this paper highlights the growing concern of deepfakes, particularly in tech-limited environments like Bangladesh, where misinformation can spread rapidly due to limited technological infrastructure. The survey findings reveal significant awareness gaps and the potential societal harm caused by deepfakes, including emotional and financial distress. It underscores the urgent need for public education, effective detection tools, and cross-disciplinary collaboration to mitigate these risks. By fostering resilience against deepfakes through prevention, detection, and mitigation strategies, communities can better navigate the ethical and societal challenges posed by this emerging technology.

References

Achyut, T. S. 2023. DeepFake Deception: A Comprehensive Analysis of DeepFake Technology and its Effects on Ethics, Politics and Society. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, 07(09).

Akhtar, Z.; Pendyala, T. L.; and Athmakuri, V. S. 2024. Video and Audio Deepfake Datasets and Open Issues in Deepfake Technology: Being Ahead of the Curve. *Forensic Sciences*, 4(3): 289–377.

AL-KHAZRAJI, S. H.; SALEH, H. H.; KHALID, A. I.; and MISHKHAL, I. A. 2023. Impact of Deepfake Technology on Social Media: Detection, Misinformation and Societal Implications. *The Eurasia Proceedings of Science Technology Engineering and Mathematics*, 23: 429–441.

Alanazi, S.; and Asif, S. 2024. Exploring deepfake technology: creation, consequences and countermeasures. *Human-Intelligent Systems Integration*.

Chadha, A.; Kumar, V.; Kashyap, S.; and Gupta, M. 2021a. *Deepfake: An Overview*, 557–566. Springer Singapore. ISBN 9789811607332.

Chadha, A.; Kumar, V.; Kashyap, S.; and Gupta, M. 2021b. *Deepfake: An Overview*, 557–566. Springer Singapore. ISBN 9789811607332.

Christofoletti, R. 2024. Trust in Media and journalism credibility in the sea of misinformation. *The International Review of Information Ethics*, 33(1).

Cochran, J. D.; and Napshin, S. A. 2021. Deepfakes: Awareness, Concerns, and Platform Accountability. *Cyberpsychology, Behavior, and Social Networking*, 24(3): 164–172.

de Almeida, P. G. R.; dos Santos, C. D.; and Farias, J. S. 2021. Artificial Intelligence Regulation: a framework for governance. *Ethics and Information Technology*, 23(3): 505–525.

Dolhansky, B.; Bitton, J.; Pflaum, B.; Lu, J.; Howes, R.; Wang, M.; and Ferrer, C. C. 2020. The DeepFake Detection Challenge (DFDC) Dataset.

Doğan Akkaya, F. 2024. Deepfake Dilemmas: Imagine Tomorrow's Surveillance Society through Three Scenarios. *Journal of Economy Culture and Society*, 0(0): 0–0.

Dudka, Y. 2023. The Manipulation of Public Consciousness: How Propaganda Methods Threaten the Stability of States and Regions. *The International Journal of Humanities amp; Social Studies*.

Ekpang, J. E.; Iyorza, S.; and Ekpang, P. O. 2023. SOCIAL MEDIA AND ARTIFICIAL INTELLIGENCE: PERSPECTIVES ON DEEPFAKES' USE IN NIGERIA'S 2023 GENERAL ELECTIONS. *Kampala International University Interdisciplinary Journal of Humanities and Social Sciences*, 4(2): 109–124.

Fagni, T.; Falchi, F.; Gambini, M.; Martella, A.; and Tesconi, M. 2021. TweepFake: About detecting deepfake tweets. *PLOS ONE*, 16(5): e0251415.

Farish, K. 2019. Do deepfakes pose a golden opportunity? Considering whether English law should adopt California's publicity right in the age of the deepfake. *Journal of Intellectual Property Law amp; Practice*, 15(1): 40–48.

Fehring, J. A.; and Bonaci, T. 2023. It looks like me, but it isn't me: On the societal implications of deepfakes. *IEEE Potentials*, 42(5): 33–38.

Frank, J.; and Schönherr, L. 2021. WaveFake: A Data Set to Facilitate Audio Deepfake Detection.

Gregory, S. 2023. Fortify the Truth: How to Defend Human Rights in an Age of Deepfakes and Generative AI. *Journal of Human Rights Practice*, 15(3): 702–714.

