Examining the Role of Deepfake Technology in Organized Fraud: Legal, Security, and Governance Challenges

Leo S.F. Lin*

Senior Lecturer, Charles Sturt University, Bathurst, Australia

Abstract: Deepfake technology has evolved astonishingly by applying artificial intelligence (AI) to inspire ultra-realistic audio and video content. Initially praised for its legitimate use cases in entertainment and education, deepfake technology has increasingly become a tool for organized fraud and other malicious purposes. This paper investigates the role of deepfake technology in enabling identity theft, financial fraud, and unlawful activities. By conducting a qualitative comparative analysis of three cases, this paper analyzes deepfakes' legal, security, and governance aspects, indicating that deepfakes have posed a massive threat at the national and global levels. Also, this research demonstrates how the current regulatory regimes cannot adequately mitigate these emerging threats. The results expose glaring deficiencies in accountability and enforcement, which are made even more glaring by the global character of the internet and the accelerative pace of technological innovation. This research provides implications of deepfake technology in organized fraud and offers policy recommendations to mitigate the threats and prevent misuse of deepfake technology in the future.

Keywords: Deepfake technology, artificial intelligence, organized fraud, identity theft, law enforcement, global governance.

1. INTRODUCTION

Artificial intelligence (AI) has revolutionized deepfake technologies by enabling the creation of hyper-realistic audio and visual content. A deepfake is a digital photo, video, or sound file of a real person edited to create an extremely realistic but false depiction of them doing or saying something they did not do or say[1].Deepfake technology creates synthetic media that people cannot differentiate from authentic content (Gil, Virgili-Gomà, López-Gil, & García, 2023). Initially, experts lauded deepfake technology for its legitimate applications in the entertainment, healthcare, and education sectors; Nevertheless, it quickly evolved from an innovative tool into a potentially dangerous technology. The capacity to distort video and audio material raises serious concerns about the authenticity and trustworthiness of media, especially as deepfakes can plausibly represent people saying or doing things that they did not, undermining the foundations of evidence and reality itself (De Ruiter, 2021; Fallis, 2021).

Organized crime groups have exploited the dark side of deepfake as another powerful means of deception. With the emergence of deepfakes, criminal organizations use them for a multitude of hidden motives: identity theft, fraud, extortion, and the spread of disinformation. The technology's ability to produce convincing impersonations raises substantial threats to individual privacy and security,and to trust in societal media and institutions (de Rancourt-Raymond & Smaili, 2023; Gupta, Chugh, Dhall, & Subramanian, 2020; Timmerman et al., 2023). For example, agents of influence have used deepfakes to target political candidates, as seen during the Russian invasion of Ukraine, where they employed the technology to erode confidence in political figures (Twomey et al., 2023).

This paper examines the legal, security, and governance dimensions of deepfake technology in the context of organized fraud. Moreover, the fast evolution of deepfake technology can potentially outpace the current legal frameworks that can hardly adapt to address the apparent challenges that deepfake technology presents (Tuysuz & Kılıç, 2023). Existing laws and traditional law enforcement approaches are often ill-suited to address the complexities of crimes associated with deepfakes, resulting in significant gaps in accountability and enforcement against organized crime (Lin, 2023). Moreover, given the international nature of the internet, governance is fragmented — the spectrum of approaches to digital content regulation is so broad that legal responses are riddled with loopholes, allowing criminals to circumvent them with ease.

Therefore, this research aims to identify legal, security, and governance challenges and propose policy recommendations by analyzing case studies. This paper aims to answer the following research questions:

*Address correspondence to these authors at the Senior Lecturer, Charles Sturt University, Bathurst, Australia; E-mail: lin.leonidas@gmail.com

[1]See Deepfake-position-statement _v2.pdf

1. How do criminal groups use deepfakes for organized fraud and deception, including their modus operandi, victim profiles, and the level of harm caused?

2. What are the legal, security, and governance gaps and challenges?

3. What measures can be implemented to mitigate these threats?

The paper is structured as follows: In the Introduction, the paper provides the problem statement and research goals. The Literature Review discusses the development of deepfake technology and how it is misused. The Methods describe the case-oriented qualitative comparative analyzes approach, the data collection, as well as analysis procedures. The Analysis section covers three case studies – Arup Case, UAE Case, and Celebrity Scams on Social Media – with comparative analyzes of the cases. The Discussion explores the legal, security, and governance challenges deepfake technology presents, along with implications and policy recommendations. This paper concludes with a summary of the outcomes and future trajectory for research.

2. LITERATURE REVIEW: DEEPFAKE TECHNOLOGY AND ITS MISUSE

Deepfakes are the product of artificial intelligence (AI) applications that merge, combine, replace, and superimpose images and video clips to create fake videos that appear authentic (Maras & Alexandrou, 2019). Advanced technologies leveraging artificial intelligence generate deepfake products, with machine learning often exploited. AI is the "computational models of human behavior and thought processes that are constructed to behave rationally and intelligently" through simulations of human behavior (Maras, 2017). A subfield of AI, machine learning allows computer systems to learn from examples, data, and experience directly… [and] carry out complex tasks by learning from data rather than through pre-programmed instructions' (Group, 2017). Such systems improve performance over time, using the lessons learned from experience to tweak their behavior in response to the performance with fresh data about the world (Maras & Alexandrou, 2019).

Deepfakes also refer to (social) media—primarily videos, audio, or images—produced by AI methods that modify real media to create a false sense of actual events or statements that never happened. The most important parts of deepfake technology are generative adversarial networks (GANs); there are two neural networks — a generator that generates fake content and a discriminator that examines whether the content is real or fake. This adversarial process continually evolves towards better quality for the generated media to the point where it becomes arduous to tell the difference between real and fake (Alhaji, Celik, & Goel, 2024).