Haimson, O. L.; Delmonaco, D.; Nie, P.; and Wegner, A. 2021. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2): 1–35.

Hancock, J. T.; and Bailenson, J. N. 2021. The Social Impact of Deepfakes. *Cyberpsychology, Behavior, and Social Networking*, 24(3): 149–152.

Heidari, A.; Navimipour, N. J.; Dag, H.; Talebi, S.; and Unal, M. 2024. A Novel Blockchain-Based Deepfake Detection Method Using Federated and Deep Learning Models. *Cognitive Computation*, 16(3): 1073–1091.

Hoofnagle, C. J.; van der Sloot, B.; and Borgesius, F. Z. 2019. The European Union general data protection regulation: what it is and what it means. *Information amp; Communications Technology Law*, 28(1): 65–98.

Hummer, D.; and J. Rebovich, D. 2023. *Identity theft and financial loss*, 38–53. Edward Elgar Publishing. ISBN 9781800886643.

Jiang, L.; Yang, X.; Yu, C.; Wu, Z.; and Wang, Y. 2024. Advanced AI Framework for Enhanced Detection and Assessment of Abdominal Trauma: Integrating 3D Segmentation with 2D CNN and RNN Models. In *2024 3rd International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC)*, 337–340. IEEE.

Ki Chan, C. C.; Kumar, V.; Delaney, S.; and Gochoo, M. 2020. Combating Deepfakes: Multi-LSTM and Blockchain as Proof of Authenticity for Digital Media. In *2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G)*, 55–62. IEEE.

Li, H. Y.; An, J. T.; and Zhang, Y. 2021. Ethical Problems and Countermeasures of Artificial Intelligence Technology. *E3S Web of Conferences*, 251: 01063.

Mania, K. 2022. Legal Protection of Revenge and Deepfake Porn Victims in the European Union: Findings From a Comparative Legal Study. *Trauma, Violence, amp; Abuse*, 25(1): 117–129.

Masood, M.; Nawaz, M.; Malik, K. M.; Javed, A.; Irtaza, A.; and Malik, H. 2022. Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence*, 53(4): 3974–4026.

Misirlis, N.; and Munawar, H. B. 2023. From deepfake to deep useful: risks and opportunities through a systematic literature review.

Mubarak, R.; Alsboui, T.; Alshaikh, O.; Inuwa-Dutse, I.; Khan, S.; and Parkinson, S. 2023. A Survey on the Detection and Impacts of Deepfakes in Visual, Audio, and Textual Formats. *IEEE Access*, 11: 144497–144529.

Noor, L.; Malahat, I.; and Noor, H. 2024. The Socio-Political Implications of Deepfakes in Developing Countries.

Patel, Y.; Tanwar, S.; Gupta, R.; Bhattacharya, P.; Davidson, I. E.; Nyameko, R.; Aluvala, S.; and Vimal, V. 2023. Deepfake Generation and Detection: Case Study and Challenges. *IEEE Access*, 11: 143296–143323.

Qureshi, K. 2024. A Multidisciplinary Threats to Emerging Cybersecurity, Legal and Ethical Threats Posed by Deepfake Technology. *International Journal for Research in Applied Science and Engineering Technology*, 12(9): 574–580.

Rana, M. S.; Nobi, M. N.; Murali, B.; and Sung, A. H. 2022. Deepfake Detection: A Systematic Literature Review. *IEEE Access*, 10: 25494–25513.

Romero Moreno, F. 2024. Generative AI and deepfakes: a human rights approach to tackling harmful content. *International Review of Law, Computers amp; Technology*, 38(3): 297–326.

Samuel-Okon, A. D.; Akinola, O. I.; Olaniyi, O. O.; Olateju, O. O.; and Ajayi, S. A. 2024. Assessing the Effectiveness of Network Security Tools in Mitigating the Impact of Deepfakes AI on Public Trust in Media. *Archives of Current Research International*, 24(6): 355–375.

Shoaib, M. R.; Wang, Z.; Ahvanooey, M. T.; and Zhao, J. 2023. Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models.

Sippy, T.; Enock, F.; Bright, J.; and Margetts, H. Z. 2024. Behind the Deepfake: 8

Solaiman, M.; and Rana, M. S. 2024. Enhancing Global Security: A Robust CNN Model for Deepfake Video Detection. In *2024 7th International Conference on Information and Computer Technologies (ICICT)*, 238–243. IEEE.