Deepfake technology has morphed from sophisticated in-house tools requiring technical expertise to applications anyone can use. In the early days, generating deepfakes required substantial computational power and technical know-how, often possessed only by researchers and technically adept users. However, the accessibility of deepfake tools and technology has exploded with the emergence of open-source software and mobile apps that allow anyone with a smartphone or computer to generate realistic deepfakes with just a few taps and swipes. This shift has raised concerns over its misuse potential, as the technology can be deployed for both benign and nefarious purposes, which gives rise to the dual-use nature (Brandqvist, 2024; Ussenova, 2023).

Deepfake technology presents a dual-use dilemma, as there are legitimate uses and the risk of illegal exploitation. For instance, deepfake technology can be used in creative industries, such as the movie and entertainment industries, to extend the capacity of storytelling and visual effects. Deepfake technology has, for example, been used to resurrect dead actors in films or to produce content that makes audiences laugh. Then there is the upside of this technology,a separate lie weaponized for malevolent use, including identity theft, fraud, and misinformation. The risk of deepfakes being used to enable social engineering attacks, where individuals are manipulated into disclosing sensitive information, adds another layer to the ethical dimension of this technology (de Rancourt-Raymond & Smaili, 2023; Ramluckan, 2024).

Deepfake technology has been widely exploited to create pornographic videos and images since 2018. This includes the ability to produce explicit content featuring celebrities, politicians, acquaintances, or adversaries without their consent. Notable victims of deepfake videos include Aubrey Plaza, Daisy Ridley, Gal Gadot-Varsano, Natalie Portman, Scarlett Johansson, Meghan Markle, and Taylor Swift (Cuthbertson, 2018). A deepfake video on Reddit targeted former U.S. First Lady Michelle Obama,

superimposing her face onto the body of a pornographic actress with a similar facial structure (Farokhmanesh, 2018). Also, voice cloning, a sophisticated type of deepfake technology, is increasingly used in organized fraud. Scammers can generate convincing fake calls to manipulate victims into transferring money via wire transfers, gift cards, or cryptocurrency. Criminals take audio from public platforms like YouTube or TikTok to produce clones of the voices of celebrities, officials, and even ordinary people (Lin, Aslett, Mekonnen, & Zecevic, 2024). In addition to deepfake creators, other digital media manipulation tools exist and are being developed to modify user videos, voices,and images (Maras & Alexandrou, 2019).

Another question is that the implications of deepfake technology go beyond isolated individual use cases; they threaten society as a whole. With the globalized digital world, emergening technologies such as deepfakes have created opportunities for organized crime groups to pose security threats to countries (Lin, 2022). With the rise of deepfakes, distinguishing between reality and fiction is becoming increasingly complex, resulting in what is commonly known as "truth decay" (Chesney & Citron, 2018; Kozemczak, 2019). This loss of trust in (social) media can carry major ramifications, especially in the political sphere, where deepfakes are designed to delegitimize political actors or manipulate elections (Diakopoulos & Johnson, 2021). The key challenge is to find common ground between the functionality of deepfake and their implications for harm and to provide relevant regulatory and governance standards to address this challenge.

3. METHODS

This paper conducted a case-oriented Qualitative Comparative Analysis (QCA)to address the research questions. This method entails a detailed comparison of a small number of cases to identify patterns of conditions that are necessary for a specific outcome (Ragin, 2000). Since this study aims to analyze three deepfake cases better to understand the interaction between organized fraud and deepfake technology, this small number of cases is a suitable strategy for the comparative method (Della Porta, 2008). The case-oriented strategy analyzes a small number of cases as interpretive wholes, aiming to understand complex unities rather than relationships between variables (Ragin, 2000). This approach facilitates a snapshot mapping of the real-world implications of the multi-dimensional threat deepfakes pose in organized fraud. Additionally, thematic analysis was conducted for each case to identify key themes in organized fraud, focusing on coding key patterns related to the exploitation of deepfakes by organized crime. Thematic analysis is a qualitative method used to identify, analyze, and interpret patterns of meaning ("themes") within data (Clarke & Braun, 2017).

This paper utilized secondary data, including academic literature, government documents, non-governmental organization documents, websites, news articles, and open-source information regarding case studies. Case studies illustrate real-world applications of deepfake technology in organized crime, such as identity theft and fraud (Westerlund, 2019). By triangulating data from multiple sources, the findings are better validated, allowing for a more comprehensive analysis of the issues.

There are a few limitations to this study that deserve mention. First, the lack of primary data, such as interviews with victims or law enforcement officers, could challenge capturing first-hand perspectives on the real-world impact of deepfake-enabled organized fraud. Moreover, the rapid evolutionary pace of deepfake technology limits the ability to keep abreast of developments and emerging threats for future studies—all too often, findings have become dated in the short space from the time of investigation to publication. Third, given limited time and resources, the present study analyzes only three well-known cases for conclusions. Collecting a larger number of cases for comprehensive analysis might be beneficial, increasing the results' generalizability and robustness. The challenges and constraints of each method call for ongoing and adaptable research to understand the changing dynamics of deepfake technology and its use in organized crime.

4. ANALYSIS OF DEEPFAKE TECHNOLOGY IN ORGANIZED FRAUD: CASE STUDIES

Deepfake technology has emerged as a double-edged threat to security because it facilitates deceptive activities and enables social engineering attacks. Deepfake technology becoming more accessible and easier to use creates serious dangers that can cause destructive impacts on human societies. This section provides a comparative analysis of three well-known cases to understand how deepfake technology has been exploited, highlighting specific case studies and their implications.

CASES

Case 1: The Arup Case (Deepfake Videos)

Incident

The Arup case in Hong Kong involved a sophisticated cyber scam utilizing deepfake technology, making it one of the world's most significant cases. The scam occurred in January 2024, when Arup's Hong Kong branch employee was deceived into transferring HK$200 million (US$25.6 million) to fraudsters. The perpetrators employed advanced deception using a digitally cloned version of the company's Chief Financial Officer (CFO) to issue false instructions during a video conference. The scam was reported to the Hong Kong police, marking the first instance of a deepfake video call being used for fraud[2].

Victim

The main victim of the scam was Arup, a UK-based multinational design and engineering consultancy featured in the construction of landmark projects such as the Sydney Opera House and the Bird's Nest stadium for the 2008 Beijing Olympic Games. The direct victim at Arup was a staff member in its Hong Kong office's finance department. An employee was deceived into believing that they were communicating with the CFO and senior staff from the UK headquarters, leading to the company suffering a major monetary loss. The company's East Asia chair, Andy Lee, resigned later, but the company did not mention any direct connection between his departure and the incident[3].

Modus Operandi

The scammers had a multi-layered approach to their operations. It started with a phishing email to the finance employee, disguised as one from Arup's CFO, asking to operate a sensitive transaction. The criminals staged a video conference using deepfake technology to make the request seem legitimate. On the call, the hackers digitally cloned the CFO, among other fake coworkers, who convincingly said and did things, using artificial intelligence to replicate their voices, faces, and mannerisms. Thinking the interaction was legitimate, the employee made 15 financial transfers to five Hong Kong bank accounts. The scam came to light only after the employee sought follow-up clarification from Arup's headquarters.

Financial Loss

The Arup deepfake scam resulted in a financial loss of HK$200 million (US$25.6 million). While the extent of the loss was major, Arup's global finances and business practices were reportedly unaffected. It reassured that its systems remained secure, emphasizing that the scam exposed vulnerabilities in human decision-making.

Law Enforcement Investigations

The police in Hong Kong classified the case as "obtaining property by deception" and began an ongoing investigation[4].Hong Kong police said they had made six arrests concerning such scams[5].Authorities noted that the threat from deepfakes to corporate fraud was growing, with experts cautioning that scam attempts could become more frequent and sophisticated. Hong Kong police recommended that companies tighten their cyber security measures and reinforce employees' awareness of these scams. Rob Greig, the global Chief Information Officer for Arup, said cyber attacks were becoming more sophisticated and added that the growing tactics of cybercriminals mean companies should not be complacent[6].

(2). Case 2: The UAE Case (Voice Cloning)

Incident

In early 2020, cybercriminals in the United Arab Emirates executed a complex financial heist using artificial intelligence voice cloning technology. Fraudsters used voice cloning technology to mimic a company director and fool a bank manager into approving multiple financial transactions. This high-tech scam became one of the largest AI-driven voice deepfake fraud cases when perpetrators stole approximately US$35 million[7].

Victim

The primary victim of this cyber heist was a bank in the UAE. The fraud targeted a branch manager who was manipulated into believing that he was acting on the legitimate instructions of a company director. The

[2]See https://oecd.ai/en/incidents/83641
[3]See https://www.indianweb2.com/2024/05/in-one-of-worlds-biggest-known-deepfake.html
[4]See https://www.theguardian.com/technology/article/ 2024/may/17/uk-engineering- arup- deepfake-scam-hong-kong-ai-video
[5]See https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html
[6]See https://www.theguardian.com/technology/article/2024/may/17/uk-engineering-arup-deepfake-scam-hong-kong-ai-video
[7]See https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/

director's voice was familiar to the manager, which increased the scam's credibility. The company associated with the director, as well as the bank that processed the transfers, were both affected by the incident. Given the transnational nature of the stolen funds, the financial institutions receiving the transferred amounts also became indirect victims.

Modus Operandi

The scammers' modus operandi depended on AI-based voice cloning technology known as "deep voice" technology. The cybercriminals generated a fake voice of the company director from speech samples that they probably secured through public domains or intercepted communications. The scammers used the branch manager to validate financial transfers by employing the cloned voice to instruct the manager during the alleged company acquisition. The fraud gained additional legitimacy when emails from an alleged lawyer, Martin Zelner, arrived with instructions and documentation supporting the fraudulent request. The bank manager processed the transfers due to the familiar voice and supporting documents but remained oblivious to the fraudulent nature of the request[8].

Financial Loss

The total financial loss for this case constituted approximately $35 million. The stolen money was transferred through bank accounts in numerous countries, and $400,000 ended up in Centennial Bank accounts in the United States. Investigators suspected the leftover money was scattered through numerous international bank accounts under multiple aliases, making tracing and recovery difficult[9].

Law Enforcement Investigations

The United Arab Emirates authorities initiated an investigation after the heist and requested help from the United States to locate the $400,000 which had been deposited into US bank accounts. The investigation concluded that no fewer than 17 suspects were considered to be criminal activity participants. The international scope of the scam and advanced AI-based voice cloning technology presented substantial obstacles for law enforcement officials. The Dubai Public Prosecution Office directed the investigation while working with US authorities to follow and locate the offenders. According to this case, financial fraud using deepfake technologies has become a significant threat, revealing weaknesses in traditional banking authentication methods.

(3). Case 3: Celebrity Scams on Social Media (Deepfake Content)

Incident

Celebrity deepfake scams on social media like Facebook and Instagram have emerged as a sophisticated and pervasive cybercrime. Scammers use generative artificial intelligence (AI) to craft ultra-realistic images of famous individuals. By producing counterfeit video, audio, and image content that looks like it features celebrities, scammers seek to manipulate public trust to trap unsuspecting victims in fraudulent activities.There has been an alarming increase in AI abuse cases in recent years which often target well-known celebrities including Taylor Swift. Explicit AI-generated images of celebrities caused widespread backlash and raised concerns over how generative AI tools are misused[10]. The White House and U.S. Congress have turned their focus on tackling the rapid spread of deepfake technology and its dangerous consequences after this and other incidents[11]. Another incident involved an Australian citizen who lost $80,000 in cryptocurrency after seeing a deepfake Elon Musk video interview on social media, clicking the link, and registering his details online[12].