Tan, C.; Zhao, Y.; Wei, S.; Gu, G.; Liu, P.; and Wei, Y. 2024. Frequency-Aware Deepfake Detection: Improving Generalizability through Frequency Space Domain Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(5): 5052–5060.

Vaccari, C.; and 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).

Veerasamy, N.; and Pieterse, H. 2022. Rising Above Misinformation and Deepfakes. *International Conference on Cyber Warfare and Security*, 17(1): 340–348.

Verdoliva, L. 2020. Media Forensics and DeepFakes: An Overview. *IEEE Journal of Selected Topics in Signal Processing*, 14(5): 910–932.

Wang, T.; Liao, X.; Chow, K. P.; Lin, X.; and Wang, Y. 2024. Deepfake Detection: A Comprehensive Survey from the Reliability Perspective. *ACM Computing Surveys*, 57(3): 1–35.

Whittaker, L.; Mulcahy, R.; Letheren, K.; Kietzmann, J.; and Russell-Bennett, R. 2023. Mapping the deepfake landscape for innovation: A multidisciplinary systematic review and future research agenda. *Technovation*, 125: 102784.

Whyte, C. 2020. Deepfake news: AI-enabled disinformation as a multi-level public policy challenge. *Journal of Cyber Policy*, 5(2): 199–217.

Widder, D. G.; Nafus, D.; Dabbish, L.; and Herbsleb, J. 2022. Limits and Possibilities for "Ethical AI" in Open Source: A Study of Deepfakes. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 2035–2046. ACM.

Yadlin-Segal, A.; and Oppenheim, Y. 2020. Whose dystopia is it anyway? Deepfakes and social media regulation. *Convergence: The International Journal of Research into New Media Technologies*, 27(1): 36–51.

Appendix: Questionnaire

Here is the full survey questionnaire:

1. Age:
   Options: i) Under 18 ii) 18-24 iii) 25-34 iv) 35-44 v) 45 and above
2. Gender:
   Options: i) Male ii) Female iii) Other (Please write your answer .....)
3. Highest Education Level:
   Options: i) No formal education ii) Primary school iii) Secondary school iv) Higher education (undergraduate, postgraduate) v) Other (Please write your answer .....)
4. Occupation:
   Options: i) Student ii) Corporate professional iii) Educator iv) Media/Journalism v) IT/Technology vi) Other (Please write your answer .....)
5. What type of area do you currently live in?
   Options: i) Urban ii) Suburban iii) Rural iv) Other (Please write your answer .....)
6. Country:
7. Have you heard of the term **"deepfake"**? Do you know what it means?
   Options: i) Yes ii) No
8. How would you describe your level of understanding of deepfakes?
   Options: i) Heard the term but not sure what it means properly ii) Somewhat knowledgeable iii) Very knowledgeable
9. Where have you learned about deepfakes?
   Options: i) Social media platforms (e.g., Facebook, Twitter, TikTok) ii) News outlets (online, TV, or print) iii) Educational resources (e.g., articles, courses, or lectures) iv) Friends or family v) Online videos or tutorials (e.g., YouTube) vi) Personal or professional experience vii) I have not heard about deepfakes before this survey
10. Have you ever encountered a deepfake (in videos, audio, or images)?
    Options: i) Yes ii) No

11. If yes, where did you encounter it?
Options: i) Social media ii) News media iii) Video platforms iv) Other (Please write your answer .....)

12. When you see a video or image online, how confident are you that it is authentic? (Not confident at all to Very confident)
Options: i) 1 ii) 2 iii) 3 iv) 4 v) 5

13. If yes, in which context did you encounter it? (Select all that apply)
Options: i) Social media (e.g., fake videos/images of public figures) ii) Entertainment (e.g., movies, memes, parodies) iii) Politics (e.g., fake speeches or announcements) iv) Cybersecurity (e.g., scams, phishing, identity theft) v) Personal experience vi) Other (Please write your answer .....)