Victim

Social media users with strong celebrity interests and celebrities represent the main groups targeted by these scams. The Australian Competition and Consumer Commission's Scam Watch reports that older adults commonly face financial scams. Older adults aged 65 years or older experience higher rates of financial losses from online scams than other demographics[13]. Fake celebrity meet-and-greet offers to lure people into signing for merchandise purchases and exclusive investment deals or giveaways.

[8]See https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/

[9]See https://www.newser.com/story/312228/scammers-clone-execs-voice-steal-35m-from-bank.html

[10]See https://www.usatoday.com/story/entertainment/celebrities/2024/01/25/taylor-swift-artificial-intelligence-images/72350439007/

[11]See https://www.congress.gov/118/meeting/house/116778/documents/HHRG-118-JU03-20240202-SD002.pdf

[12]See https://www.theguardian.com/australia-news/2024/mar/01/scams-promoted-in-fake-news-articles-and-deepfake-videos-cost-australians-more-than-8m-last-year

[13]See https://www.accc.gov.au/media-release/criminals-targeting-victims-of-previous-scams-promising-financial-recovery#:~:text=Australians%20aged%2065%20and%20older,to%20get%20their%20money%20back

Scammers use emotional manipulation and fake celebrity endorsements to create convincing fraudulent schemes.

Modus Operandi

Scammers deploy advanced AI technologies to generate deepfake content that convincingly mimics celebrities' facial features, voice, and behavioral patterns. Scammers distribute deepfake content throughout social media platforms and initiate direct contact with victims in comments and messages or through targeted advertising. Scammers deploy the "celeb-bait" scam, which involves impersonating celebrities who promise exclusive opportunities to potential victims[14].When claiming these offers, scammers typically ask victims to provide personal details, follow phishing links, or process payments. Some fraudulent schemes advertise opportunities to meet famous people and exclusive merchandise or financial guidance. Scammers use famous personalities to advertise fake products, including dietary supplements, weight-loss gummies, and seemingly real cryptocurrency websites or mobile phone applications.

Financial Loss

Celebrity deepfake scams have caused significant financial damage. Scamwatch data reveals that Australians 65 years old and older experienced over 72,500 reported scams in 2023, resulting in financial losses exceeding $120 million[15].Scammers create losses beyond just stealing money because they can gather victims' data and use it for future fraudulent activities.One report found that deepfake incidents increased by 700% in fin tech in 2023[16]. Globally, fraud losses due to AI deepfakes are expected to triple to USD 40 billion by 2027[17].

Law Enforcement Investigations

Law enforcement agencies, social media companies, and financial institutions have caught on to the increasing threat of celebrity deepfake scams. Moreover, to tackle the problem, Meta (the parent company of Facebook, formerly known as Facebook Inc.) is joining forces with the Australian Financial Crimes Exchange (AFCX) to launch the Fraud Intelligence Reciprocal Exchange (FIRE). The initiative enables banks like the Commonwealth Bank of Australia and ANZ to pass on insights about scam trends to Meta[18]. The partnership allows for improved tracking, reporting, and removal of fraudulent content. Meta has already deleted over 9,000 spam accounts and 8,000 AI-generated scams from Facebook and Instagram as part of its enforcement[19]. There have also been heightened efforts to inform the public and warn them to be on the lookout for scams, including campaigns encouraging social media users to check the authenticity of celebrity accounts and be wary of deals that seem too good to be true.

Comparative Analysis of the Three Cases

The use of deepfake technology in cybercrime has introduced new dimensions of risk for individuals, financial institutions, and corporations. The cases of Arup, the UAE voice cloning scam, and celebrity deepfake scams on Facebook illustrate the diverse methods and impacts of these crimes. This section discusses the similarities and differences in these cases, focusing on incident characteristics, victims, modus operandi, financial losses, and law enforcement responses.

(1). Incident Characteristics

Similarities: All three cases involve the strategic use of generative AI to manipulate human perception. By producing hyper-realistic audio or audio-visual material, the attackers duped victims into thinking that they were dealing with legitimate agents. The Arup case and the UAE case both relied on deepfake technology to recreate top officials of the companies, and the celebrity scam case relied on the likenesses of popular public figures to elicit trust from the general public. These cases illustrate how deepfake technology can be monetized through direct interface with employees (as in Arup and UAE) or by taking aesthetic advantage of a mass audience (as in celebrity scams).

Differences: The nature of the deception differed significantly. The Arup example used an utterly immersive video conference with AI-generated

[14]See https://ia.acs.org.au/article/2024/meta-and-australian-banks-tackle-ai-celeb-bait-scams.html
[15]See https://nationalseniors.com.au/news/latest-news/share-a-story-stop-a-scam
[16]See https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html
[17]See https://www.forbes.com/sites/chriswestfall/2024/11/29/ai-deepfakes-of-elon-musk-on-the-rise-causing-billions-in-fraud-losses/
[18]See https://www.mi-3.com.au/03-10-2024/meta-and-afcx-team-pilot-scam-reporting-channel-australian-bankshttps://www.mi-3.com.au/03-10-2024/meta-and-afcx-team-pilot-scam-reporting-channel-australian-banks
[19]See https://www.theguardian.com/technology/2024/oct/02/more-than-9000-scam-facebook-pages-deleted-after-australians-lose-millions-to-celebrity-deepfakes

likenesses of executives from the company. The UAE example was purely voice cloning-based, without any visual aspect, and the celebrity scam employed pre-recorded advocates through video, image, or text. While the Arup and UAE cases targeted internal organization actors—its employees and bank managers with access to enormous financial resources, celebrity scams targeted the general public, specifically vulnerable population segments such as older adults.

Victims

Similarities: Each case involved unsuspecting victims being misled by credible impersonations of trusted figures. The human factor played a major role, with criminals taking advantage of the confidence in well-known voices, images, or versions of brands. The Arup case was aimed at a finance employee, the UAE case tricked a bank manager, and the celebrity scams focused on the general public. In each case, the attackers manipulated basic social engineering principles such as authority and familiarity to convince victims to take action.

Differences: The primary victims differ in scope and status. The Arup and UAE examples feature direct financial losses to specific organizations (Arup and a UAE bank). In contrast, celebrity scams generally had a large number of victims in the general public, with older adults disproportionately affected. The financial repercussions, too, are different. While Arup and the UAE bank lost significant sums of money, fakecelebrity scams have victims worldwide.

Modus Operandi

*Similarities:*Deepfake tech (voice, video, image) can emulate aspects of human behavior to deceive perception across all three cases. The aggressors spent much time imitating the conduct, look, and style of correspondence of huge names (Arup and UAE) or famous people (big-name tricks). Finally, all three cases had a multi-step approach to the scam, e.g., using emails followed by video calls followed by text-based interactions to reinforce the credibility of the scam.

Differences: The Arup case is notable for involving real-time manipulation of multi-modal (video and audio) deepfakes during a live video conference. The UAE case was based only on voice cloning, which is simpler and less demanding in technology than real-time synthesis of a person's face in a video frame. In contrast, celebrity scams exploit pre-recorded or pre-generated content to disseminate phoney endorsements over social media. However, while the above scams involving Arup and UAE depended on "live" engagement with the targets, the celebrity scam mechanism was more passive, using fake content across different social media channels to entice a large number of victims.

(4). Financial Losses

Similarities: All cases resulted in substantial financial losses. Each crime relied on the manipulation of human error and misplaced trust. The monetary scale of loss highlights the high potential impact of deepfake-enabled fraud. In all three instances, multiple transactions were involved in scam losses which indicates deliberate fund distribution across various accounts or targeting of multiple victims by the fraudsters to avoid detection..

Differences: The Arup and UAE cases related to a massive, collective transfer of funds totalling HK$200 million (US$25.6 million) and US$35 million, respectively, whereas celebrity scams resulted in large but not just individual losses. The transnational character of the funds' journey was particularly apparent in the UAE case, where stolen funds flowed into many bank accounts worldwide. While the Arup case's financial losses were limited to movements within the Hong Kong banking system, celebrity scam losses stretched across the globe, hitting several individuals (e.g. Including celebrities and citizens) instead of one organization.

(5). Law Enforcement Investigations

Similarities: Law enforcement responses in all three cases underscore the reactive nature of combating crime in the face of emerging technologies. Hong Kong, the UAE, and other jurisdictions affected by deepfake fraudulent schemes worked to recover losses, identify fraudsters, and warn the general public and industry participants about the risks of deepfake fraud. All three cases involved investigative challenges, especially the role of transnational financial transactions and jurisdictional constraints.

Differences: The scope of investigations and the involvement of third-party institutions differ. The Arup case was investigated in Hong Kong under the charge of "obtaining property by deception." Investigators

Table 1: Comparative Analysis of AI-Driven Cybercrime Cases – Deepfake Videos, Voice Cloning, and Celebrity Deepfake Scams

Theme \Case	Deepfake Videos (Arup Case)	Voice Cloning (UAE Case)	Celebrity Deepfake Scams (Facebook/Instagram)
Nature of the Scam	Use of deepfake video to impersonate an executive in video calls to deceive employees or stakeholders.	Use AI-generated voice to impersonate a company's director in phone calls to facilitate fraudulent transfers.	Use AI-generated deepfake videos or images of celebrities to promote fake investment schemes on social media.
Primary Technology Used	Deepfake video (visual AI synthesis)	Voice cloning (audio AI synthesis)	Deepfake video or image (visual AI synthesis)
Method of Deception	Visual impersonation of senior executives in live or pre-recorded video calls.	Audio impersonation of a high-level executive's voice in real-time phone calls.	Use celebrity images/videos to endorse fraudulent schemes, often appearing as "ads" on social media.
Target Victim(s)	Corporate employees, company stakeholders, and financial controllers.	Bank employees or financial staff responsible for authorizing large fund transfers.	The general public, mainly social media users, is susceptible to "get-rich-quick" schemes.
Perpetrator Strategy	Convince employees or financial controllers to transfer funds or approve financial transactions.	Directly instruct bank employees to process high-value transfers to criminal-controlled accounts.	Lure victims into investing in fraudulent schemes with the promise of high returns, often using celebrity trust as bait.
Main Objective	Financial fraud (unauthorized transfers of company funds).	Financial fraud (large-sum wire transfers to fraudulent accounts).	Financial fraud (inducing mass online payments or investments in fake opportunities).
Level of Sophistication	High – Requires live deepfake video synchronization and realistic audio-visual coordination.	High – Requires advanced real-time voice synthesis and impersonation.	Medium – Uses pre-made video/image deepfakes, often distributed via ads or posts on social media platforms.
Impact on Victims	Loss of corporate funds, damage to corporate reputation, and erosion of trust in video conferencing tools.	Loss of funds from bank accounts, breach of trust, and operational disruption.	Financial losses from fraudulent investments, public mistrust of celebrity endorsements, and reputational damage for social media platforms.
Scale of Losses	Large-scale corporate fraud with high-value transfers.	Multi-million dollar thefts (e.g., $35 million in the UAE case).	Cumulative losses from numerous victims worldwide, often in smaller amounts but affecting large numbers of people.
Countermeasures	Use of video authentication tools, multi-factor verification, and employee awareness training.	Voice recognition security, stricter protocols for financial transfers, and staff training on voice-cloning risks.	Takedown scam campaigns by social media platforms, public awareness campaigns, and enhanced AI detection tools for deepfakes.