14. How much deepfake content do you think is available on social media?
Options: i) Very little ii) 1 iii) 2 iv) 3 v) 4 vi) Almost all content

15. Did any deepfake content ever have a significant impact on you?
Options: i) No, it was harmless ii) Unsure iii) Yes, it was harmful

16. If it was harmful, what type of harm did it cause?
Options: i) No, it was harmless ii) Yes, it was emotionally distressing iii) Yes, it was physically distressing iv) Yes, it caused financial loss or harm v) Other (Please write your answer .....)

17. Did any deepfake content ever have a significant impact on your friends or family?
Options: i) No, it was harmless ii) Unsure iii) Yes, it was harmful

18. How harmful do you think deepfakes can be? (Very harmful to Not harmful at all)
Options: i) 1 ii) 2 iii) 3 iv) 4 v) 5

19. Do you think deepfakes could be used to mislead people in your community?
Options: i) Yes ii) No iii) Not sure

20. How likely do you think deepfakes are to influence public opinion or decisions in your community? (Very likely to Not likely at all)
Options: i) 1 ii) 2 iii) 3 iv) 4 v) 5

21. How concerned are you about the misuse of deepfakes? (Not concerned at all to Very concerned)
Options: i) 1 ii) 2 iii) 3 iv) 4 v) 5

22. Do you trust information from the following sources? (Rate each between: Very trustworthy, Somewhat trustworthy, Not trustworthy)
Options: i) News outlets (TV, online news) ii) Social media platforms (Facebook, Twitter, etc.) iii) Teachers/Seniors iv) Family/friends/neighbours

23. How do you typically verify the information you receive online? (Select all that apply)
Options: i) I don't verify the information ii) Cross-check with other sources iii) Ask others if they think it's real iv) Other (Please write your answer .....)

24. How confident are you in your ability to detect a deepfake? (Not confident at all to Very confident)
Options: i) 1 ii) 2 iii) 3 iv) 4 v) 5

25. How confident are you in your friends' or family's ability to detect a deepfake? (Not confident at all to Very confident)
Options: i) 1 ii) 2 iii) 3 iv) 4 v) 5

26. What do you think is the most effective way to combat deepfake-related issues? (Select all that apply)
Options: i) Hardening AI Use Laws and Regulations ii) Improved AI detection tools iii) Stricter laws and regulations iv) Public awareness campaigns v) Individual responsibility (e.g., verifying content before sharing) vi) Other (Please write your answer .....)

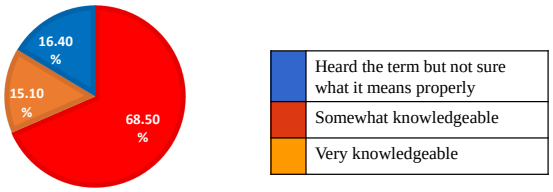
27. What strategies do you think would help protect people from the negative impacts of deepfakes? (Select all that apply)
Options: i) Social Campaigns (Education on identifying deepfakes) ii) Improved regulations on social media iii) More reliable fact-checking systems iv) Improving rules and regulation regarding AI/AI Tools v) Other (Please write your answer .....)

28. Do you believe your workplace or educational institution provides sufficient training or awareness about deepfakes?
Options: i) Yes ii) No iii) Unsure

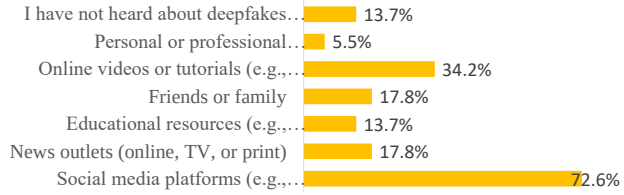
29. What would you recommend improving trust in the media in your community? (Select all that apply)
Options: i) Greater transparency from media sources ii) Public awareness campaigns iii) Technology to identify and flag deepfakes iv) Stricter laws and regulations v) Other (Please write your answer .....)

30. How do you think AI can combat deepfakes or misinformation? (Select all that apply)
Options:
    - I don't believe AI can help combat deepfakes or misinformation
    - Developing deepfake detection tools to identify manipulated content
    - Creating AI systems that flag suspicious content in real-time
    - Educating the public through AI-powered awareness campaigns
    - Enhancing media verification and fact-checking with AI
    - Using AI to track the origin and spread of misinformation
    - Collaborating with social media platforms to remove deepfake content
    - Implementing AI-based authentication systems for media
    - Other (Please write your answer .....)