Made by the author.

working on the UAE case had to coordinate between UAE law enforcement, the Dubai Public Prosecution Office, and US authorities to trace funds associated with US bank accounts. In contrast to the Arup and UAE cases, where corporations lost money, the celebrity scam resulted in structural reforms to online safety and regulatory responses.

5. DISCUSSION

Legal, Security and Governance Perspectives

National-Level Dimensions in Combating Crime

At the national level, a major challenge is the absence of comprehensive legislation addressing the production and misuse of AI-generated content. Many legal frameworks have yet to evolve to address the issue adequately, and gaps in accountability remain even where they are well established. The evolving landscape of AI-based impersonation and digital manipulation cannot be neatly contained within existing laws. A good instance to support this is the UAE case, where AI voice cloning was utilized to execute a significant financial fraud. The crime was serious, but current cybercrime laws had no explicit offence provisions to respond to deepfake-related crimes. Victims of such crimes often struggle to navigate evolving legal grey areas, with few avenues available to pursue justice or compensation. The unclear nature of existing laws diminishes protection for victims while enabling offenders to take advantage of legal loopholes.

A major problem exists in inconsistent legal penalties and enforcement for AI-enabled crimes

between different jurisdictions. The lack of uniform national laws leads to an uncoordinated enforcement system that allows criminals to avoid prosecution by taking advantage of legal loopholes in certain regions. This lack of consistency in penalties and mechanisms of enforcement also allows "jurisdiction shopping or forum shopping," allowing offenders to shop around for the jurisdiction with the least restrictive laws to commit cross-border deepfake scams (Rawat, 2021). The growth of celebrity deepfake scams on social media platforms such as Facebook, for example, demonstrates how scammers take advantage of these legal inconsistencies to target individuals across jurisdictions with varying levels of protections. While some countries have imposed strict penalties on cybercrime-related offences — their law in many cases — some have not aligned their legal frameworks with the new threat landscape. Moreover,*inconsistent enforcement standards hinder effective compliance and impede efforts to coordinate a global response to AI-driven crime*.

Technical and investigative challenges further complicate the fight against deepfake-related crimes. Unlike other cybercrimes, deepfake crimes are challenging to prevent and solve, requiring advanced technical tools and forensic experience to identify, track, and assign the content to its source. Often, law enforcement has a considerable struggle to access the technical tools needed to combat crimes of this nature. The challenges are highlighted by the Arup case in Hong Kong. In this case, investigators struggled to track the criminals behind a HK$200 million deepfake scheme due to the complexity of the machine-generated material. A significant deepfake technological breakthrough enabled criminals to execute attacks while preventing investigators from recognizing perpetrators or finding evidence. Many law enforcement agencies lack the technical capacity to combat increasingly sophisticated AI-enabled criminal tactics.

Regulatory overlaps and compliance issues add additional complexity to national-level governance of deepfake technology. AI-specific regulations, however, frequently come into contact with already established legal paradigms revolving around privacy, data protection, intellectual property rights, and ethical utilization of AI. Overlap regarding regulatory requirements creates uncertainty, suppressing innovation as companies struggle to clarify a confusing compliance landscape. In addition to privacy and data protection laws, the companies developing AI systems must comply with AI ethics guidelines and intellectual property rules that can impose competing obligations. The Arup case underlines the value of clarity in regulatory oversight. In this case, cybercriminals did not take advantage of system vulnerabilities but rather the holes in human decision-making, revealing how far the current compliance safeguards fall short in addressing human-centered security weaknesses. When regulatory measures are applied effectively, they can promote responsible AI development and make financial systems more resilient against criminal exploitation.

Global-Level Dimensions in Combating Crime

Misusing deepfake technologies presents a substantial worldwide risk to security measures and democratic systems while eroding public trust. Findings from the Australian Strategic Policy Institute (ASPI) and the European Union Agency for Law Enforcement Cooperation (Europol) show that deepfakes are used for fraud and propaganda to spread disinformation. Europol's report, *Facing reality? Law Enforcement and the Challenge of Deepfakes* warns that deepfakes may soon become a "staple tool" for organized crime[20]. The Arup and UAE cases further corroborate this foreshadowing — evidence of how AI-enabled fraud can now undermine financial systems on an unprecedented scale. Beyond financial crime, deepfakes can thwart elections and warp public sentiment, undermining confidence in democracy and governance. As confidence in information declines, so does the quality of public discourse, which means that the danger posed by deepfakes is much more than the sum of isolated criminal acts.

A significant barrier remains because there is no unified international regulatory system governing AI ethics and accountability alongside safety measures. Such jurisdictional discrepancies create loopholes that criminals can exploit (e.g., the proliferation of celebrity deepfake scams on social media sites). These platforms are global but subject to different jurisdictions' fragmented legal standards. Furthermore, without international legal standards aligning these, it is pretty much impossible to set clear rules of engagement for investigation and prosecution. The dearth of standardization and uniformity on this issue hinders the establishment of global best practices for managing deepfake-related risks. It exposes social

[20]See https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes

media users and financial institutions to potential exploitation.

International jurisdictional issues and cross-border cooperation requirements make solving AI-driven cybercrime more difficult. Perpetrators often commit crimes in multiple countries. Jurisdictional complexities were on full display in the case of the voice cloning heist that the UAE investigated, where the large fund was traced to US bank accounts, requiring collaboration with US authorities. The investigation faced obstacles from conflicting jurisdictional authority and procedural requirements. These examples highlight the necessity for coordinated efforts while cross-border legal collaboration remains slow and obstructed by bureaucratic procedures and inconsistent legal systems. The lack of rules and protocols for jurisdictional alignment creates significant obstacles for international authorities who wish to investigate and prosecute deepfake-related crimes.

International efforts to fight AI-driven cybercrime face reduced effectiveness because of insufficient data-sharing mechanisms. Data protection laws create barriers for law enforcement agencies that prevent them from accessing vital information quickly. The Arup investigation showed that delays in cross-border data sharing prevented authorities from successfully catching the criminals. Privacy regulations in certain nations place more importance on safeguarding personal information than on sharing vital data needed to investigate international criminal activities. Law enforcement agencies face siloed operations without real-time data-sharing capabilities, allowing perpetrators to exploit investigation delays. This will necessitate creating shared datapolicies between countries that respect the loss of privacy and the desire and need for rapid data sharing. Establishing such frameworks would allow investigations to respond more quickly and help bolster global efforts to bring AI-fueled criminals to justice.

Implications and Recommendations

The Discussion demonstrates deepfake technology's diverse obstacles and the immediate need for preventive action. The development of deepfake technology is outpacing the potential for harmful use. Combating this threat requires tailored recommendations across policy, technology, and capacity-building endeavours. This paper has identified implications and offered suggestions to help counteract the risks associated with deepfakes.

- *Establishing Global Standards for Deepfake Governance:* The proliferation of deepfakes necessitates the establishment of harmonized global standards for their governance. Such standards should include codes of ethics, security protocols, and governance standards that all nations can implement. Current policies tend to be inconsistent and poorly coordinated internationally, leaving regulatory land vacant by criminals. A common framework sets the tone for governments, corporations, and individuals to align on AI content generation expectations.

- *Defining the Illegal Use of Deepfake Technolog:* The misuse of deepfake technology demands clear legal definitions and assigned responsibilities. Current legislation fails to define deepfakes properly, leaving criminal charges for malicious use ambiguous. Targeted punishment of fraudulent misuse of deepfake technology for defamation and disinformation requires specific legal definitions to fill current legal gaps.

- *Enhancing Law Enforcement Training in Digital Forensics and AI-DrivenCrime:* Law enforcement agencies must enhance their capacity to address deepfake-related crimes. Investments in specialized training on digital forensics and AI-related crime detection should support this initiative. Developing knowledge in these areas will train law enforcement to be more knowledgeable and prepared to identify and investigate deepfake-related crimes and the legal implications of such crimes. Increased technical knowledge amongst law enforcement will fill expertise gaps and allow for more enforcement and prosecution of offenders.

- *Raising Public Awareness to Mitigate Susceptibility to Deepfake Scams:* Public awareness programs are critical tools to reduce individual vulnerability to deepfake scams. The campaigns teach people to detect deepfakes and develop digital literacy and critical thinking abilities. The best strategy to reduce misinformation and fraud involves public education. Through an all-behavioral approach, people learn to recognize deepfake risks, which improves their ability to spot harmful content and reduces their chances of mistreatment.

- *Promoting Public-Private Partnerships for Deepfake Detection Tools:* Investment in

sophisticated deepfake detection tools is critical for countering the spread of manipulated content. Doing so would have a key role in public-private partnerships in developing advanced detection algorithms and technologies capable of real-time detection of deepfakes. Combining the tools and knowledge of both sectors results in more effective detection systems. Once created, these systems can be plugged into social media platforms, news outlets, and other places where deepfakes might appear.

- *Balancing Privacy, Freedom of Expression, and Security:* AI technologies like deepfakes require regulation, which presents technological and regulatory challenges because they involve ethical and human rights issues concerning privacy, freedom of expression and security. It is difficult but necessary to strike a balance between these conflicting interests. Although technologies waiting on artificial intelligence hover on security and public safety, they could also be exploited for surveillance and repression, especially by authoritarian regimes. Lacking sufficient safeguards, the very tools leveraged to identify or surveil deepfakes could be turned against dissenters and erode civil liberties in the process.

- *Mitigating Risks of Surveillance and Authoritarian Abuse:* The use of AI for unethical surveillance poses a direct threat to privacy rights and civil liberties. Populations could be tracked and monitored, and dissent was repressed by implementing government AI-driven technologies. To counter the abuse, human-centric ethical guidelines must put human concerns first and prevent their misuse. We can avoid overreach and surveillance by ensuring the responsible development and use of these tools.

6. CONCLUSION

The Arup, UAE, and celebrity deepfake scams demonstrate the broad applicability of deepfake technology in cybercrime, with unique differences in methods, victims, and financial consequences. The Arup and the UAE cases show how criminals can leverage trusted internal actors to forfeit the financial security of an organization. Conversely, celebrity scams are all about manipulating public trust on a large scale. The analysis also identifies significant differences in modus operandi, with live, multi-modal deepfake technology being the most sophisticated observed to date, as demonstrated in the Arup case. The law enforcement responses underscore the ongoing difficulty of jurisdictional boundaries, technological sophistication, and the need for international cooperation to counter transnational financial fraud. These cases illustrate how rising threats from deepfakes necessitate more vigorous work around governance and cybersecurity and public awareness campaigns to combat these threats.

This rapid evolution of deepfake technology poses significant challenges and risks in organized fraud, necessitating urgent attention from legal, security, and governance perspectives. This paper has shown that criminal organizations can weaponize deepfakes, facilitating identity theft, financial fraud, and disinformation, thus eroding trust in digital media and institutions. The study examined how these threats emerge through practical examples and identified the shortcomings of existing legal frameworks that frequently fail to keep pace with technology's rapid development.

The fight against deepfake technology abuse demands the creation of precise legal definitions and responsibilities which define acceptable usage boundaries. Expanding law enforcement proficiency in digital forensics and AI-related criminal investigations will boost detection capabilities and prosecution effectiveness. Establishing standardized international regulations for deepfake technology will create a unified framework to address these issues across different legal systems effectively. Public awareness programs must be established to teach people how to recognize and critically assess deepfake content to reduce their susceptibility to fraudulent schemes and false information. Society should receive education on deepfake technology basics and training to develop critical thinking skills to prevent potential dangers.