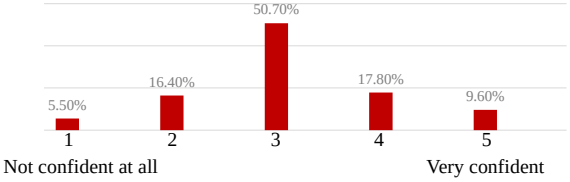
How would you describe your level of understanding of deepfakes?



Where have you learned about deepfakes?

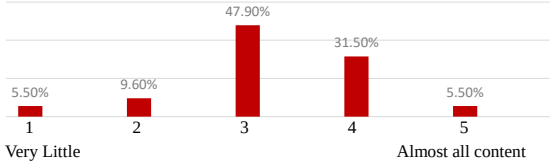


When you see a video or image online, how confident are you that it is authentic?

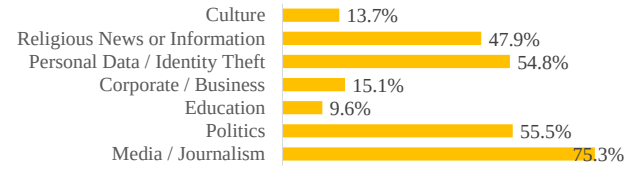


If you are unsure about the authenticity of a media piece, what do you typically do?

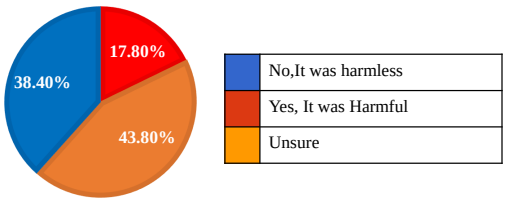
How much deepfake content do you think is available on social media?



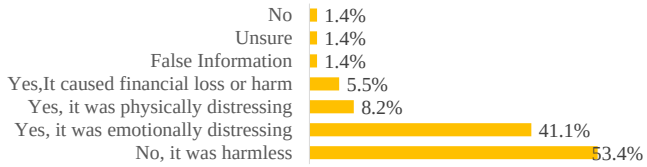
Which sectors do you think are most vulnerable to deepfake misuse? (Select all that apply)



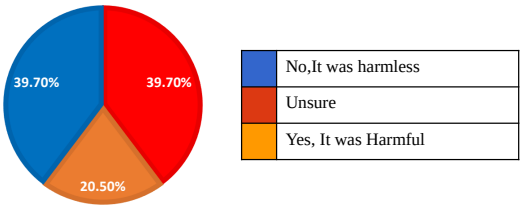
Did any deepfake content ever have a significant impact on you?



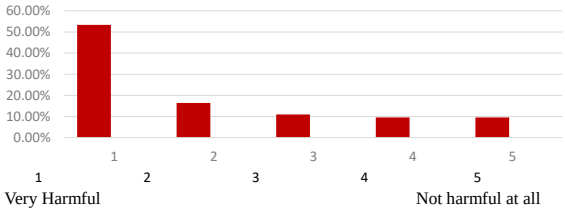
If it was harmful, what type of harm it caused?



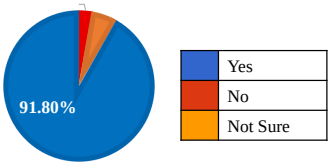
Did any deepfake content ever have a significant impact on your friends or family?



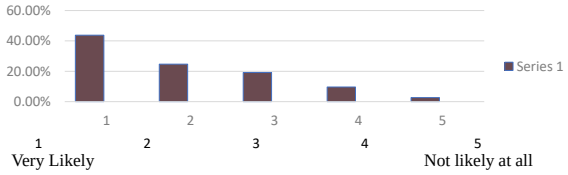
How harmful do you think deepfakes can be?



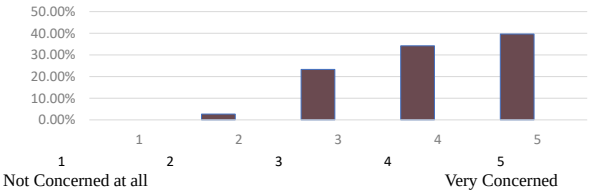
Do you think deepfakes could be used to mislead people in your community?



How likely do you think deepfakes are to influence public opinion or decisions in your community?