Future research directions should focus on several key areas to address the challenges posed by deepfake technology in organized fraud. Both longitudinal studies that closely follow the changing effects of deepfakes over time and comparative analyses of the legal framework across jurisdictions could illuminate potential best practices for a more harmonized international framework. Research into advanced technological detection methods must accompany studies of psychological and social effects to understand deepfakes' impact on public trust and

behaviour. Evaluating current policies and awareness campaigns can guide the creation of targeted strategies to advance digital literacy and critical thinking skills. Combining insights from legal studies with technological expertise and psychological and sociological research helps understand deepfake challenges and create comprehensive solutions.

CONFLICT OF INTEREST STATEMENT

The author declares no known financial or personal relationships that could have influenced the work reported in this paper.

REFERENCES

Alhaji, H. S., Celik, Y., & Goel, S. (2024). An Approach to Deepfake Video Detection Based on ACO-PSO Features and Deep Learning. *Electronics, 13*(12), 2398. https://doi.org/10.3390/electronics13122398

Brandqvist, J. (2024). The cybersecurity threat of deepfake.

Chesney, R., & Citron, D. K. (2018). 21st century-style truth decay: Deep fakes and the challenge for privacy, free expression, and national security. *Md. L. Rev., 78*, 882. https://doi.org/10.2139/ssrn.3213954

Clarke, V., & Braun, V. (2017). Thematic analysis. *The journal of positive psychology, 12*(3), 297-298. https://doi.org/10.1080/17439760.2016.1262613

Cuthbertson, A. (2018). What is Deepfake porn? AI brings face-swapping to a disturbing new level. *Newsweek, Feb*.

de Rancourt-Raymond, A., & Smaili, N. (2023). The unethical use of deepfakes. *Journal of Financial Crime, 30*(4), 1066-1077. https://doi.org/10.1108/JFC-04-2022-0090

De Ruiter, A. (2021). The distinct wrong of deepfakes. *Philosophy & Technology, 34*(4), 1311-1332. https://doi.org/10.1007/s13347-021-00459-2

Della Porta, D. (2008). 11 Comparative analysis: case-oriented versus variable-oriented research. *Approaches and methodologies in the social sciences*, 198. https://doi.org/10.1017/CBO9780511801938.012

Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New media & society, 23*(7), 2072-2098. https://doi.org/10.1177/1461444820925811

Fallis, D. (2021). The epistemic threat of deepfakes. *Philosophy & Technology, 34*(4), 623-643. https://doi.org/10.1007/s13347-020-00419-2

Farokhmanesh, M. (2018). Deepfakes Are Disappearing from Parts of the Web, But They're Not Going Away. *The Verge*.

Gil, R., Virgili-Gomà, J., López-Gil, J.-M., & García, R. (2023). Deepfakes: evolution and trends. *Soft Computing, 27*(16), 11295-11318. https://doi.org/10.1007/s00500-023-08605-y

Group, R. S. W. (2017). Machine learning: The power and promise of computers that learn by example. *Technical report*.

Gupta, P., Chugh, K., Dhall, A., & Subramanian, R. (2020). *The eyes know it: Fakeet-an eye-tracking database to understand deepfake perception.* Paper presented at the Proceedings of the 2020 international conference on multimodal interaction. https://doi.org/10.1145/3382507.3418857

Kozemczak, V. (2019). Deep Fakes: Preserving Truth & Human Rights in an Era of Truth Decay.

Lin, L. S. (2022). Globalization of Crime and Digitized Societies: A Recent Survey *Evolution of Digitized Societies Through Advanced Technologies* (pp. 153-163): Springer. https://doi.org/10.1007/978-981-19-2984-7_13

Lin, L. S. (2023). Theories and Analytical Framework *Asian Organized Crime and the Anglosphere* (pp. 19-42): Springer. https://doi.org/10.1007/978-3-031-41482-4_2

Lin, L. S., Aslett, D., Mekonnen, G., & Zecevic, M. (2024). The dangers of voice cloning and how to combat it.

Maras, M.-H. (2017). Social media platforms: Targeting the 'found space'of terrorists. *Journal of Internet Law, 21*(2), 3-9. https://doi.org/10.1177/1365712718807226

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

Ragin, C. C. (2000). *Fuzzy-set social science*: University of Chicago Press.

Ramluckan, T. (2024). *Deepfakes: The legal implications.* Paper presented at the International Conference on Cyber Warfare and Security. https://doi.org/10.34190/iccws.19.1.2099

Rawat, M. (2021). Transnational cybercrime: Issue of jurisdiction. *Issue 2 Int'l JL Mgmt. & Human., 4*, 253.

Timmerman, B., Mehta, P., Deb, P., Gallagher, K., Dolan-Gavitt, B., Garg, S., & Greenstadt, R. (2023). Studying the Online Deepfake Community. *Journal of Online Trust and Safety, 2*(1). https://doi.org/10.54501/jots.v2i1.126

Tuysuz, M. K., & Kılıç, A. (2023). Analyzing the Legal and Ethical Considerations of Deepfake Technology. *Interdisciplinary Studies in Society, Law, and Politics, 2*(2), 4-10. https://doi.org/10.61838/kman.isslp.2.2.2

Twomey, J., Ching, D., Aylett, M. P., Quayle, M., Linehan, C., & Murphy, G. (2023). Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. *PLoS one, 18*(10), e0291668. https://doi.org/10.1371/journal.pone.0291668

Ussenova, M. (2023). Legal awareness and precautions: safeguarding against misuse of DeepFake technology. *Scientific Collection «InterConf»*(161), 79-82.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology innovation management review, 9*(11). https://doi.org/10.22215/timreview/1282

Received on 15-02-2025 Accepted on 01-04-2025 Published on 16-04-2025

https://doi.org/10.6000/2817-2302.2025.04.02