How concerned are you about the misuse of deepfakes?



Do you trust information from the following sources?

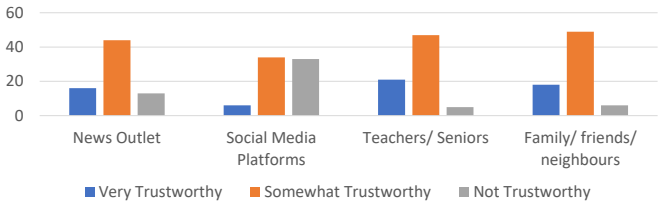
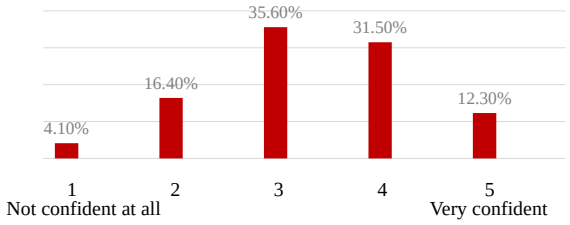
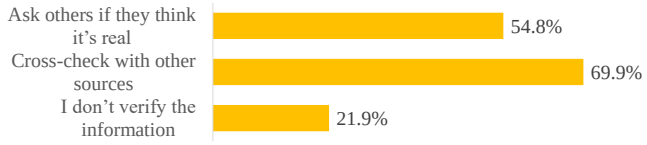


Figure 4: Data Collection: Statistics (Part 1).

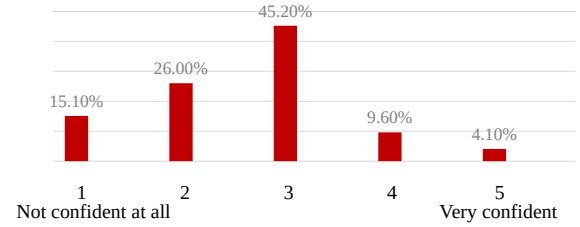
How confident are you in your ability to detect a deepfake?



How do you typically verify the information you receive online?

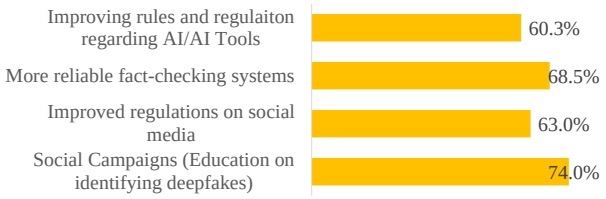


How confident are you in your friends' or family's ability to detect a deepfake?

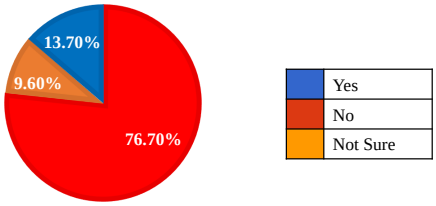


What do you think is the most effective way to combat deepfake-related issues?

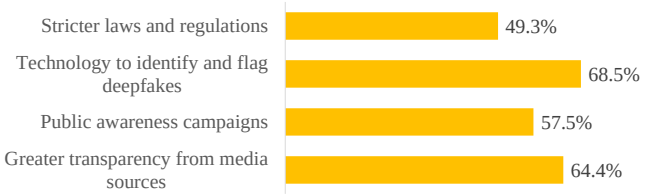
What strategies do you think would help protect people from the negative impacts of deepfakes?



Do you believe your workplace or educational institution provides sufficient training or awareness about deepfakes?



What would you recommend improving trust in the media in your community?



How do you think AI can combat deepfakes or misinformation?

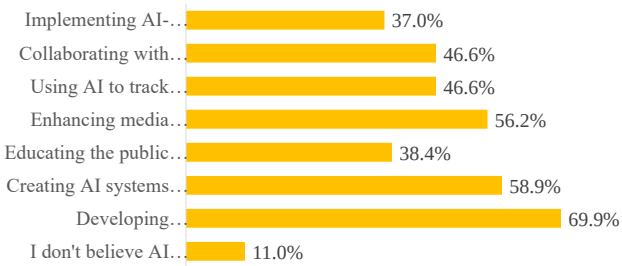


Figure 5: Data Collection: Statistics (Part 2).